Harmonic Analysis

Jacob Denson

January 31, 2022

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 sove 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 on the line.

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. In particular, it is essential to utilize all the convergence techniques developed through the rigorous study of analysis. Under pointwise convergence, the representation of a function by Fourier series need not be unique. Uniform convergence is more useful, and uniform convergence holds for all smooth functions, but does not hold if we only assume a function is continuous. Thus we must introduce more subtle methods.

Chapter 1

Introduction

One fundamental family of oscillatory functions in mathematics are the trigonometric functions

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

The value ϕ is the *phase* of the oscillation, C is the *amplitude*, and $s/2\pi$ is the *frequency* of the oscillation. These oscillatory functions occur in many situations; for instance, in the study of the solution of the harmonic oscillator. The main topic of Fourier analysis is to study how well one may represent a general function as an analytical combination of these trigonometric functions. In the periodic setting, we fix a function $f : \mathbf{R} \to \mathbf{C}$ 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 coefficients 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, given a function f, is to find a family $\{A_m\}$, $\{B_m\}$, and C which 'represents' the 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, Fourier found a way to formally associate a formal trigonometric series with any integrable periodic function. If the function is differentiable, then the trigonometric series gives a trigonometric expansion for the function. But even if this series does not give a trigonometric expansion for this function, the series itself still reflects many important properties of the function, which are of interest independant of their convergence to the function f.

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

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

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

We note that these values may still be defined even if f is not a trigonometric polynomial. Thus given *any* periodic integrable function f, a reasonable candidate for the coefficients is given by the values A_m, B_m, and C above. Unlike when f is a trigonometric polynomial, we can have infinitely many non-zero coefficients.

There is an additional choice of oscillatory functions, which replaces the sine and cosine with a single family of trigonometric functions, and thus gives a more notationally convenient analysis. For $\xi, t \in \mathbf{R}$, we let $e_\xi(t) = e^{2\pi \xi i t}$. For each integer $n \in \mathbf{Z}$, e_n 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 we can use orthogonality to find a natural choice of an expansion

$$f(t) \sim \sum_{n \in \mathbf{Z}} C_n e^{2\pi n i t},$$

given by setting

$$C_n = \int_0^1 f(t)\overline{e_n(t)}\, dt = \int_0^1 f(t) e^{-2\pi i n t}\, dt.$$

Euler's formula $e^{nit} = \cos(nt) + i\sin(nt)$ 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 used in Fourier analysis today.

To summarize, we have shown a periodic integrable 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. Given a periodic, integrable 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 basic purposes, the exponential notation is just a simple way to represent the oscillations of sines and cosines in a unified manner.

1.3 The Fourier Transform

For a general function $f : \mathbf{R} \to \mathbf{C}$, we cannot rely *just* on orthogonality, because the functions $\sin(2\pi mx)$ 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

$$f(t) \sim \int_{-\infty}^{\infty} \widehat{f}(\xi) e_\xi(t)\, d\xi.$$

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 non-compactness causes issues which are not present in the case of Fourier series, and so the Fourier series theory is often a simpler theory to begin with.

1.4 Multidimensional Theory

Finally, we note that the Fourier series and Fourier transform are not relegated to a one dimensional theory. If $f : \mathbf{R}^d \to \mathbf{C}$ is periodic, in the sense that $f(x+n) = f(x)$ for each $x \in \mathbf{R}^d$ and $n \in \mathbf{Z}^d$, then we can consider the natural higher dimensional Fourier series

$$f(t) \sim \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e_n(t)$$

where for each $\xi \in \mathbf{R}^d$, $e_\xi : \mathbf{R}^d \to \mathbf{C}$ is the function given for each $t \in \mathbf{R}^d$ by setting $e_\xi(t) = e^{2\pi i \xi \cdot t}$, and

$$\widehat{f}(n) = \int_{[0,1]^d} f(t) \overline{e_n(t)}\, dt$$

Similarily, for $f : \mathbf{R}^d \to \mathbf{C}$, we can consider the Fourier inversion formula

$$f(t) \sim \int_{\mathbf{R}^d} \widehat{f}(\xi) e_\xi(x)\, d\xi$$

where for each $\xi \in \mathbf{R}^d$,

$$\widehat{f}(\xi) = \int_{\mathbf{R}^d} f(t)\overline{e_\xi(t)}$$

The basic theory of Fourier series and the Fourier transform in one dimension extends naturally to higher dimensions, as do the basic theories of orthogonality. On the other hand, the theory of convergence in higher dimensions requires much greater regularity in higher dimensions and many fundamental questions about the convergence of Fourier series here more nuance than in the lower dimensional theory.

1.5 Examples of Expansions

Before we get to the real work, let's start by computing some examples of Fourier series and examples of the Fourier transform. We also illustrate the convergence properties of these series, which we shall look at in more detail later.

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

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

Thus

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

Thus we have a formal representation

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

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

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

In this case, we still have

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

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

Example. *The tent function*

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

is even, and therefore has a purely real Fourier expansion

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

Thus we obtain an expansion

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

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

Example. *Consider the characteristic function*

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

Then

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

Hence we may write

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

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

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

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

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

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

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

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

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

This means

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

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

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

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

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

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

Chapter 2

Fourier Series

Let us now focus on the theory of *Fourier series* 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 $|t| = \min_{n \in \mathbf{Z}} |t+n|$ for $t \in \mathbf{T}$. The Lebesgue measure on $\mathbf{R}$ induces a natural Borel measure on $\mathbf{T}$, such that for any periodic function $f : \mathbf{T} \to \mathbf{C}$,

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

It will also be of interest to consider the higher dimensional torii $\mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$, which naturally has the induced product metric and measure from $\mathbf{T}$. For each $f \in L^1(\mathbf{T}^d)$, we associate the *formal trigonometric series*

$$\sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot t}$$

where for each $n \in \mathbf{Z}^d$,

$$\widehat{f}(n) = \int_{\mathbf{T}^d} f(t) e^{-2\pi i n \cdot t}\, dt.$$

If $\{\widehat{f}(n)\}$ is an absolutely summable sequence, then one can interpret the formal trigonometric series nonformally as an infinite series, and we would then hope that for each $t \in \mathbf{T}^d$,

$$f(t) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot t}.$$

It turns out that, under the assumption of absolute summability, this equation does hold provided f is a continuous function on $\mathbf{T}^d$. We will eventually see that the condition that the Fourier series of f is absolutely summable under the assumption that $f \in C^\infty(\mathbf{T}^d)$. We will be able to prove these facts immediately after we prove some basic symmetry properties of the Fourier series.

2.1 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, if $h = f + g$, then $\widehat{h} = \widehat{f} + \widehat{g}$. Linearity is *essential* to most methods in this book; many problems about nonlinear transforms remain unsolved. The Fourier series is also stable under various transformations which occur in analysis, which makes the Fourier series tractable to analyze, and therefore useful. We summarize these properties here:

- Given $f \in L^1(\mathbf{T}^d)$, define $\text{Con}f, \text{Ref}f \in L^1(\mathbf{T}^d)$ by setting $\text{Con}f(x) = \overline{f(x)}$ and $\text{Ref}f(x) = f(-x)$. Then

  $$\widehat{\text{Con}f} = (\text{Con} \circ \text{Ref})\widehat{f}.$$

  and

  $$\widehat{\text{Ref}f} = \text{Ref}\widehat{f}.$$

  As a corollary, if f is real-valued, then

  $$\widehat{f} = \widehat{\text{Con}f} = (\text{Con} \circ \text{Ref})\widehat{f}$$

  In other words, for each $n \in \mathbf{Z}^d$,

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

  It also follows from the reflection symmetry that if $f \in L^1(\mathbf{T}^d)$ is odd, then $\widehat{f}$ is odd, and if $f \in L^1(\mathbf{T}^d)$ is even, $\widehat{f}$ is even.

- For each $s \in \mathbf{R}^d$, and $m \in \mathbf{Z}^d$, and any $f \in L^1(\mathbf{T}^d)$, define the translation and frequency modulation operators Trans_s and Mod_m by setting

  $$(\text{Trans}_s f)(t) = f(t-s) \quad \text{and} \quad (\text{Mod}_m f)(t) = e_m(t) f(t).$$

  Similarly, for each function $C : \mathbf{Z}^d \to \mathbf{C}$, for each $m \in \mathbf{Z}^d$ and $\xi \in \mathbf{R}$, define

  $$(\text{Trans}_m C)(n) = C(n-m) \quad \text{and} \quad (\text{Mod}_\xi C)(n) = e_\xi(n) C(n).$$

  Then for any $f \in L^1(\mathbf{T}^d)$, $\widehat{\text{Trans}_s f} = \text{Mod}_{-s} \widehat{f}$, and $\widehat{\text{Mod}_m f} = \text{Trans}_{-m} \widehat{f}$.

- An easy integration by parts shows that if $f \in C^\infty(\mathbf{T}^d)$, then for any $k \in \{1, \dots, d\}$,

  $$\widehat{D^k f}(n) = 2\pi i n_k \widehat{f}(n)$$

  for each $n \in \mathbf{Z}^d$. The proof follows from an easy integration by parts, so the claim is actually true for any $f \in L^1(\mathbf{T}^d)$ with a weak derivative $D^k f$ in $L^1(\mathbf{T}^d)$. Iterating this argument shows that, assuming the required weak derivatives exist,

  $$\widehat{D^\alpha f}(n) = (2\pi i n)^\alpha \widehat{f}(n).$$

Remark. We note that if $f \in L^1(\mathbf{T})$ is even, then $\widehat{f}$ is even, so formally

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

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.

2.2 Unique Representation of a Function?

To study the convergence properties of Fourier series, we begin by studying whether a function is uniquely determined by it's Fourier coefficients, which would be certainly true if the Fourier series held. However, such a statement is clearly cannot be true for all $f \in L^1(\mathbf{T}^d)$, since the Fourier coefficients of a function depend only on the *distributional* properties of f, i.e. those that can be obtained through integration. In particular, if two integrable functions f and g agree on a set of measure zero, then they have the same Fourier coeffients depends only on the equivalence class of f in $L^1(\mathbf{T}^d)$, with functions identified if they are equal almost everywhere. Nonetheless, if $f \in C(\mathbf{T}^d)$ then there is no way to edit f on a set of measure zero while preserving continuity. Thus we can hope for unique Fourier coefficients in the setting of continuous functions.

Theorem 2.1. *Suppose $f \in L^1(\mathbf{T}^d)$. If $\widehat{f}(n) = 0$ for all $n \in \mathbf{Z}^d$, then f vanishes at all it's continuity points.*

Proof. It suffices to prove that if $f \in L^1(\mathbf{T}^d)$ 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 $\delta > 0$ such that if $|x| \leqslant \delta$, $f(x) > f(0)/2$. Consider the trigonometric polynomial

$$g(x) = \prod_{k=1}^{d} [\varepsilon + \cos(2\pi x_k)] = \prod_{k=1}^{d} \left[\varepsilon + \frac{e^{2\pi i x_k} + e^{-2\pi i x_k}}{2} \right],$$

and where $\varepsilon > 0$ is small enough that if $|x| \geqslant \delta$, then $g(x) \leqslant B < 1$. We can then choose $0 < \eta < \delta$ such that if $|x| < \eta$, $g(x) \geqslant A > 1$. Finally, if

δ is sufficiently small, we also have $g(x) > 0$ if $0 \leqslant |x| \leqslant \delta$. The series of trigonometric polynomials $g_n(x) = g(x)^n$ satisfy

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

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

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

On the other hand,

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

Thus we conclude

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

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 u and v are continuous at x, so we can apply the real-valued case to complete the proof in the case of complex values. $\square$

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

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

Corollary 2.3. *If $f \in C(\mathbf{T}^d)$ and $\widehat{f} \in L^1(\mathbf{Z}^d)$, then for each $x \in \mathbf{Z}^d$,*

$$f(x) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot x}$$

Proof. Since $\widehat{f} \in L^1(\mathbf{Z}^d)$, the sum

$$g(x) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot x}$$

converges *uniformly*. In particular, this implies that g is a continuous function. Moreover, it allows us to conclude that $\widehat{g}(n) = \widehat{f}(n)$ for each $n \in \mathbf{Z}$. But this means $f = g$. $\square$

In the next section we will show that if $f \in C^m(\mathbf{T}^d)$, then

$$\widehat{f}(n) = O(\langle n \rangle^{-m}).$$

In particular, if $m \geqslant d+1$, then the Fourier series of f is integrable. Moreover, if $f \in C^\infty(\mathbf{T})$, then using the Fourier series equation for the derivative of a function, for each multi-index α, we conclude that for all $x \in \mathbf{R}^d$

$$(D^\alpha f)(x) = \sum_{n \in \mathbf{Z}^d} (2\pi i n)^\alpha \widehat{f}(n) e^{2\pi i n \cdot x}.$$

On the other hand, suppose $\{a_n : n \in \mathbf{Z}^d\}$ such that $|a_n| \lesssim_m \langle n \rangle^{-m}$ for all $m > 0$, then the infinite sum

$$\sum_{n \in \mathbf{Z}^d} a_n e^{2\pi i n \cdot x}$$

and all it's derivatives converge uniformly to an infinitely differentiable function with the Fourier coefficients $\{a_n\}$. Thus there is a perfect duality between infinitely differentiable functions and arbitrarily fast decaying sequences of integers. In more advanced contexts, like distribution theory, this duality is very useful for studying the Fourier transform in a much more general setting.

2.3 Quantitative Bounds on Fourier Coefficients

There are various reasons why one would not be completely satisfied by the convergence result above. Unlike with the case of a Taylor series, the Fourier series can be applied to a much more general family of situations. There is no hope of the Fourier series being integrable *and* obtaining a pointwise convergence result unless we are dealing with continuous functions, because any absolutely summable trigonometric series sums up to a continuous function. Thus we must analyze non-integrable families of coefficients if we are to obtain deeper convergence properties of the Fourier series for non-continuous functions.

On the other hand, in practical contexts, one might argue that the functions dealt with can be assumed arbitrarily smooth, so the picture established in the last section seems rather complete. However, even if this is true it is still important to study more *qualitative questions* about the Fourier series. Instead of taking the infinite Fourier series, we take a finite sum. For a function $f \in L^1(\mathbf{T})$, it is most natural to consider the partial sums

$$S_N f(x) = \sum_{n=-N}^{N} \widehat{f}(n) e^{2\pi i n x}.$$

In higher dimensions, no canonical 'cutoff' exists. Two possible options are *spherical summation*

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

and *square summation*

$$\sum_{n_1, \dots, n_d = -N}^{N} \widehat{f}(n) e_n.$$

There are subtle differences in these operators which cause problems in the higher dimensional theory. Right now, all we assume is that we are consider an increasing family of sets $\{E_N\}$ in $\mathbf{Z}^d$ with $\lim_{N \to \infty} E_N = \mathbf{Z}^d$, and we then define

$$S_N f = \sum_{n \in E_N} \widehat{f}(n) e_n$$

for $f \in L^1(\mathbf{T}^d)$. A natural question now is whether $S_N f$ is *qualitatively similar* to the function f globally rather than just pointwise. The most natural way to measure how similar two functions are from the perspective of analysis is via measuring the differences with respect to a suitable *norm*. For instance, under the assumptions of the last section, we not only get pointwise convergence at each point, but *uniform convergence*.

Theorem 2.4. *Suppose $f \in C(\mathbf{T})$ and $\widehat{f} \in L^1(\mathbf{Z})$. Then*

$$\lim_{N \to \infty} \|S_N f - f\|_{L^\infty(\mathbf{R})}.$$

In other words, $S_N f$ converges uniformly *to f instead of pointwise.*

Proof. We know that for each $x \in \mathbf{T}^d$,

$$f(x) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot x}.$$

A simple application of the triangle inequality shows that

$$|f(x) - S_N f(x)| \leqslant \sum_{n \notin E_N} |\widehat{f}(n)|.$$

Since the Fourier coefficients are absolutely summable, for each $\varepsilon > 0$, there is N_0 such that for $N \geqslant N_0$,

$$\sum_{N \notin E_N} |\widehat{f}(n)| \leqslant \varepsilon,$$

and thus $\|f - S_N f\|_{L^\infty(\mathbf{T}^d)} \leqslant \varepsilon$. □

Another question one might ask is the *rate of convergence* of the function f. In this situation, things are quite bad even in the setting of the previous setting. For general elements of $C(\mathbf{T})$ with integrable Fourier coefficients, the convergence of $\|f - S_N f\|_{L^\infty(\mathbf{T}^d)}$ as $N \to \infty$ can be *as slow as any convergent sequence*.

Theorem 2.5. *Let $\{a_n : n \in \mathbf{Z}^d\}$ be any sequence of coefficients with* $\lim_{|n| \to \infty} a_n = 0$*. Then there exists $f \in C(\mathbf{T}^d)$ such that $\widehat{f} \in L^1(\mathbf{Z})$, but $\widehat{f}(n) = a_n$ for infinitely many $n \in \mathbf{Z}^d$.*

Proof. For each $1 \leqslant k < \infty$, pick n_k such that $|a_{n_k}| \leqslant 1/2^k$ and such that the family $\{n_k\}$ is distinct. Then define

$$f(x) = \sum_{k=1}^{\infty} a_{n_k} e^{2\pi i n_k \cdot x}.$$

The absolute convergence of the right hand side shows $f \in C(\mathbf{T}^d)$, and that $\widehat{f}(n_k) = a_{n_k}$ for each $1 \leqslant k \leqslant \infty$. □

A natural question is whether we *can* get quantitative convergence results for functions under additional assumptions. For instance, do we get faster convergence rates if $\|f\|_{L^\infty(\mathbf{T}^d)}$ is small (i.e. we have uniform control on the magnitude of f) rather than just if $\|f\|_{L^1(\mathbf{T}^d)}$ is small.

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*

$$\|\chi_I - S_N \chi_I\|_{L^2(\mathbf{T})} \sim 1/\sqrt{N},$$

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 about $1/\varepsilon^2$ 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 $f \in C^{m+1}(\mathbf{T}^d)$, then we will see

$$\|f - S_N f\|_{L^2(\mathbf{T})} \lesssim 1/N^{m/2},$$

so we require significantly less notes to simulate this sound, i.e. $\varepsilon^{-2/m}$. In this case a piano can simulate these sounds much more accurately.

Another question is whether $S_N f$ is stable under pertubations. For instance, if we replace f with a function g close to f the original function, is $S_N f$ close to $S_N g$? This is of interest in many partical applications, where error terms are inherently present. If an operator is unstable under pertubations that it is unpractical to use it in an application to a real life situation. Again, the best way to measure the error terms are using an appropriate norm space.

These examples show that working with certain norms is an important way to understand the deeper properties of the Fourier series. It is an important property of norm spaces that most questions are equivalent to questions in the *completion* of that norm space. For instance, if one wants to use the norm $\|\cdot\|_{L^1(\mathbf{T}^d)}$ to analyze the space $C(\mathbf{T}^d)$, most questions are equivalent to questions about the completion of $C(\mathbf{T}^d)$, i.e. the space $L^1(\mathbf{T}^d)$ of all integrable functions. Moreover, working in the completion of a space enables us to employ many functional analysis arguments which make working with the more general space essential to many modern arguments. Despite the fact that we will be analyzing functions that one never deals with in 'practical situations', using these functions is a useful tool to determine the quantitative behaviour of more regular functions with respect to a norm.

2.4 Boundedness of Partial Sums

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 $\|\widehat{f}\|_{L^q(\mathbf{Z}^d)} \lesssim \|f\|_{L^p(\mathbf{T}^d)}$ for particular values of p and q. This does not explicitly answer a question about convergence, but still shows that the Fourier series is stable under small pertubations in the norm on $L^p(\mathbf{T}^d)$. The first inequality we give is trivial, but is certainly tight, e.g. for $f(t) = e_n(t)$.

Theorem 2.6. *For any $f \in L^1(\mathbf{T}^d)$, $\|\widehat{f}\|_{L^\infty(\mathbf{T}^d)} \leqslant \|f\|_{L^1(\mathbf{T}^d)}$.*

Proof. We just take absolute values into the oscillatory integral defining the Fourier coefficients, calculating that for any $n \in \mathbf{Z}^d$,

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

which was the required bound. □

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^1 norms of the phase spaces, other than by taking a naive absolute bound into the L^1 norm. The only L^p norm where we can get a completely satisfactory bound is for $p = 2$, where we can use Hilbert space techniques; this should be expected to be very useful since orthogonality was implicitly used to define the Fourier series.

Theorem 2.7. *For any function $f \in L^2(\mathbf{T}^d)$, $\|\widehat{f}\|_{L^2(\mathbf{Z}^d)} = \|f\|_{L^2(\mathbf{T}^d)}$.*

Proof. With respect to the normalized inner product on the space $L^2(\mathbf{T}^d)$,the calculations of the last chapter tell us that the exponentials $\{e_n : n \in \mathbf{Z}^d\}$ are an orthonormal family of functions, in the sense that for distinct pair

$n, m \in \mathbf{Z}^d$, $(e_n, e_m) = 0$ and $(e_n, e_n) = 1$. Since $\widehat{f}(n) = (f, e_n)$, we apply Bessel's inequality to conclude

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

The exponentials $\{e_n\}$ are actually an orthonormal basis for $L^2(\mathbf{T}^d)$; there are many ways to see this (the Stone-Weirstrass theorem, for instance). The most convenient way for us will be to note that if $f \in C^\infty(\mathbf{T}^d)$, then we have shown that if $(f, e_n) = 0$ for all $n \in \mathbf{Z}^d$, then $f = 0$. But $S_N f$ converges to f in $L^2(\mathbf{T}^d)$ (it actually converges uniformly), and so

$$\begin{aligned} \|f\|_{L^2(\mathbf{T}^d)} = \lim_{N\to\infty} \|S_N f\|_{L^2(\mathbf{T}^d)} &= \lim_{N\to\infty} \left(\sum_{n \in E_N} |\widehat{f}(n)|^2 \right)^{1/2} \\ &= \left(\sum_{n \in \mathbf{Z}^d} |\widehat{f}(n)|^2 \right)^{1/2} = \|\widehat{f}\|_{L^2(\mathbf{Z}^d)}. \end{aligned}$$

This is Parseval's inequality for $C^\infty(\mathbf{T}^d)$. Now a density argument will give the general result. If $f \in L^2(\mathbf{T}^d)$ is a general element, then for each $\varepsilon > 0$ we can find $f_\varepsilon \in C^\infty(\mathbf{T}^d)$ such that $\|f_\varepsilon - f\|_{L^2(\mathbf{T}^d)} \leqslant \varepsilon$. Then Bessel's inequality

$$|\|f\|_{L^2(\mathbf{T}^d)} - \|\widehat{f}\|_{L^2(\mathbf{T}^d)}| \leqslant |\|f\|_{L^2(\mathbf{T}^d)} - \|f_\varepsilon\|_{L^2(\mathbf{T}^d)}| + |\|\widehat{f_\varepsilon}\|_{L^2(\mathbf{Z}^d)} - \|\widehat{f}\|_{L^2(\mathbf{Z}^d)}| \leqslant 2\varepsilon.$$

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

This equality makes the Hilbert space $L^2(\mathbf{T}^d)$ 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 2.8. *If $1 \leqslant p \leqslant 2$, and $f \in L^p(\mathbf{T}^d)$, then $\|\widehat{f}\|_{L^{p*}(\mathbf{Z}^d)} \leqslant \|f\|_{L^p(\mathbf{T}^d)}$.*

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

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

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.

2.5 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 norm spaces, and we want to show that for each $x \in X$, $\lim_{n\to\infty} T_n(x) = T(x)$, where T is another bounded operator. Suppose there is a dense set $X_0 \subset X$ such that for each $x_0 \in X_0$, $\lim_{n\to\infty} T_n(x_0) = T(x_0)$, and the family of operators $\{T_n\}$ are *uniformly* bounded in operator norm. Then for any $x \in X$,

$$\|T_n(x) - T(x)\| \leqslant \|T_n(x) - T_n(x_0)\| + \|T_n(x_0) - T(x_0)\| + \|T(x_0) - T(x)\|.$$

If we choose x_0 such that $\|x - x_0\| \leqslant \varepsilon$, 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 $n \to \infty$. If we are working in a Banach space, the uniform boundedness says obtaining a uniform operator norm bound on $\{T_n\}$ is the *only* way to obtain this convergence.

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

Proof. We claim the lemma is true for the characteristic function χ_I of a cube I. If $I = [a_1, b_1] \times \cdots \times [a_d, b_d]$, then it is simple to calculate that

$$\widehat{\chi_I}(n) = \prod_{k=1}^{d} \frac{e_n(-b_k) - e_n(-a_k)}{-in} = O(1/n)$$

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 linear functions on $L^1(\mathbf{T}^d)$. Since $\lim_{|n|\to\infty} \Lambda_n(f) = 0$ for any step function f, and the step functions are dense in $L^1(\mathbf{T}^d)$, we conclude that

$$\lim_{|n|\to\infty} \Lambda_n(f) = 0$$

for all $f \in L^1(\mathbf{T}^d)$. □

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 $O(1/n)$. For instance, we have shown that there are continuous functions whose Fourier decay is arbitrarily slow. This is precisely the penalty for using a soft type analytical argument. Nonetheless, for smoother functions, we can obtain a uniform decay rate, which is our goal in the next section.

2.6 Smoothness and Decay

The next theorem obtains sharper bounds for smoother functions, and is an instance of a general phenomenon relating the duality because decay and smoothness in phase and frequency space.

Theorem 2.10. *If $f \in C^m(\mathbf{T}^d)$, then for each $n \in \mathbf{Z}^d$,*

$$|\widehat{f}(n)| \lesssim_{d,m} |n|^{-m} \max_{1 \leqslant i \leqslant d} \|\partial_i^m f\|_{L^1(\mathbf{T}^d)}.$$

Proof. We have

$$\widehat{\partial_i^m f}(\xi) = (2\pi i \xi_i)^m \widehat{f}(\xi).$$

Thus

$$|\widehat{f}(\xi)| \leqslant \frac{|\partial_i^m f|(\xi)|}{(2\pi|\xi_i|)^m} \leqslant \frac{\|\partial_i^m f\|_{L^1(\mathbf{T}^d)}}{(2\pi|\xi_i|)^m}.$$

But taking infima over all $1 \leqslant i \leqslant d$, we find

$$|\widehat{f}(\xi)| \leqslant \frac{\max_{1\leqslant k\leqslant d} \|\partial_i^m f\|_{L^1(\mathbf{T}^d)}}{[2\pi \max |\xi_i|]^m} \leqslant \frac{d^{1/2}}{(2\pi)^m} \frac{\max_{1\leqslant i\leqslant d} \|\partial_i^m f\|_{L^1(\mathbf{T}^d)}}{|\xi|^m}. \qquad \square$$

On the other hand, if $|\widehat{f}(n)| \lesssim 1/|n|^{d+m}$, it is easy to see from the pointwise convergence of the Fourier series that $f \in C^m(\mathbf{R}^d)$. Note, however, that the introduction of the factor of d here gives a large gap between obtaining decay from smoothness and smoothness and decay when d is large, which is often a tricky problem to control when studying problems using harmonic analysis.

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 $x, h \in \mathbf{T}^d$. We define

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

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

Theorem 2.11. *If $f \in C^{0,\alpha}(\mathbf{T}^d)$, then $|\widehat{f}(n)| \lesssim_d \|f\|_{C^{0,\alpha}(\mathbf{T}^d)} |n|^{-\alpha}$ for all $n \in \mathbf{Z}^d$.*

Proof. Fix $n \in \mathbf{Z}^d$. Then there is some $k \in \{1, \dots, d\}$ such that $|n_k| \gtrsim_d |n|$. We calculate that by periodicity,

$$\widehat{f}(n) = -\int_{\mathbf{T}^d} f(x + e_k/n_k) \overline{e_n(x)}\, dx,$$

so

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

Thus taking in absolute values and applying Hölder continuity gives

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

We also have a weaker converse statement, which shows f is Hölder continuous if it's Fourier series decays fast enough.

Theorem 2.12. *Fix $f \in L^1(\mathbf{T}^d)$. Then*

$$\|f\|_{C^{0,\alpha}(\mathbf{T}^d)} \lesssim_d \sup_{n \in \mathbf{Z}^d} |n|^{d+\alpha} |\widehat{f}(n)|.$$

Proof. Let $A = \sup_{n \in \mathbf{Z}^d} |n|^{d+\alpha} |\widehat{f}(n)|$. Then $\widehat{f} \in L^1(\mathbf{Z}^d)$, so the Fourier inversion formula implies that for almost every $x \in \mathbf{T}^d$,

$$f(x) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot x}.$$

Then for $|h| < 1$,

$$f(x+h) - f(x) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n \cdot x} \left(e^{2\pi i n \cdot h} - 1 \right).$$

Now $|e^{2\pi i n \cdot h} - 1| \lesssim \min(1, |n||h|)$, so

$$\left| \sum_{|n| \leqslant 1/|h|} \widehat{f}(n) e^{2\pi i n \cdot x} \left(e^{2\pi i n \cdot h} - 1 \right) \right| \leqslant A|h| \sum_{|n| \leqslant 1/|h|} \frac{1}{|n|^{d-1+\alpha}} \lesssim_d A|h||h|^{\alpha - 1} = A|h|^\alpha$$

and

$$\begin{aligned} \left| \sum_{|n| \geqslant 1/|h|} \widehat{f}(n) e^{2\pi i n \cdot x} \left(e^{2\pi i n \cdot h} - 1 \right) \right| &\leqslant 2A \sum_{|n| \geqslant 1/|h|} 1/|n|^{d+\alpha} \\ &\lesssim_d 2A|h|^\alpha. \end{aligned}$$

Combining these two calculations shows that

$$|f(x+h) - f(x)| \lesssim_d A|h|^\alpha,$$

so $\|f\|_{C^{0,\alpha}(\mathbf{T}^d)} \lesssim_d A$. $\square$

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

$$\widehat{\mu}(n) = \int_{\mathbf{T}^d} e^{-2\pi i n \cdot x} d\mu(x).$$

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 finite measures. One can verify that

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

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. Indeed, it is not too difficult to prove that a finite Borel measure μ on $\mathbf{T}^d$ is absolutely continuous with respect to the Lebesgue measure if and only if

$$\lim_{|y|\to 0} \int_{\mathbf{R}^d} d|\mu(x+y) - \mu(x)| \to 0,$$

(we show that integrable functions satisfy this property in the next section) which shows that integrable functions are precisely the measures such that, in a certain sense, $\mu(x+y) \approx \mu(x)$ for small y.

2.7 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 $f, g \in L^1(\mathbf{T}^d)$, we define

$$(f * g)(x) = \int_{\mathbf{T}^d} f(y) g(x-y)\, day.$$

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

Lemma 2.13. *For any $1 \leqslant p < \infty$, and $f \in L^p(\mathbf{T}^d)$,*

$$\lim_{h\to 0} Trans_h f = f$$

in $L^p(\mathbf{T}^d)$.

Proof. If f is $C^1(\mathbf{T}^d)$, then $|f(x+h) - f(x)| \lesssim_f h$ uniformly in x, implying that $\|\text{Trans}_h f - f\|_{L^p(\mathbf{T}^d)} \leqslant \|\text{Trans}_h f - f\|_{L^\infty(\mathbf{T}^d)} \lesssim_f h$, and so $\text{Trans}_h f \to f$ in all the spaces $L^p(\mathbf{T}^d)$. We have $\|\text{Trans}_h f\|_{L^p(\mathbf{T}^d)} = \|f\|_{L^p(\mathbf{T}^d)}$, so the operators $\{\text{Trans}_h\}$ are uniformly bounded. Since $C^1(\mathbf{T}^d)$ is dense in $L^p(\mathbf{T}^d)$ for $1 \leqslant p < \infty$, we conclude that $\lim_{h\to 0} \text{Trans}_h f = f$ for all $f \in L^p(\mathbf{T}^d)$. $\square$

Theorem 2.14. *Convolution has the following properties:*

- *If $f \in L^p(\mathbf{T}^d)$ and $g \in L^q(\mathbf{T}^d)$, for $1/p + 1/q = 1$, then $f * g$ is uniformly continuous.*
- *If $f \in L^p(\mathbf{T}^d)$ and $g \in L^q(\mathbf{T}^d)$, 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 convolution integral formula almost everywhere, and*

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

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

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

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

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

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

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

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

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

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

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

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

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 $t \in E + F$. $\square$

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

Example. *Given a function $f \in L^1(\mathbf{T}^d)$ we define the* autocorrelation function

$$R(\tau) = \int_{\mathbf{T}^d} f(t+\tau)\overline{f(t)}\, dt.$$

Then R is the convolution of $f(t)$ with $g(t) = \overline{f(-t)}$. Thus for $f \in L^1(\mathbf{T}^d)$, $R \in L^1(\mathbf{T}^d)$, and

$$\widehat{R}(n) = \widehat{f}(n)\overline{\widehat{f}(n)} = |\widehat{f}(n)|^2.$$

The function $\widehat{R}$ is known as the power spectrum *of f.*

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 $\delta > 0$,

$$\lim_{n \to \infty} \int_{|t| > \delta} |K_n(t)| \to 0. \tag{2.1}$$

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

Example. *The simplest way to obtain a good kernel is to fix $K \in L^1(\mathbf{T}^d)$ with*

$$\int_{\mathbf{T}^d} K(x)\, dx = 1,$$

and to define

$$K_n(x) = \begin{cases} n^d \cdot K(nx) & : |x_1|, \dots, |x_d| \leqslant 1/n, \\ 0 & : \textit{otherwise}. \end{cases}$$

Then $\|K_n\|_{L^1(\mathbf{T})} = 1$ for all $n > 0$, and K_n is eventually supported on every small ball around the origin, which implies (2.1). *If $K \in L^\infty(\mathbf{T}^d)$, then the resulting sequence $\{K_n\}$ is also an approximation to the identity.*

Theorem 2.15. *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 $f \in C(\mathbf{T}^d)$, and $K_n * f$ converges to f in $L^p(\mathbf{T}^d)$ if $f \in L^p(\mathbf{T}^d)$, 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 $C(\mathbf{T}^d)$. But a density argument allows us to conclude that $K_n * f \to f$ in $L^p(\mathbf{T})$ for each $f \in L^p(\mathbf{T}^d)$, for $1 \leqslant p < \infty$. To obtain pointwise convergence for t in the Lebesgue set of f, we calculate

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

Let $A(\delta) = \delta^{-d} \int_{|s|<\delta} |f(t-s) - f(t)|$. 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 $|K_n(s)| \lesssim n^d$,

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

Thus we have a bound

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

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

$$|(K_n*f)(t)-f(t)| \lesssim_d \sum_{k=0}^{m} \frac{A(2^k/n)}{2^{dk}} + \|A\|_\infty \sum_{k=m}^{\infty} \frac{1}{2^{dk}} = \sum_{k=0}^{m} \frac{A(2^k/n)}{2^{dk}} + O_d\left(1/2^{dm}\right).$$

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

2.8 The Dirichlet Kernel

For simplicity, let us now focus exclusively on the case $d = 1$ with the canonical summation operators S_N. For $f \in L^1(\mathbf{T})$, we calculate that

$$(S_N f)(t) = \sum_{n=-N}^{N} \widehat{f}(n) e_n(t) = \int_{\mathbf{T}^d} f(x) \left(\sum_{n=-N}^{N} e_n(t-x) \right) dx.$$

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

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

and so $S_N f = f * D_N$. Thus analyzing this convolution is *key* to understanding the partial summation operators.

Remark. In the higher dimensional case, we can consider the operators

$$K_N(t) = \sum_{n \in E_N} e_n(t).$$

The behaviour of these functions is highly dependant on the choice of the sets E_N, and we thus leave the higher dimensional analysis to a different time.

Theorem 2.16. *For any integer N and $t \in \mathbf{T}$,*

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

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

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

□

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

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

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

Proof. If we consider the linear functionals $\Lambda_N f = (S_N f)(0) = (f * D_N)(0)$ on $C(\mathbf{T})$. If we let f_N be a continuous function approximating $\text{sgn}(D_N)$ for each N, then $|\Lambda_N f_N| \gtrsim \log N \cdot \|f_N\|_{L^\infty(\mathbf{T})}$. 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 $N \to \infty$. □

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 $f \in L^p(\mathbf{T})$, 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 $1 < p \leqslant \infty$, the Fourier series of f converges almost everywhere to f. The multivariate picture is more complicated and many questions remain open today; tensoring shows that $S_N f$ converges to f in $L^p(\mathbf{T}^d)$ if $f \in L^p(\mathbf{T}^d)$, *provided that we interpret $S_N f$ as a square summation*, and in 1970 Charles Fefferman showed that for square summation $S_N f$ converges to f almost everywhere. On the other hand, in 1971 Charles Fefferman showed that for spherical summation the *only* place we have norm convergence is in $L^2(\mathbf{T}^d)$. It remains an open question whether the partial spherical summation $S_N f$ converges to f almost everywhere.

2.9 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.*

Abel summation is even more general than Cesaro summation, as the following theorem shows.

Theorem 2.18. *A Cesaro summable sequence is Abel summable.*

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

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

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

$$\begin{aligned}
\left| (1-r)^2 \sum_{k=0}^{\infty} (k+1)\sigma_k r^k \right| &\leqslant (1-r)^2 \left(\sum_{k=0}^{N} (k+1)|\sigma_k| r^k + \varepsilon \sum_{k=N+1}^{\infty} (k+1) r^k \right) \\
&= (1-r)^2 \left(\sum_{k=0}^{N} (k+1)(|\sigma_k| - \varepsilon) r^k + \varepsilon \left[\frac{r^{n+1}}{1-r} + \frac{1}{(1-r)^2} \right] \right) \\
&\leqslant (1-r)^2 M \sum_{k=0}^{N} (k+1) r^k + \varepsilon r^{n+1}(1-r) + \varepsilon \\
&\leqslant (1-r)^2 M \frac{(N+1)(N+2)}{2} + \varepsilon r^{n+1}(1-r) + \varepsilon
\end{aligned}$$

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

2.10 Fejer Summation

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

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

where we have introduced a new kernel F_N, called the *Fejer kernel*. Here, we have a simple formula for the Cesaro means, i.e.

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

Thus the oscillations of the Dirichlet kernel are slightly dampened, and as a result, we can easily see that F_N is an approximation to the identity.

Theorem 2.19 (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 $f \in L^p(\mathbf{T})$.*

If we look at the Fourier expansion of the trigonometric polynomial $\sigma_N(f)$, viewing σ_N as a *Fourier multiplier operator*, we see that

$$\sigma_N f = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N} \right) \widehat{f}(n) e_n.$$

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.

Corollary 2.20. *If $f \in L^1(\mathbf{T})$ and $\widehat{f} = 0$, 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 $L^1(\mathbf{T})$, which means that $f = 0$ in $L^1(\mathbf{T})$, so $f = 0$ almost everywhere. $\square$

This corollary is often more useful than the more technical convergence statements due to it's relative simplicity. We will later see this result is also true for $d > 1$, via use of the Poisson summation formula for the Fourier transform.

Example. *We say $f \in L^1(\mathbf{T}^d)$ is* band limited *if it's Fourier series is supported on finitely many points. If $\{S_N\}$ is defined as before, and N is suitably large that E_N contains the support of $\widehat{f}$, then*

$$\widehat{f} = \widehat{f} \cdot \mathbf{I}_{E_N} = \widehat{f}\widehat{K_N} = \widehat{f * K_N}.$$

It thus follows from the previous result that $f = f * K_N$ almost everywhere. But this means we can adjust f on a set of measure zero such that $f \in C^\infty(\mathbf{T}^d)$.

2.11 Abel Summation

Let us now consider the Abel sum of the Fourier integrals. We begin by focusing on the one-dimensional case, as in the last section. Thus for $f \in L^1(\mathbf{T})$ we have

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

Thus, if we define the *Poisson kernel*

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

For each $r < 1$, this series converges uniformly for $t \in \mathbf{T}$, so P_r is a well-defined continuous function, and the uniform convergence shows that $A_r(f) = P_r * f$. As with the Fejer kernel, the family $\{P_r\}$ is *also* a good kernel as $r \to 1$. 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

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

Moreover,

$$\|P_r\|_{L^\infty(\mathbf{T})} \leqslant \frac{1-r^2}{1-2r+r^2} \leqslant \frac{2}{1-r}.$$

Thus the Poisson kernel is an approximation to the identity; the oscillation in the kernel cancels out as $r \to 1$.

Theorem 2.21. *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. We will see the kernel return when we study application of harmonic analysis to partial differential equations.

2.12 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 2.22. *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 2.23. *If $f \in L^1(\mathbf{T})$ 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. $\square$

2.13 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 2.24 (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. ☐

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 2.25 (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

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

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. ☐

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 2.26 (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

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

□

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 2.27. *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

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

2.14 Gibbs Phenomenon

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, the sequence $\{f_k\}$ does not converge to f.

Theorem 2.28. *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 *T*chebyshev 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. $\square$

Another proof uses the family of *Landau kernels*

$$L_n(x) = c_n \cdot \begin{cases} (1-x^2)^n & : -1 \leqslant x \leqslant 1 \\ 0 & : |x| \geqslant 1 \end{cases}$$

where c_n is chosen so that $\int_{-1}^1 L_n(x) = 1$. It is simple to show that the family $\{L_n\}$ is a

$$\begin{aligned} \int_{-1}^1 (1-x^2)^n \, dx \geqslant &= \sum_{k=0}^\infty \int_{1/2^{k+1} \leqslant |x| \leqslant 1/2^k} (1-x^2)^n \\ &\gtrsim \sum_{k=0}^\infty \frac{(1-1/4^{k+1})^n}{2^k} \\ &= \sum_{k=0}^\infty \exp\left(n \log(1 - 1/4^{k+1}) - k \log(2) \right) \\ &\geqslant \sum_{k=0}^\infty \exp\left(-n/4^k \right) / 2^k \\ &\gtrsim \sum_{k=\log_4 n}^\infty 1/2^k \gtrsim 1/2^{\log_4 n} = 1/n^{1/2} \end{aligned}$$

Thus $\|L_n\|_{L^\infty(\mathbf{R})} \leqslant n^{1/2}$ which can be used to show the family $\{L_n\}$ is an approximation to the identity. An important fact here is that if f is supported on $[-1/2, 1/2]$, then $L_N * f$ agrees with a polynomial on $[-1/2, 1/2]$, which can be used to approximate f by a polynomial on this region.

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

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

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 The Poisson Equation

Consider Poisson's equation

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial u^2}{\partial y^2} = 0$$

on the unit disk. Solutions are called *harmonic*. We can reduce this equation to a problem about Fourier series by writing

$$u(re^{2\pi i t}) = \sum_{n=0}^{\infty} a_n(r) e^{2\pi n i t}.$$

We consider a boundary condition, that $u(e^{2\pi i t}) = f(t)$ for some function $f(t)$ on $\mathbf{T}$. Working formally, noting that in radial coordinates,

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

and then taking Fourier series on each side, we find that for each $n \in \mathbf{Z}$,

$$a_n''(r) + a_n'(r)/r - 4\pi^2 n^2 a_n(t)/r^2 = 0.$$

The only *bounded* solution to this differential equation subject to the initial condition $a_n(1) = \widehat{f}(n)$ is $a_n(t) = \widehat{f}(n) r^{|n|}$. Thus we might guess that

$$u(re^{2\pi it}) = \sum_{n \in \mathbf{Z}} \widehat{f}(n) r^{|n|} e^{2\pi nit} = (P_r * f)(x),$$

where P_r is the Poisson kernel. Working backwards through this calculation shows that if $f \in L^1(\mathbf{T})$, then the function $u(re^{2\pi it}) = (P_r * f)(t)$ lies in $C^\infty(\mathbf{D})$ and

$$\lim_{r \to 1} \int_{\mathbf{T}} |u(re^{2\pi it}) - f(t)|\, dt = 0.$$

The next theorem shows this is the *only* harmonic function with this propety.

Theorem 3.3. *Suppose $f \in L^1(\mathbf{T})$. Then the function $u : \mathbf{D}^\circ \to \mathbf{C}$ defined for $r > 0$ and $t \in \mathbf{T}$ by setting*

$$u(re^{2\pi it}) = (A_r f)(t)$$

is the unique *harmonic function in $C^2(\mathbf{D}^\circ)$ such that*

$$\lim_{r \to 1} \int \left| u(re^{2\pi it}) - f(t) \right| dt = 0.$$

Proof. Suppose $u \in C^2(\mathbf{D})$ is harmonic. Then we can find functions $a_n(r)$ for each $n \in \mathbf{Z}$ such that

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

where

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

Because $u \in C^2(\mathbf{D})$, we see that $a_n \in C^2((0,1))$ and $a_n(r)$ is bounded as $r \to 0$. Interchanging integrals shows that

$$a_n''(r) + (1/r) a_n'(r) - (n^2/r^2) a_n(r) = 0.$$

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

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

so

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

In particular, the theorem above gives us a map from $L^1(\mathbf{T})$ to the space of harmonic functions on the interior of the unit disk. This is a very handy idea in classical harmonic analysis, and is exploited to it's fullest extent in the theory of Hardy spaces.

3.5 The Heat Equation on a Torus

Recall the heat equation. We are given an initial temperature distribution on $\mathbf{T}^d$. We wish to study the propogation of this temperature over time. If we let $u(x,t)$ denote the temperature density at $x \in \mathbf{T}^d$ and at time t, then this temperature evolves under the heat equation

$$\frac{\partial u}{\partial t} = \Delta u.$$

We let $f(x) = u(x,0)$ denote the initial heat distribution. To solve this heat equation, we expand u in a Fourier series, i.e. writing

$$u(x,t) = \sum_{n \in \mathbf{Z}^d} a_n(t) e^{2\pi i n \cdot x}.$$

We then formally find that for each $n \in \mathbf{Z}^d$,

$$a_n'(t) = -4\pi^2 |n|^2 a_n(t),$$

which we can solve to give

$$a_n(t) = \widehat{f}(n) e^{-4\pi^2 |n|^2 t}.$$

In particular, we would expect the solution to the heat equation would be given by letting

$$u(x,t) = \sum_{n \in \mathbf{Z}^d} \widehat{f}(n) e^{-4\pi^2 |n|^2 t} e^{2\pi n i t}.$$

As with Poisson's equation on the disk, we can write this as

$$u(x,t) = (H_t * f)(x)$$

where H_t is the *heat kernel*

$$H_t(x) = \sum_{n \in \mathbf{Z}^d} e^{-4\pi^2 |n|^2 t} e^{2\pi n i t}.$$

The rapid convergence of this sum implies that $H_t \in C^\infty(\mathbf{T}^d)$ and that $\widehat{H_t}(n) = e^{-4\pi^2 |n|^2}$ for each $n \in \mathbf{Z}^d$. To study this partial differential equation, it suffices to study the heat kernel H_t. Unlike in the case of the Poisson kernel however, we have no explicit formula for the heat kernel, which makes the kernel a little harder to work with.

Lemma 3.4. *The family $\{H_t : t > 0\}$ is an approximation to the identity.*

Proof. The Poisson summation formula implies that

$$H_t(x) = \frac{1}{(4\pi t)^{d/2}} \sum_{n \in \mathbf{Z}^d} e^{-|x+n|^2/4t}.$$

This shows that $H_t(x) \geqslant 0$, and that

$$\int_{\mathbf{T}^d} H_t(x) = \frac{1}{(4\pi t)^{d/2}} \int_{\mathbf{R}^d} e^{-|x|^2/4t}\, dx = \int_{\mathbf{R}^d} e^{-\pi |x|^2}\, dx = 1.$$

We claim that for $|x| \leqslant 1/2$,

$$\left| H_t(x) - \frac{e^{-x^2/4t}}{(4\pi t)^{d/2}} \right| \lesssim_d e^{-c/t},$$

where $c > 0$ is a universal constant. To prove this, we note this difference is equal to

$$
\begin{aligned}
(4\pi t)^{-d/2} \left| \sum_{n \neq 0} e^{-|x+n|^2/4t} \right| &\lesssim t^{-d/2} \sum_{n \neq 0} e^{-c'|n|^2/4t} \\
&\lesssim t^{-d/2} e^{-c'/2t} \sum_{n \neq 0} e^{-c'|n|^2/2} \\
&\lesssim_d t^{-d/2} e^{-c'/2t} \lesssim_d e^{-c/t}.
\end{aligned}
$$

This implies that for any fixed $\delta > 0$,

$$
\begin{aligned}
\int_{|x|>\delta} H_t(x) &\lesssim t^{-d/2} \int_{|x|>\delta} e^{-|x|^2/4t}\, dx + e^{-c/t} \\
&\lesssim_d t^{-d/2} e^{-\delta^2/4t} + e^{-c/t}
\end{aligned}
$$

which tends to zero as $t \to \infty$. Thus we have proved that H_t is an approximation to the identity. $\square$

Theorem 3.5. *For any $f \in L^1(\mathbf{T}^d)$, for $1 \leqslant p < \infty$. Then the function*

$$u(x,t) = (H_t * f)(x)$$

lies in $C^\infty(\mathbf{T}^d \times (0,\infty))$, and for $t > 0$ solves the heat equation. Moreover, u is the unique solution to the heat equation in $C^2(\mathbf{T}^d \times (0,\infty))$ such that

$$\lim_{t \to 0^+} \int_{\mathbf{T}^d} |u(t,x) - f(x)|\, dx = 0.$$

Proof. We have already shown the former statement by the fact that $\{H_t : t > 0\}$ is an approximation to the identity. To prove the latter statement, given $u \in C^2(\mathbf{T}^d \times (0,\infty))$, we can take a Fourier series, letting

$$a_n(t) = \int_{\mathbf{T}^d} u(x,t) e^{-2\pi i n \cdot x}\, dx.$$

Then $a_n \in C^2((0,\infty))$ and differentiation under the integral sign shows that $a_n'(t) = -4\pi^2 a_n(t)$, so that $a_n(t) = c_n e^{-4\pi^2 t}$ for some c_n. But $a_n(t) \to \widehat{f}(n)$ as $t \to 0$ uniformly in n by the convergence assumption, so $c_n = \widehat{f}(n)$. But this implies that $u(x,t) = (H_t * f)(x)$ for each $x \in \mathbf{T}^d$, since both sides have the same Fourier series for all $t > 0$. $\square$

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

$$f(x) = \sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{2\pi i k x}$$

is the Fourier inversion formula

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

where for each real number ξ, we define

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

The function $\widehat{f}$ is known as the *F*ourier 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. The fact that $\mathbf{R}$ is non-compact and has infinite measure adds some difficulty to the study of the Fourier transform over the Fourier series. For instance, since $L^p(\mathbf{R}^d)$ is not included in $L^q(\mathbf{R}^d)$ for $p \neq q$, which makes it more difficulty to perform a qualitative analysis of convergence in this setting. Nonetheless, the Fourier transform has many properties as the Fourier series. We add an additional difficulty by also analyzing the Fourier transform on $\mathbf{R}^d$, which, given $f : \mathbf{R}^d \to \mathbf{C}$, considers the quantities

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

for $\xi \in \mathbf{R}^d$. The basic theory of the Fourier transform in one dimension is essentially the same as the theory of the Fourier transform in d dimensions, though as d increases certain more technical considerations such as pointwise convergence become more difficult to understand.

4.1 Basic Calculations

In order to interpret the Fourier transform as an absolutely convergent integral, we require that we are dealing with integrable assumptions. Thus we analyze functions 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. However, one novel phenomenon in the basic theory is that the Fourier transform of an integrable function is continuous and vanishes at ∞.

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. $\square$

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

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

  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

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

  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 $(\mathrm{Trans}_y f)(x) = f(x - y)$. If $\xi \in \mathbf{R}^d$, then we define $(\mathrm{Mod}_\xi f)(x) = e^{2\pi i \xi \cdot x} f(x)$. We then find that

  $$\begin{aligned} \widehat{\mathrm{Trans}_y} f(\xi) &= \int f(x - y) e^{-2\pi i \xi \cdot x}\, dx \\ &= e^{-2\pi i \xi \cdot y} \int f(x) e^{-2\pi i \xi \cdot x}\, dx = (\mathrm{Mod}_{-y} \widehat{f})(\xi). \end{aligned}$$

and

$$\widehat{\text{Mod}_\xi f}(\eta) = \int e^{2\pi i \xi \cdot x} f(x) e(-\eta \cdot x)\, dx = \widehat{f}(\eta - \xi) = (\text{Trans}_\xi \widehat{f})(\eta).$$

Thus we conclude $\mathcal{F} \circ \text{Trans}_y = \text{Mod}_{-y} \circ \mathcal{F}$, and $\mathcal{F} \circ \text{Mod}_\xi = \text{Trans}_\xi \circ \mathcal{F}$.

- A very related property to the translational symmetry of the Fourier transform is related to the convolution

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

of two functions $f, g \in L^1(\mathbf{R}^d)$. 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.

Another related property to the translation symmetry is that the Fourier transform behaves well with respect to differentiation. If $f \in L^1(\mathbf{R}^d)$ has a weak derivative $D^\alpha f \in L^1(\mathbf{R}^d)$, then

$$\widehat{D^\alpha f}(\xi) = (2\pi i \xi)^\alpha \widehat{f}(\xi).$$

In particular, this is true if f is a *Schwartz function*, i.e. an element of

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

which is often a natural place to consider the Fourier transform. Conversely, if $f \in L^1(\mathbf{R}^d)$, and $x^\alpha f \in L^1(\mathbf{R}^d)$ for some multi-index α, then $\widehat{f}$ has a weak derivative $D^\alpha \widehat{f}$ in $L^1(\mathbf{R}^d)$, and

$$D^\alpha \widehat{f}(\xi) = (-\widehat{2\pi i x)^\alpha f}(\xi).$$

In particular, this means that the Fourier transform of a compactly supported element of $L^1(\mathbf{R}^d)$ lies in $C^\infty(\mathbf{R}^d)$, and all deriatives of the Fourier transform are integrable.

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

$$\begin{aligned}
\widehat{f \circ T}(\xi) &= \int f(Tx) e^{-2\pi i \xi \cdot x}\, dx \\
&= \frac{1}{|\det(T)|} \int f(y) e^{-2\pi i \xi \cdot T^{-1} y)}\, dy \\
&= \frac{1}{|\det(T)|} \int f(y) e^{-2\pi i T^{-T} \xi \cdot y}\, dy \\
&= \frac{1}{|\det(T)|} (\widehat{f} \circ T^{-T})(\xi).
\end{aligned}$$

Thus we conclude that if $T^* : L^1(\mathbf{R}^d) \to L^1(\mathbf{R}^d)$ is the operator defined by setting $T^*(f) = f \circ T$, then

$$\mathcal{F} \circ T^* = \frac{1}{|\det(T)|} \cdot (T^{-T})^* \circ \mathcal{F}.$$

This property indicates the 'cotangent' and symplectic properties of the Fourier transform. One way to think of this property is that one can think of the frequency variable as 'cotangent' to the spatial variable, since if we have a coordinate change $y = Tx$, and we define the 'coordinatized' Fourier transforms $\mathcal{F}_x$ and $\mathcal{F}_y$ by setting

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

and

$$\mathcal{F}_y f(\eta) = \int f(T^{-1} y) e^{-2\pi i y \cdot \eta}\, dy,$$

then the transformation formula tells us that

$$\mathcal{F}_y = \mathcal{F}_x \circ (T^{-1})^* = |\det(T)| \cdot (T^T)^* \circ \mathcal{F}_x,$$

i.e. $\mathcal{F}_y f(\eta) = |\det(T)| \cdot \mathcal{F}_x f(\xi)$, where $\eta = T^T \xi$. By interpreting ξ and η as *cotangent* vectors to the x and y coordinates respectively, one can therefore use this symmetry property to define a version of the Fourier transform that is invariant under area preserving changes of coordinates.

- As a special case of the theorem above, if $a \in \mathbf{R}$ and $\text{Dil}_a : L^1(\mathbf{R}^d) \to L^1(\mathbf{R}^d)$ is the operator defined by setting

$$(\text{Dil}_a f)(x) = f(x/a),$$

then

$$\widehat{\text{Dil}_a f} = a^d \cdot \text{Dil}_{1/a} \widehat{f}$$

As we increase a, the values of f are traced over more quickly, and so it is natural for the support of f to lie over larger frequencies. Note that this dilation preserves the L^∞ norm of f, but the Fourier transform preserves the L^1 norm. We could have alternatively consider the L^1 preserving dilation $f \mapsto a^{-d} f(x/a)$, which on the frequency side of things preserves the L^∞ norm. We can consider various magnitude adjustments; for instance, $f \mapsto a^{-d/2} f(x/a)$ preserves the L^2 norm in both space and frequency. But regardless, as we concentrate f in a smaller neighborhood, $\widehat{f}$ is dilated so it's support lies in a larger and larger neighborhood.

- Another special case is that 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. In particular, the space of radial functions is an *invariant subspace* with respect to the Fourier transform, which becomes important in more advanced theories, such as the theory of spherical harmonics.

4.2 The Fourier Algebra

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 $\mathbf{A}(\mathbf{R}^d) = C_0(\mathbf{R}^d)$. Current research cannot give a satisfactory description of the elements of $\mathbf{A}(\mathbf{R}^d)$, and a simple characterization is unlikely. 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

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

for all $\xi \in \mathbf{R}$, since

$$\lim_{b \to \infty} \int_1^b \frac{|\sin(2\pi\xi)|}{\xi \log|\xi|} = \infty. \qquad \square$$

On the other hand, for a finite measure μ on $\mathbf{R}^d$, we can define the Fourier transform to be the continuous function

$$\widehat{\mu}(\xi) = \int_{\mathbf{R}^d} e^{-2\pi i \xi \cdot x} d\mu(x).$$

In this case, the family of continuous functions which are the Fourier transforms of finite measures is precisely the family of $f \in C(\mathbf{R}^d)$ which are *positive definite*, in the sense that for each $x_1, \dots, x_N \in \mathbf{R}^d$ and $\xi_1, \dots, \xi_N \in \mathbf{C}$,

$$\sum_{i=1}^N \sum_{j=1}^N f(x_i - x_j)\xi_i\xi_j \geqslant 0.$$

The theorem, proved by Bochner, is best addressed in the more general case of harmonic analysis on locally compact abelian groups, and so we leave the proof of this for another time.

4.3 Basic Convergence Properties

As we might expect from the Fourier series theory, if $f \in C(\mathbf{R}) \cap L^1(\mathbf{R}^d)$ and $\widehat{f} \in L^1(\mathbf{R}^d)$, then the formula

$$f(x) = \int_{\mathbf{R}^d} \widehat{f}(\xi)e(\xi \cdot x)\, dx$$

holds for all $x \in \mathbf{R}^d$. Unlike in the case of the Fourier series, we cannot test our function against trigonometric polynomials. On the other hand, we have a multiplication formula which often comes in useful.

Theorem 4.4 (The Multiplication Formula)**.** *If $f, g \in L^1(\mathbf{R}^d)$,*

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

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

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

In particular, if $f \in L^1(\mathbf{R}^d)$ and $\widehat{f} = 0$, then for any $g \in L^1(\mathbf{R}^d)$,

$$\int_{\mathbf{R}^d} f(x)\widehat{g}(x)\,dx = 0.$$

If x_0 is a continuity point of f, then it suffices to choose a function g such that the majority of the mass of $\widehat{g}$ is concentrated at the point x_0. A natural choice here is to use a *Gaussian function*.

Let $g(x) = e^{-\pi x^2}$. Then $g \in L^1(\mathbf{R})$. Then $g'(x) = -2\pi x g(x)$, and since $xg \in L^1(\mathbf{R})$,w e conclude that

$$\frac{d\widehat{g}(\xi)}{d\xi} = -2\pi\xi\widehat{g}(\xi).$$

Since $\widehat{g}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2} = 1$, we conclude from solving the ordinary differential equation that

$$\widehat{g}(\xi) = e^{-\pi\xi^2} = g(\xi).$$

Tensorizing, it follows that if $g(x) = e^{-\pi|x|^2}$ is the element of $L^1(\mathbf{R}^d)$, then

$$\widehat{g}(\xi) = e^{-\pi|\xi|^2} = g(x).$$

In particular, if for $x_0 \in \mathbf{R}^d$ and $\delta > 0$, we define

$$g_{x_0,\delta}(\xi) = e^{-2\pi i x_0\cdot\xi)}g(\delta\xi)$$

then the symmetries of the Fourier transform imply that

$$\widehat{g_{x_0,\delta}}(\xi) = \delta^{-d}e^{-(\pi/\delta^2)|\xi - x_0|^2}.$$

Thus we conclude that if $\widehat{f} = 0$, then for any x_0 and δ,

$$\delta^{-d}\int_{\mathbf{R}^d} f(x)e^{-(\pi/\delta^2)|\xi - x_0|^2} = 0.$$

A simple approximation as $\delta \to 0$ gives the following result.

Theorem 4.5. *Suppose $f \in L^1(\mathbf{R}^d)$ and $\widehat{f} = 0$. Then f vanishes at any of it's continuity points. In particular, if $f \in L^1(\mathbf{R}^d) \cap C(\mathbf{R}^d)$ and $\widehat{f} = 0$, then $f = 0$.*

As in the case of the Fourier series, if $f \in L^1(\mathbf{R}^d)$ and $\widehat{f} \in L^1(\mathbf{R}^d)$, then the multiplication formula implies that for any $g \in L^1(\mathbf{R}^d)$ with $\widehat{g} \in L^1(\mathbf{R}^d)$,

$$\int_{\mathbf{R}^d} \widehat{\widehat{f}}(x) g(x)\, dx = \int_{\mathbf{R}^d} \widehat{f}(\xi)\widehat{g}(\xi)\, d\xi = \int_{\mathbf{R}^d} f(x)\widehat{\widehat{g}}(x)\, dx.$$

If $g = g_{x_0,\delta}$ for some x_0 and δ, then it is simple to calculate that $\widehat{\widehat{g}}(x) = g(-x)$. Thus we conclude that for any such function,

$$\int_{\mathbf{R}^d} \widehat{\widehat{f}}(x) g(x)\, dx = \int_{\mathbf{R}^d} f(-x) g(x)\, dx.$$

In particular, a similar approximation technique to the last theorem gives the Fourier inversion theorem.

Theorem 4.6. *Suppose $f \in L^1(\mathbf{R}^d)$ and $\widehat{f} \in L^1(\mathbf{R}^d)$. Then for any continuity point x of f,*

$$f(x) = \int_{\mathbf{R}^d} \widehat{f}(\xi) e^{2\pi i \xi \cdot x}\, dx.$$

In particular, if we also assume $f \in C(\mathbf{R}^d)$, then the inversion formula holds everywhere.

4.4 Alternative Summation Methods

As with the Fourier series, we can obtain results for more general functions by 'dampening' the integration factor. To do this, we consider 'alternate integral' methods which can define the integral of a measurable function that is not necessarily absolutely integrable.

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

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

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

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

Thus $A_\delta f$ represents the Abel sums of the Fourier inversion formula.

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

$$\int_{\mathbf{R}^d} f(x) e^{-\delta|x|}\, dx \to \int_{\mathbf{R}^d} f(x)\, dx.$$

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

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

Then as $\delta \to 0$, $G_\delta f$ represents the Gauss sums of the Fourier inversion formula.

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

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

which are analogous to the Fejér sums in the periodic setting.

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

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

These integrals can be written as the integral of $f\chi_{[-R,R]}$, and so in a generalized sense, we can integrate a function f if $f\chi_{[-R,R]}$ 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 4.7. *Suppose $f \in L^1_{loc}(\mathbf{R})$, that*

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

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

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

Proof. Let

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

For each $x \geqslant 0$, write

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

Then F is continuous and differentiable almost everywhere, 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 $s > 0$,

$$\begin{aligned}
\int_{-s}^{s} f(x) e^{-\delta x^2}\, dx &= \int_{0}^{s} [f(x) + f(-x)] e^{-\delta x^2}\, dx \\
&= F(s) e^{-\delta s^2} + 2\delta \int_0^s x F(x) e^{-\delta x^2}\, dx.
\end{aligned}$$

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

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

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

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

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

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

So the third term is equal to $|A||1 - e^{-\delta t^2}|$ and so 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 $\varepsilon \to 0$. $\square$

Abel summation is even more general than Gauss summation.

Lemma 4.8. *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 $\Phi \in C_0(\mathbf{R}^d)$ 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 4.9 (The Multiplication Formula)**.** *If $f, g \in L^1(\mathbf{R}^d)$,*

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

Proof. If $f, g \in L^1(\mathbf{R}^d)$, then $\widehat{f}$ and $\widehat{g}$ are bounded, continuous functions on $\mathbf{R}^d$. 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

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

Thus if we define $K_\delta^\Phi(x) = \delta^{-d}\widehat{\Phi}(-x/\delta)$, 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 $\{\|K_\delta\|_{L^1(\mathbf{R}^d)}\}$ 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 $f \in L^p(\mathbf{R}^d)$, 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 $\xi \in \mathbf{R}^d$,
$$|K_\delta(\xi)| \lesssim \frac{\delta}{|\xi|^{d+1}}.$$

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. For a particular function Φ, the family $\{K_\delta^\Phi\}$ forms a good kernel as $\delta \to 0$ if $\widehat{\Phi} \in L^1(\mathbf{R}^d)$ and

$\int \widehat{\Phi}(\xi)\, d\xi = 1$, and forms an approximation to the identity if we assume in addition that $\Phi \in C^{d+1}(\mathbf{R}^d)$. Thus we conclude that as $\delta \to 0$, if Φ satisfies the appropriate conditions then as $\delta \to 0$, the summations $S_\delta(f, \Phi)$ converge to f in the appropriate sense as considered above.

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

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

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 $\mathbf{R}^d$ 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 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

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

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

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

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

Thus the Abel mean is the Fourier inverse of the Poisson kernel on the upper half plane $\mathbf{H}^{d+1}$. We note that the Poisson summation formula shows that for $d = 1$, the Poisson kernel on $\mathbf{T}$ is the periodization of the Poisson kernel on $\mathbf{R}$.

In order to conclude $\{P_\delta\}$ is a good kernel, it now suffices to verify that

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

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

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

Setting $r = \tan u$, we find

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

But we can now show by induction that

$$\frac{S_d}{2} = S_{d-1} \int_0^{\pi/2} (\sin u)^{d-1}\, du.$$

Using the values $S_0 = 2$, $S_1 = 2\pi$, and $S_2 = 4\pi$, the theorem certainly holds for $d = 1$ and $d = 2$. For $d > 2$, integration by parts and induction shows that

$$\begin{aligned} S_{d-1} \int_0^{\pi/2} (\sin u)^{d-1}\, du &= S_{d-1} \frac{d-2}{d-1} \int_0^{\pi/2} (\sin u)^{d-3}(t)\, dt. \\ &= \frac{d-2}{d-1} \frac{S_{d-1} S_{d-2}}{2 S_{d-3}} \\ &= \frac{d-2}{d-1} \frac{\pi^{d/2} \pi^{d/2-1/2}}{\pi^{d/2-1}} \frac{\Gamma(d/2-1)}{\Gamma(d/2)\Gamma(d/2-1/2)} \\ &= \frac{\pi^{d/2+1/2}}{\Gamma(d/2+1/2)} = \frac{S_d}{2}. \end{aligned}$$

Thus our theorem is complete.

Example. *We note that*

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

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,

$$f(x) = \lim_{\delta \to 0} \int \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \Phi(\delta x).$$

If $\widehat{f}$ is integrable, then the bound $|\widehat{f}(\xi) e^{2\pi i \xi \cdot x} \Phi(\delta \xi)| \leqslant \|\Phi\|_\infty |\widehat{f}(\xi)|$ 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 4.10. *If* f *and* $\widehat{f}$ *are elements of* $L^1(\mathbf{R}^d)$*, then for any* x *in the Lebesgue set of* f*,*

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

Remark. We note that if $f \in L^1(\mathbf{R}^d)$, $\widehat{f} \geqslant 0$, and f is continuous at the origin, then the Fourier inversion formula and the monotone convergence theorem implies that

$$f(0) = \lim_{\delta \to 0} \int_{\mathbf{R}^d} \widehat{f}(\xi) e^{-\delta \xi}\, d\xi = \int_{\mathbf{R}^d} \widehat{f}(\xi)\, d\xi.$$

Thus $\widehat{f}$ is integrable, and so the Fourier inversion theorem holds.

As a particular example of this remark, if $f \in L^1(\mathbf{R}^d)$ then we can define the autocorrelation function

$$R(x) = \int_{\mathbf{R}^d} f(y+x) f(y)\, dy,$$

then $R \in L^1(\mathbf{R}^d)$ and $\widehat{R}(\xi) = |\widehat{f}(\xi)|^2$. Thus R is continuous at the origin if and only if $\widehat{R}$ is integrable, which, using the L^2 theory we develop in the next section, holds if and only if $f \in L^2(\mathbf{R}^d)$.

It is often useful to note that if the Fourier transform of an integrable function is non-negative, then it's Fourier transform is automatically integrable.

Theorem 4.11. *If $f \in L^1(\mathbf{R})$ is continuous at the origin, and $\widehat{f} \geqslant 0$, then $\widehat{f}$ is integrable.*

Proof. This follows because

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

By Fatou's lemma,

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

so $\widehat{f}$ is finitely integrable. ∎

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

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 weak 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 4.12. *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

$$\int_{|x| \geqslant R} \widehat{f}(\xi) e(\xi \cdot x) \leqslant \int_{|x| \geqslant R} |\widehat{f}(\xi)| \, d\xi = o(1).$$

so the pointwise convergence is uniform in x. □

Remark. This theorem also generalizes to $\mathbf{R}^d$. 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 4.13. *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. □

The corollary above is often 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 4.14. *For any δ_1, δ_2,*

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

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}. □

Many of the other convergence statements for Fourier series hold in the case of the Fourier transform. For instance, a non-periodic variant of the De la Vallee Poisson kernel shows that if $f \in L^1(\mathbf{R})$ and $\widehat{f}(\xi) = O(1/|\xi|)$, then $S_R f$ converges uniformly to f. But for the purpose of novelty, we move on to other concepts.

4.5 The L^2 Theory

There are various differences in the L^2 for the Fourier transform vs the case of Fourier series. In the compact, periodic case, $L^2(\mathbf{T}^d)$ is contained in $L^1(\mathbf{T}^d)$ and can thus be viewed as a *more regular* family of functions than the square integrable functions. In the noncompact case, $L^2(\mathbf{R}^d)$ is not contained in $L^1(\mathbf{R}^d)$, and thus is *more regular* in some respects (we have more control over singularities), but we have less control on how spread out the function is. In particular, we often have to rely on density arguments, working in the space $L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, which is a dense subspace of $L^2(\mathbf{R}^d)$.

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

$$\sum_{n \in \mathbf{Z}^d} |\widehat{f}(n)|^2 = \int_{\mathbf{T}^d} |f(x)|^2 \, dx$$

Let us try and extend this to $\mathbf{R}^d$. A natural formula is to expect that

$$\int_{\mathbf{R}^d} |\widehat{f}(\xi)|^2 \, d\xi = \int_{\mathbf{R}^d} |f(x)|^2 \, dx.$$

In order to interpret the right hand side as a finite quantity, we must assume $f \in L^2(\mathbf{R}^d)$, and to interpret the left hand side, we must assume $f \in L^1(\mathbf{R}^d)$. A result of our calculation will show that under these assumptions, $\widehat{f} \in L^2(\mathbf{R}^d)$, and that the formula holds.

Theorem 4.15. *If $f \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, 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

$$\int_{\mathbf{R}^d} |\widehat{f}(\xi)|^2 \, d\xi = \int_{\mathbf{R}^d} \widehat{f}(\xi)\overline{\widehat{f}(\xi)} \, d\xi = \int_{\mathbf{R}^d} f(x)\overline{f(x)} \, dx = \int_{\mathbf{R}^d} |f(x)|^2 \, dx. \quad \square$$

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

Corollary 4.16. *If $f \in L^1(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$ for $1 \leqslant p \leqslant 2$, then*

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

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

Though the integral formula of an element of $L^2(\mathbf{R}^d)$ does not make sense, the bounds above provide a canonical way to define the Fourier transform of an element of $L^p(\mathbf{R}^d)$, for $1 \leqslant p \leqslant 2$. The space $L^1(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$ is a dense subset of $L^p(\mathbf{R}^d)$, so we can use the Hahn-Banach theorem to define the Fourier transform $\mathcal{F} : L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ as the *unique* bounded operator agreeing with the integral formula on the common domain. A more explicit way to define the Fourier transform is as the L^2 limit of the bounded Fourier transform operators; for each $f \in L^2(\mathbf{R}^d)$, and $R > 0$, $f\mathbf{I}_{B_R} \in L^1(\mathbf{R}^d)$, where B_R is the ball of radius R about the origin. It follows that if we define

$$\mathcal{F}_R f(\xi) = \int_{|x| \leqslant R} f(x)e^{-2\pi i \xi \cdot x}.$$

then since $\lim_{R\to\infty} \|\mathbf{I}_{B_R} f - f\|_{L^2(\mathbf{R}^d)} = 0$, the L^2 continuity of the Fourier transform implies that $\mathcal{F}_R f$ converges to $\widehat{f}$ in $L^2(\mathbf{R}^d)$.

The main way to obtain results about the Fourier transform of square integrable functions is by a density argument. For instance, suppose we wish to prove that for $f, g \in L^2(\mathbf{R}^d)$,

$$\int_{\mathbf{R}^d} \widehat{f}(\xi)\widehat{g}(\xi)\, d\xi = \int_{\mathbf{R}^d} f(x)g(x)\, dx.$$

This equality certainly holds by the multiplication formula if $f, \widehat{g} \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$. We also find that both sides are continuous as bilinear functionals, by applying the Cauchy-Schwartz inequality and the isometry of the Fourier transform. Since any element f of $L^2(\mathbf{R}^d)$ can be approximated in the L^2 norm by an element of $L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, and since any element g of $L^2(\mathbf{R}^d)$ can be approximated by functions with $\widehat{g} \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, the theorem holds in general. In particular, this shows that the extension of the Fourier transform to $L^2(\mathbf{R}^d)$ remains unitary.

Another approximation argument can be used to obtain convergence results in $L^2(\mathbf{R}^d)$ for the Fourier transform. If we let

$$S_\delta(f,\Phi) = \int \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \Phi(\delta\xi)\, d\xi$$

which is well defined for a particular function $\Phi \in L^2(\mathbf{R}^d)$, then a density argument again shows that $S_\delta(f,\Phi) = K_\delta^\Phi * f$, where K_δ^Φ is defined as in the last section. Provided that we also have $\Phi \in L^1(\mathbf{R}^d)$ and $\int \widehat{\Phi}(\xi)\, d\xi = 1$, then we conclude that $S_\delta(f,\Phi) \to f$ in $L^2(\mathbf{R}^d)$, and that if $\Phi \in C^{d+1}(\mathbf{R}^d)$, then $S_\delta(f,\Phi) \to f$ almost everywhere. In particular, one can use the Gauss, Abel, and Fejer sums here to get L^2 convergence.

Unlike in the case of Fourier series, where the L^2 theory gives an isometry between $L^2(\mathbf{T}^d)$ and $L^2(\mathbf{Z}^d)$, in the case of the Fourier transform the Fourier transform gives a unitary operator from $L^2(\mathbf{R}^d)$ to itself, and thus we can consider the spectral theory of such an operator. The Fourier inversion formula implies that the Fourier transform has order four. Thus the only eigenfunctions of the Fourier transform correspond to eigenvalues in $\{1,-1,i,-i\}$. We have seen $e^{-\pi x^2}$ is an eigenfunction with eigenvalue one. If we consider the family of all *Hermite polynomials*

$$H_n(x) = \frac{(-1)^n}{n!} e^{\pi x^2} \frac{d^n}{dx^n}\left(e^{-\pi x^2}\right).$$

One can also see that

$$\sum_{n=0}^{\infty} (-t)^n/n!$$

$$\sum_{n=0}^{\infty} H_n(x) \frac{t^n}{n!} = e^{-\pi x^2 - (2\pi)^{1/2} tx + t^2}$$

TODO PROVE ORTHOGONALITY AND COMPLETENESS. which satisfy $\widehat{H_n} = (-i)^n H_n$, then we obtain an orthonormal basis of eigenfunctions. In higher dimensions, a basis of eigenfunctions for $L^2(\mathbf{R}^d)$ is given by taking tensor products of Hermite polynomials.

4.6 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 $f \in L^1(\mathbf{R})$, let $f_r(x) = f(rx)$. Then we find $\widehat{f_r}(\xi) = r^{-d} \widehat{f}(\xi/r)$, and so*

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

In order for a bound to hold in terms of p and q uniformly for all values of r, we need $r^{-d/p} = r^{d/q-d}$, which means $1/q + 1/p = 1$, so p and q must be conjugates of one another. In the case of $\mathbf{T}$, a function analogous to f_r can only be defined for small value of r, and a uniform estimate can then only hold if $1/p + 1/q \geqslant 1$.

If $p > 2$, then for the value q with $1/p + 1/q = 1$, we have $q < p$. It is a principle of Littlewood that translation invariant operators cannot satisfy a L^p to L^q bound.

Theorem 4.17. *Suppose $T : L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ is a translation invariant continuous operator, where $q < p$. Then $T = 0$.*

Proof. If T was nonzero, we could pick some $f_0 \in L^p(\mathbf{R}^d)$ such that $T f_0 \neq 0$. Rescaling T and f_0, we may assume without loss of generality that

$\|f_0\|_{L^p(\mathbf{R}^d)} = \|Tf_0\|_{L^q(\mathbf{R}^d)} = 1$. Furthermore, by truncation we may assume that f_0 has compact support on some ball B_R. But then the supports of the functions $\mathrm{Trans}_{2Rn} f_0$ and f_0 are disjoint for $n \in \mathbf{Z}^d$, so for any choice of coefficients $\{a_n\}$,

$$\left\| \sum_{n \in \mathbf{Z}^d} a_n \cdot \mathrm{Trans}_{2Rn} f_0 \right\|_{L^p(\mathbf{R}^d)} = \left(\sum |a_n|^p \right)^{1/p}.$$

Assume that at most N of the coefficients a_n are nonzero. We cannot necessarily assume that Tf_0 has compact support but the majority of the mass of Tf_0 can still be concentrated on a compact set. For any $\varepsilon > 0$ we can choose R large enough that

$$\left(\int_{|x| \geqslant R} |Tf_0(x)|^q \right)^{1/q} \leqslant \varepsilon.$$

Now for each B_R and $m \in \mathbf{Z}^d$,

$$\left(\int_{x \in 2Rm + B_R} \left| \sum_{n \in \mathbf{Z}^d} a_n \mathrm{Trans}_{2Rn} Tf_0(x) \right|^q dx \right)^{1/q} \geqslant \left(|a_m|^q - \varepsilon \sum_{n \neq m} |a_n|^q \right)^{1/q}.$$

If, for a *fixed* sequence $\{a_n\}$, we choose

$$\varepsilon \leqslant \frac{0.5}{\max_{n \in \mathbf{Z}^d} |a_n| \cdot \left(\sum_{n \in \mathbf{Z}^d} |a_n|^q \right)^{1/q}}.$$

Then we find

$$\left(\int_{x \in 2Rm + B_R} \left| \sum_{n \in \mathbf{Z}^d} a_n \mathrm{Trans}_{2Rn} Tf_0(x) \right|^q dx \right)^{1/q} \geqslant 0.5^{1/q} |a_m|$$

and so summing over all m, we conclude that

$$\| \sum_{n \in \mathbf{Z}^d} a_n \mathrm{Trans}_{2Rn} Tf_0 \|_{L^q(\mathbf{R}^d)} \geqslant 0.5^{1/q} \left(\sum_{n \in \mathbf{Z}^d} |a_n|^q \right)^{1/q}.$$

Thus we conclude that for *any* sequence $\{a_n\}$ in $l^q(\mathbf{Z}^d)$,

$$\left(\sum_{n \in \mathbf{Z}^d} |a_n|^q \right)^{1/q} \lesssim_q \left(\sum_{n \in \mathbf{Z}^d} |a_n|^p \right)^{1/p}.$$

where the constant is independent of the sequence. For $q < p$ this is impossible. $\square$

We can also provide a family of functions whose Fourier transforms contradict an extension of the Hausdorff Young inequality for $p > 2$.

Example. *Consider the family of functions $f_s(x) = s^{-d/2} e^{-\pi |x|^2/s}$, where $s = 1 + it$ for some $t \in \mathbf{R}$. One can easily calcluate using analytic continuation and the Fourier transform for the Gaussian that $\widehat{f_s}(\xi) = e^{-\pi s |\xi|^2}$. We calculate*

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

whereas $\|\widehat{f_s}\|_q = q^{-d/2}$. Thus to be able compare the two quantities as $t \to \infty$, we need $d/p - d/2 \leqslant 0$, so $p \leqslant 2$. As $t \to \infty$, $|f_s(x)| \sim t^{-d/2} e^{-\pi |x/t|^2}$, 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. One can obtain analogous examples in $\mathbf{T}^d$ to this example, by applying the Poisson summation formula to the functions f_s and noting that the L^p and L^q norms also follows approximately the same formulas as above.

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.

4.7 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 is a function on $\mathbf{R}$, 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, under an appropriate regularity condition so that we can interpret these formulas correctly, they give the same function.

Theorem 4.18. *Suppose $f \in L^1(\mathbf{R}^d)$. Then the series*

$$\sum_{n \in \mathbf{Z}^d} \mathit{Trans}_n f$$

converges absolutely in $L^1[0,1]^d$ to a function $g \in L^1[0,1]^d$ with $\widehat{g}(n) = \widehat{f}(n)$ for each $n \in \mathbf{Z}^d$.

Proof. The fact that the sum converges absolutely in $L^1[0,1]$ follows because

$$\sum_{n \in \mathbf{Z}^d} \|\text{Trans}_n f\|_{L^1[0,1]} = \|f\|_{L^1(\mathbf{R}^d)}.$$

But the absolute convergence in L^1 also justifies the calculation that for each $n \in \mathbf{Z}^d$

$$\begin{aligned} \int_{[0,1]^d} \sum_{m \in \mathbf{Z}^d} (\text{Trans}_n f)(x) e^{2\pi n i x}\, dx &= \sum_{m \in \mathbf{Z}^d} \int_{[0,1]^d} f(x+m) e^{2\pi n i (x+m)}\, dx \\ &= \int_{\mathbf{R}^d} f(x) e^{2\pi n i x}\, dx = \widehat{f}(n). \end{aligned}$$

$\square$

We can obtain a much more powerful version of this result if we assume that there is $\delta > 0$ such that

$$|f(x)| \lesssim \frac{1}{1+|x|^{d+\delta}} \quad \text{and} \quad |\widehat{f}(\xi)| \lesssim \frac{1}{1+|x|^{d+\delta}}.$$

Then we see that the two functions

$$g_1(x) = \sum_{n\in\mathbf{Z}^d} f(x+n) \quad \text{and} \quad g_2(x) = \sum_{n\in\mathbf{Z}^d} \widehat{f}(n) e^{2\pi i n\cdot x}$$

are continuous functions on $\mathbf{T}^d$ with the same Fourier coefficients. It thus follows that $g_1 = g_2$, i.e. that for each $x \in \mathbf{R}^d$,

$$\sum_{n\in\mathbf{Z}} f(x+n) = \sum_{n\in\mathbf{Z}} \widehat{f}(n) e^{2\pi n i x}.$$

In particular, this holds if $f \in \mathcal{S}(\mathbf{R})$.

TODO: Also prove this statement under the assumption that f has bounded variation and $f(t) = [f(t+) + f(t-)]/2$ for all $t \in \mathbf{R}$.

4.8 Convergence of Fourier Series via Poisson Summation

TODO

4.9 Radial Functions

Suppose $f \in L^1(\mathbf{R}^d)$ is a radial function. Then $\widehat{f}$ is also a radial function. In particular, if we let

$$\|u\|_{L^1([0,\infty),r^{d-1})} = \int_0^\infty r^{d-1} u(r)\, dr$$

then we have a transform $u \mapsto \tilde{u}$ from $L^1([0,\infty), r^{d-1})$ to $L^\infty[0,\infty)$ where if $f(x) = u(|x|)$, then $\widehat{f}(\xi) = \tilde{u}(|\xi|)$. In particular, we calculate quite simply that

$$\tilde{u}(s) = V_d \int_0^\infty r^{d-1} u(r) \left(\int_{S^{d-1}} e^{-2\pi i x_1 s}\, dx \right).$$

If one recalls the Bessel functions $\{J_s\}$, then we have

$$\tilde{u}(s) = 2\pi s^{1-d/2} \int_0^\infty r^{d/2} u(r) J_{d/2-1}(2\pi s r)\, dr.$$

If one recalls some Bessel function asymptotics, then one can actually gain some interesting results for the *averaging operator*

$$Af(x) = \fint_{S^{d-1}} f(x-y)\, d\sigma(y)$$

Example. *Suppose* $f_R(x) = \mathbf{I}_{|x| \leqslant R}$. *Then*

$$\widehat{f}(\xi) = 2\pi |\xi|^{1-d/2} \int_0^R r^{d/2} J_{d/2-1}(2\pi s r)\, dr.$$

The

4.10 Poisson Integrals

s

Chapter 5

Applications of the Fourier Transform

5.1 Applications to Partial Differential Equations

Just as the Fourier series can be used to obtain periodic solutions to certain partial differential equations, the Fourier transform can be used to obtain more general solutions to partial differential equations on $\mathbf{R}^d$. To begin with, we study the heat equation on $\mathbf{R}^d$, i.e. we study solutions to the partial differential equation

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

Formally taking Fourier transforms in the spatial variable gives

$$\frac{\partial \widehat{u}(\xi, t)}{\partial t} = -4\pi^2 |\xi|^2 \widehat{u}(\xi, t)$$

which, if we are given $u(x,0) = f(x)$, gives that

$$\widehat{u}(\xi, t) = \widehat{f}(\xi) e^{-4\pi^2 |\xi|^2 t}.$$

Thus, taking the inverse Fourier transform, we might expect the solution to the heat equation to be given by the formula

$$u(x,t) = (H_t * f)(x)$$

where

$$H_t(x) = \frac{1}{(4\pi t)^{d/2}} e^{-|x|^2/4t}.$$

The rapid decay of H_t for large x shows that for any $1 \leqslant p \leqslant \infty$ and $f \in L^p(\mathbf{R}^d)$, u is well defined by this formula, lies in $C^\infty(\mathbf{T}^d)$, and solves the heat equation, with the appropriate norm convergence as $t \to 0$. However, in this case it is not so easy to conclude that u is the unique solution to this equation satisfying the initial conditions, since one cannot necessarily take the Fourier transform of u.

We can get slightly more results if we consider the *steady state* heat equation on the upper half plane $\mathbf{H}^d$, i.e. we study functions $u(x,t)$, for $x \in \mathbf{R}^d$ and $t > 0$, such that $\Delta u = 0$, subject to the initial condition that $u(x,0) = f(x)$. Working formally with the Fourier transform leads to the equation

$$\widehat{u}(\xi,t) = e^{-2\pi t|\xi|x}\widehat{f}(\xi)$$

Thus $u(x,t) = (f * P_t)(x)$, where P_t is the Poisson kernel. If $f \in L^1(\mathbf{R}^d)$, it is easy to see that

5.2 Shannon-Nyquist Sampling Theorem

Often, in applications, one deals with band limited function, i.e. functions whose Fourier transforms are compactly supported. For simplicity, we work solely with functions f on $\mathbf{R}$ satisfying a decay condition

$$|f(t)| \lesssim \frac{1}{(1+|t|)^{1+\delta}}.$$

It follows that $f \in L^p(\mathbf{R}^d)$ for each $1 \leqslant p \leqslant \infty$. Suppose that in addition, $\widehat{f}$ is supported on $[-1/2, 1/2]$. It follows that, $\widehat{f} \in L^p(\mathbf{R}^d)$ for each $1 \leqslant p \leqslant \infty$. In particular, it follows that f is smooth, if we alter it on a set of measure zero. Now taking Fourier series on $[-1/2, 1/2]$, noting that $f \in L^1(\mathbf{Z})$ because of it's decay, we find that for each $\xi \in \mathbf{R}$,

$$\widehat{f}(\xi) = \mathbf{I}(|\xi| \leqslant 1/2) \sum_{n=-\infty}^{\infty} f(n) e^{-2\pi i n \xi}.$$

But now we conclude by the Fourier inversion formula that

$$\begin{aligned} f(x) &= \int_{-1/2}^{1/2} \left(\sum_{n=-\infty}^{\infty} f(n) e^{-2\pi i n \xi} \right) e^{2\pi i \xi x}\, d\xi \\ &= \sum_{n=-\infty}^{\infty} f(n) \int_{-1/2}^{1/2} e^{2\pi i \xi (x-n)}\, d\xi \\ &= \sum_{n=-\infty}^{\infty} f(n) \cdot \frac{\sin(\pi(x-n))}{\pi(x-n)}. \end{aligned}$$

In particular, we conclude that the function f is uniquely determined by sampling it's values over the integers. In particular, if N is large, and $|x| \leqslant N/2$

$$\left| f(x) - \sum_{n=-N}^{N} f(n) \cdot \frac{\sin(\pi(x-n))}{\pi(x-n)} \right| \lesssim \frac{1}{N},$$

where the implicit constant depends on the decay of f. If we sample on a more fine set of values, then we obtain faster convergence. To do this, we instead take the Fourier series of $\widehat{f}$ on $[-\lambda/2, \lambda/2]$, noting that

$$\frac{1}{\lambda} \int_{-\lambda/2}^{\lambda/2} \widehat{f}(\xi) e^{2\pi i n \xi/\lambda}\, d\xi = \frac{f(n/\lambda)}{\lambda}$$

so that

$$\widehat{f}(\xi) = \chi(\xi) \sum_{n=-\infty}^{\infty} \frac{f(n/\lambda)}{\lambda} e^{-2\pi n \xi/\lambda}.$$

where instead of being the indicator on $[-1/2, 1/2]$, χ is the piecewise linear function equal to 1 for $|\xi| \leqslant 1/2$, and vanishing for $|\xi| \geqslant \lambda/2$. One can calculate quite easily that

$$\widehat{\chi}(x) = \frac{\cos(\pi x) - \cos(\lambda \pi x)}{\pi^2 (\lambda - 1) x^2}.$$

Thus it follows from the Fourier inversion formula that

$$\begin{aligned} f(x) &= \sum_{n=-\infty}^{\infty} \frac{f(n/\lambda)}{\lambda} \int_{-\infty}^{\infty} \chi(\xi) e^{2\pi i \xi (x - n/\lambda)}\, dx \\ &= \sum_{n=-\infty}^{\infty} \frac{f(n/\lambda)}{\lambda} \widehat{\chi}(n/\lambda - x) \\ &= \sum_{n=-\infty}^{\infty} f(n/\lambda) \frac{\cos(\pi(n/\lambda - x)) - \cos(\lambda\pi(n/\lambda - x))}{\pi^2 \lambda(\lambda-1)(n/\lambda - x)^2}. \end{aligned}$$

It follows that if $|x| \leqslant N/2\lambda$, then

$$\left| f(x) - \sum_{n=-N}^{N} f(n/\lambda) \frac{\cos(\pi(n/\lambda - x)) - \cos(\lambda\pi(n/\lambda - x))}{\pi^2 \lambda(\lambda-1)(n/\lambda - x)^2} \right| \lesssim \left(1 + \frac{1}{\lambda - 1} \right) \frac{1}{N^2}.$$

Thus the rate of convergence of this sum is much better if we *oversample* by a large value Λ.

We should not expect f to be obtainable exactly if we undersample, i.e. look at the coefficients $\{f(n/\lambda) : n \in \mathbf{Z}\}$ for some $\lambda < 1$. Thus undersampling often yields artifacts in our reconstruction. For instance, when one takes a video of periodic motion travelling at a much greater frequency than the framerate of a video. To see why this is true, we consider a distributional formulation of the Nyquist sampling theorem.

Theorem 5.1. *For any $\lambda < 1$, there exists $f_1, f_2 \in \mathcal{S}(\mathbf{R})$, with $\widehat{f_1}$ and $\widehat{f_2}$ supported on $[-1/2, 1/2]$, such that $f_1(n/\lambda) = f_2(n/\lambda)$ for any $n \in \mathbf{Z}$.*

Proof. Fix $f_0 \in \mathcal{S}(\mathbf{R})$. Then the Poisson summation formula, appropriately rescaled, tells us that for each $\xi \in \mathbf{R}$,

$$\sum_{n=-\infty}^{\infty} f(n/\lambda) e^{-2\pi n i \xi} = \lambda^d \sum_{n=-\infty}^{\infty} \widehat{f}(\xi - \lambda n).$$

One can determine all the coefficients $\{f(n/\lambda)\}$ if one knows the right hand side for all values $\xi \in \mathbf{R}$. Thus if $f_1, f_2 \in \mathcal{S}(\mathbf{R})$ are distinct functions such that $\widehat{f_1}$ and $\widehat{f_2}$ are supported on $[-1/2, 1/2]$, but are equal to one another at a periodization of scale λ, then $f_1(n/\lambda) = f_2(n/\lambda)$ for any $n \in \mathbf{Z}$. This is certainly possible if $\lambda < 1$. □

We can also get a discretized L^2 identity.

Theorem 5.2. *Suppose $f \in L^2(\mathbf{R})$ and $\widehat{f}$ is supported on $[-1/2, 1/2]$. Then*

$$\sum_{n=-\infty}^{\infty} |f(n)|^2 = \int_{-\infty}^{\infty} |f(x)|^2 \, dx.$$

Proof. Poisson summation applied to $|f(x)|^2$ implies that

$$\sum_{n=-\infty}^{\infty} |f(n)|^2 = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \widehat{f}(\xi)\overline{\widehat{f}(\xi - n)} \, dx = \int_{-1/2}^{1/2} |\widehat{f}(\xi)|^2 \, dx = \int_{-\infty}^{\infty} |f(x)|^2 \, dx.$$

$\square$

5.3 The Uncertainty Principle

The uncertainty principle gives a constraint preventing both a function and it's Fourier transform from concentrating too tightly in a particular region. In particular, if the mass of a function is concentrated in a region of radius L, it is impossible for the mass of the Fourier transform to be concentrated in a region of radius $1/L$. The most fundamental version of the uncertainty principle is due to Heisenberg.

Theorem 5.3 (Heisenberg)**.** *Suppose $\psi \in \mathcal{S}(\mathbf{R})$, and that*

$$\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1.$$

Then for any $x_0, \xi_0 \in \mathbf{R}$,

$$\left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx \right) \left(\int_{-\infty}^{\infty} \xi^2 |\widehat{\psi}(\xi)|^2 \, d\xi \right) \geqslant \frac{1}{16\pi^2} \left(\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx \right)^2.$$

Proof. Normalizing, we may assume that $\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1$ and that $x_0, \xi_0 = 0$. A density argument enables us to assume that $\psi \in \mathcal{S}(\mathbf{R})$. Integration by parts shows that

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} |\psi(x)|^2 \, dx \\ &= -\int_{-\infty}^{\infty} x \frac{d|\psi(x)|^2}{dx} = -\int_{-\infty}^{\infty} (x\psi'(x)\overline{\psi(x)} + x\overline{\psi'(x)}\psi(x)) \, dx. \end{aligned}$$

Thus

$$\begin{aligned}
1 &\leqslant 2 \int_{-\infty}^{\infty} |x||\psi(x)||\psi'(x)|\, dx \\
&\leqslant 2 \left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2\, dx \right)^{1/2} \left(\int_{-\infty}^{\infty} |\psi'(x)|^2\, dx \right)^{1/2} \\
&\leqslant 4\pi \left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2\, dx \right)^{1/2} \left(\int_{-\infty}^{\infty} \xi^2 |\widehat{\psi}(\xi)|^2\, d\xi \right)^{1/2}. \qquad \square
\end{aligned}$$

Taking ψ to be a Gaussian shows the constant in this inequality is tight. Let us explain the applications of this uncertainty principle in quantum mechanics. Here the position state of a particle is no longer given by a particular point, but instead given by a state function ψ subject to the normalization condition

$$\int_{-\infty}^{\infty} |\psi(x)|^2\, dx = 1.$$

It then follows that the position of a particle is nondeterministic, with $|\psi(x)|^2$ giving the probability density function of where the particle is located. If x_0 denotes the expected value of the particle, then the variance of the distribution is given by

$$\int_{-\infty}^{\infty} |x - x_0|^2 |\psi(x)|^2\, dx.$$

On the other hand, the *momentum* of the particule is also nonrandom, and given by $|\widehat{\psi}(\xi)|^2$. Thus the variance of the momentum, if ξ_0 is the expectation, is equal to

$$\int_{-\infty}^{\infty} |\xi - \xi_0|^2 |\widehat{\psi}(\xi)|^2\, d\xi.$$

If we view the variance as a measurement of the uncertainty of each quantity, and denote each variance by Δ_x and Δ_ξ, then the Heisenberg uncertainty principle tells us that $\Delta_x \cdot \Delta_\xi \geqslant 1/16\pi^2$ (actually we have lied by non introducing physical constants into the discussion - we actually have $\Delta_x \cdot \Delta_\xi \geqslant \hbar/16\pi^2$, where $\hbar$ is Planck's constant).

We can also rephrase the uncertainty principle in terms of the differential operator

$$L = x^2 - \frac{d^2}{dx^2}.$$

This operator is known as the *Hermite operator*. Then for any $f \in \mathcal{S}(\mathbf{R})$,

$$\begin{aligned}
(Lf, f) &= \int_{-\infty}^{\infty} x^2 |f(x)|^2 - f''(x)\overline{f(x)}\, dx \\
&= \int_{-\infty}^{\infty} x^2 |f(x)|^2\, dx + \int_{-\infty}^{\infty} |f'(x)|^2\, dx \\
&= \int_{-\infty}^{\infty} x^2 |f(x)|^2\, dx + 4\pi^2 \int_{-\infty}^{\infty} \xi^2 |\widehat{f}(\xi)|^2\, d\xi.
\end{aligned}$$

The Heisenberg uncertainty principle thus implies by Young's inequality that

$$\begin{aligned}
(f, f) &\leqslant 4\pi \left(\int_{-\infty}^{\infty} x^2 |f(x)|^2\, dx \right)^{1/2} \left(\int_{-\infty}^{\infty} \xi^2 |\widehat{f}(\xi)|^2\, d\xi \right)^{1/2} \\
&\leqslant \int_{-\infty}^{\infty} x^2 |f(x)|^2\, dx + 4\pi^2 \int_{-\infty}^{\infty} \xi^2 |\widehat{f}(\xi)|^2\, d\xi = (Lf, f).
\end{aligned}$$

Thus the operator $f \mapsto Lf - f$ is positive definite. If we consider the operator

$$Af = \frac{df}{dx} + xf \quad \text{and} \quad A^* f = -\frac{df}{dx} + xf$$

then $A^* A = L - I$. These two operators are called the *annihilation* and *creation* operators respectively.

The uncertainty principle is strongly connected to the *noncommuting* natural of a certain natural pair of operators. Consider the *position* operator

$$Xf(x) = xf(x)$$

and the *momentum* operator

$$\widehat{Df}(\xi) = \xi \widehat{f}(\xi).$$

both continuous operators on $\mathcal{S}(\mathbf{R})$. Both of these operators are self adjoint. Given an arbitrary self-adjoint operator A on $\mathcal{S}(\mathbf{R})$, and $f \in \mathcal{S}(\mathbf{R})$

with $\|f\|_{L^2(\mathbf{R})} = 1$, we define the *standard deviation* of the operator to be

$$\Delta_A f = \left(\int_{-\infty}^{\infty} (A^2 f)(x)\overline{f(x)}\, dx - \left(\int_{-\infty}^{\infty} (Af)(x)\overline{f(x)}\, dx \right)^2 \right)^{1/2}$$

For two self adjoint operators A and B, the *Robertson uncertainty relation* says that for any $f \in \mathcal{S}(\mathbf{R})$ with $\|f\|_{L^2(\mathbf{R})} = 1$,

$$\Delta_A \cdot \Delta_B \geqslant \frac{1}{2} \int (AB - BA) f(x)\overline{f(x)}\, dx.$$

Since $[X, D]$ is the identity map, we see this version of the uncertainty principle implies the Heisenberg uncertainty principle given above.

5.4 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 *c*haracteristic 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 5.4. *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 5.5. *If X and Y are independant normal distributions, then $aX + bY$ is normally distributed.*

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

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

Theorem 5.6. *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 $R_n(x) = e^{ix} - 1 - (ix) - (ix)^2/2 - \cdots - (ix)^n/n!$. Then because of oscillation and the fundamental theorem of calculus,

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

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. □

5.5 The Wirtinger Inequality on an Interval

Theorem 5.7. *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 $A \in \mathbf{C}$, $B, C \in \mathbf{R}$. □

Corollary 5.8. *Given* $f \in C^1[a,b]$ *with* $\int_a^b f(t) \, dt = 0$,

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

5.6 Energy Preservation in the String equation

Solutions to the string equation are

If $u(t,x)$

5.7 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 5.9. *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

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

□

Chapter 6

Partial Derivatives and Harmonic Functions

6.1 Conjugate Poisson Kernel

Recall from our discussion of the Poisson kernel that for $f \in L^p(\mathbf{R})$ for $1 \leqslant p \leqslant \infty$, if we define

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

then we obtain a harmonic function on the upper half plane. TODO If we define

$$\frac{\partial v}{\partial t} = (2\pi i) \int_{-\infty}^{\infty} \widehat{f}(\xi) \xi e^{2\pi i \xi x - t|\xi|}\, d\xi$$

$$\frac{\partial v}{\partial x} = -|\xi| \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{2\pi i \xi x - t|\xi|}\, d\xi.$$

Chapter 7

Finite Character Theory

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

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

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

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

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

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

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

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

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

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

7.1 Fourier Analysis on Cyclic Groups

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

7.2 An Arbitrary Finite Abelian Group

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

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

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

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

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

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

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

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

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

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

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

where λ_i are the eigenvalues of T_k, and V_{λ_i} are the corresponding eigenspaces. Then if $v \in V_{\lambda_i}$, and $j < k$,

$$T_k T_j v = T_j T_k v = \lambda_i T_j v$$

so $T_j(V_{\lambda_i}) = V_{\lambda_i}$. Now on each V_{λ_i}, we may apply the induction hypotheis to diagonalize the $T_1, \dots, T_{k-1}$. Putting this together, we simultaneously diagonalize $T_1, \dots, T_k$. $\square$

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 7.3. *Let G be a finite abelian group. Then $\Gamma(G) \cong G$, and forms an orthonormal basis for the space of complex valued functions on G. For any function $f : G \to \mathbf{C}$,*

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

In this context, we also have Parseval's theorem

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

7.3 Convolutions

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

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

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

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

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

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

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

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

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.

7.4 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}).

7.5 Dirichlet's Theorem

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

Theorem 7.4. *If m and n are relatively prime, then the set*

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

contains infinitely many prime numbers.

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

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

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

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

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

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

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

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

Theorem 7.5. *For any $s > 1$,*

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

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 8

Complex Methods

In this chapter, we illustrate the intimate connection between the Fourier transform on the real line, and complex analysis. We have already seen some aspects of this for Fourier analysis on the Torus, with the connection between power series of analytic functions on the unit disk. The main theme is that if f is a function initially defined on the real line, then the problem of extending the function to be analytic on a neighbourhood of this line is connected to to the Fourier transform of f decaying very rapidly (for instance, exponential decay).

8.1 Fourier Transforms of Holomorphic Functions

For each $a > 0$, let $S_a = \{x+iy : |y| < a\}$ denote the horizontal strip of width $2a$. The next theorem says that functions extendable to be holomorphic on the strip have exponential Fourier decay.

Theorem 8.1. *Let $f : S_a \to \mathbf{C}$ be holomorphic, integrable on each horizontal line in the strip, such that $f(x+iy) \to 0$ as $|x| \to \infty$. Then if $\widehat{f}$ is the Fourier transform of the restriction of f to the real line, then for each $b < a$,*

$$|\widehat{f}(\xi)| \lesssim_b e^{-2\pi b|\xi|}.$$

Proof. For any $b < a$, R, and $\xi > 0$, consider the contour γ_R on the rectangle with corners $-R$, 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

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

where $f_b(x) = f(x-ib)$. But $|\widehat{f_b}(\xi)| \leqslant \|f_b\|_{L^\infty(\mathbf{R})} \lesssim_b 1$, which implies that

$$|\widehat{f}(\xi)| \lesssim_b e^{-2\pi i b\xi}.$$

A similar estimate when $\xi < 0$ completes the argument. □

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

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

Conversely, if f is *any* integrable function with $|\widehat{f}(\xi)| \lesssim e^{-2\pi a|\xi|}$, then $\widehat{f}$ is integrable so the Fourier inversion formula holds. If we define

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

then this gives a holomorphic extension of f which is well defined on S_a.

Pushing this result to an extreme leads to the Paley-Wiener theorem, which gives precise conditions when a function has a compactly supported Fourier transform.

Theorem 8.2. *A function $f : \mathbf{R} \to \mathbf{C}$ is bounded, integrable, and continuous. Then f extends to an entire function on the complex plane, such that for all z,*

$$|f(z)| \lesssim e^{2\pi M|z|},$$

if and only if $\widehat{f}$ is supported on $[-M, M]$.

Proof. If $\widehat{f}$ is supported on $[-M, M]$, then the Fourier inversion formula comes into play, telling us that for all $x \in \mathbf{R}$,

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

But then we can clearly extend f to an entire function by defining

$$f(z) = \int_{-M}^{M} \widehat{f}(\xi) e^{2\pi i \xi z}\, d\xi,$$

and then $|f(z)| \leqslant e^{2\pi i M|z|} \|\widehat{f}\|_{L^1[-M,M]} \lesssim e^{2\pi i M|z|}$.

Conversely, suppose f is an entire function such that for all $z \in \mathbf{C}$,

$$|f(z)| \leqslant A g(x) e^{2\pi M|y|},$$

where $g \geqslant 0$ is integrable on $\mathbf{R}$. We also assume that $f(x+iy) \to 0$ uniformly as $x \to -\infty$, independently of y. Then a contour shift down guarantees that for any y,

$$\begin{aligned}
\widehat{f}(\xi) &= \int_{-\infty}^{\infty} f(x) e^{-2\pi i \xi x}\, dx \\
&= \int_{-\infty}^{\infty} f(x - iy) e^{-2\pi i \xi (x - iy)}\, dx \\
&\leqslant A e^{2\pi M y - 2\pi \xi y} \int_{-\infty}^{\infty} g(x)\, dx \lesssim e^{2\pi(My - \xi y)}.
\end{aligned}$$

If $\xi > M$, then taking $y \to \infty$ shows $\widehat{f}(\xi) = 0$. A contour shift up instead gives $\widehat{f}(\xi) = 0$ if $\xi < -M$. Thus the proof is completed in this case.

Now suppose the weaker condition

$$|f(z)| \leqslant A e^{2\pi M|y|}.$$

For each $\varepsilon > 0$, let

$$f_\varepsilon(z) = \frac{f(z)}{(1 - i\varepsilon z)^2}.$$

Then f_ε is analytic in the lower half plane. Moreover,

$$|f_\varepsilon(x+iy)| \lesssim_\varepsilon \frac{A e^{2\pi M|y|}}{1 + x^2}.$$

Thus we can apply the previous shifting techniques to show that $\widehat{f_\varepsilon}(\xi) = 0$ for $\xi > M$. For $x \in \mathbf{R}$, we have $|f_\varepsilon(x)| \leqslant |f(x)|$, and since $f_\varepsilon \to f$ pointwise as $\varepsilon \to 0$, we can apply the dominated convergence theorem to imply $\widehat{f_\varepsilon}(\xi) \to$

$\widehat{f}(\xi)$ for each ξ. In particular, we find $\widehat{f}(\xi) = 0$ for $\xi > M$. A similar technique with the family of functions

$$f_\varepsilon(z) = \frac{f(z)}{(1 + i\varepsilon z)^2},$$

show that $\widehat{f}(\xi) = 0$ for $\xi < -M$.

Finally, it suffices to show that the condition

$$|f(z)| \lesssim e^{2\pi M|z|}$$

implies $|f(x+iy)| \lesssim e^{2\pi M|y|}$. To prove this, we can apply a version of the Phragmén-Lindelöf on the quandrant $\{x + iy : x, y > 0\}$. Let $g(z) = f(z)e^{-2\pi i My}$. Then we have

$$|g(x)| = |f(x)| \leqslant \|f\|_{L^\infty(\mathbf{R})},$$

and

$$|g(iy)| = |f(iy)|e^{-2\pi i My} \leqslant A.$$

Since g has at most exponential growth on the quadrant, we can apply the Phragmén-Lindelöf to conclude $|g(z)| \leqslant \max(A, \|f\|_{L^\infty(\mathbf{R})})$ for all z on the quandrant. A similar argument works for the other quadrants. Thus we conclude that for all $z \in \mathbf{C}$

$$|f(z)| \leqslant \max(A, \|f\|_{L^\infty(\mathbf{R})})e^{2\pi i M|y|},$$

and so we can apply the previous cases to conclude that $\widehat{f}$ is supported on $[-M, M]$. □

Remark. The Paley-Wiener theorem has several variants. For instance, if f is continuous, integrable, and $\widehat{f}$ is integrable, and we further assume that $\widehat{f}(\xi) = 0$ for all $\xi < 0$, then for $z = x + iy$, we can define

$$f(z) = \int_0^\infty \widehat{f}(\xi) e^{2\pi i \xi z} = \int_0^\infty \widehat{f}(\xi) e^{-2\pi \xi y} e^{2\pi i \xi x}$$

to extend f to an analytic function in the upper half-plane, i.e. for $y > 0$, which is also continuous and bounded for $y \geqslant 0$. Conversely, similar techniques to those above enable us to show that if f is continuous, integrable, $\widehat{f}$ is integrable, and we can extend f to an analytic function on the open upper half plane, which is continuous and bounded on the closed half plane, then contour shifting shows that $\widehat{f}(\xi) = 0$ for $\xi < 0$.

8.2 Classical Theorems by Contours

We now prove some classical theorems of Fourier analysis using techniques of harmonic analysis, given that the functions we study have holomorphic extensions to tubes.

Theorem 8.3. *Let $f : S_b \to \mathbf{C}$ be holomorphic. Then for any $x \in \mathbf{R}$,*

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

where $\widehat{f}$ is the Fourier transform of f restricted to the real-axis.

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

$$\int_{-\infty}^{\infty} \widehat{f}(\xi) e^{-2\pi i \xi x} = \int_0^{\infty} \widehat{f}(\xi) e^{-2\pi i \xi x} + \widehat{f}(-\xi) e^{2\pi i \xi x}.$$

Now if $b < a$, we can apply a contour integral argument to conclude that

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

Thus by Fubini's theorem, for each $x_0 \in \mathbf{R}$,

$$\begin{aligned} \int_0^{\infty} \widehat{f}(\xi) e^{2\pi i \xi x_0} &= \int_0^{\infty} \int_{-\infty}^{\infty} f(x - ib) e^{2\pi i \xi [x_0 - (x - ib)]}\, dx\, d\xi \\ &= \int_{-\infty}^{\infty} f(x - ib) \left(\int_0^{\infty} e^{2\pi i \xi [x_0 - (x - ib)]}\, d\xi \right) dx \\ &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{f(x - ib)}{(x - ib) - x_0}\, dx. \end{aligned}$$

Similarily, another application of Fubini's theorem implies

$$\begin{aligned} \int_0^{\infty} \widehat{f}(-\xi) e^{-2\pi i \xi x_0}\, d\xi &= \int_0^{\infty} \int_{-\infty}^{\infty} f(x + ib) e^{-2\pi i \xi [x_0 - (x + ib)]}\, dx\, d\xi \\ &= \int_{-\infty}^{\infty} f(x + ib) \int_0^{\infty} e^{-2\pi i \xi [x_0 - (x + ib)]}\, d\xi\, dx \\ &= \frac{-1}{2\pi i} \int_{-\infty}^{\infty} \frac{f(x + ib)}{[(x + ib) - x_0]}\, dx. \end{aligned}$$

In particular, we conclude that

$$\int \widehat{f}(\xi) e^{2\pi i \xi x_0} = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z - x_0},$$

where γ is the path traces over the two horizontal strips $x + ib$ and $x - ib$. Approximating this integral by rectangles, and then apply Cauchy's theorem, we find this value is equal to $f(x)$. $\square$

We can also prove the Poisson summation formula.

Theorem 8.4. *Let $f : S_a \to \mathbf{C}$ be holomorphic. Then*

$$\sum_{n \in \mathbf{Z}} f(n) = \sum_{n \in \mathbf{Z}} \widehat{f}(n),$$

where $\widehat{f}$ is the Fourier transform of f restricted to the real line.

Proof. The function

$$\frac{f(z)}{e^{2\pi i z} - 1}$$

is meromorphic, with simple poles on $\mathbf{Z}$, with reside equal to $f(n)$ at each $n \in \mathbf{Z}$. If we apply the Residue theorem to a curve γ_N travelling around the rectangle connecting the points $N + 1/2 - ib$, $N + 1/2 + ib$, $-N - 1/2 + ib$, and $-N - 1/2 - ib$, then we conclude

$$\sum_{|n| \leqslant N} f(n) = \int_{\gamma_N} \frac{f(z)}{e^{2\pi i z} - 1}\, dz.$$

These values converge to $\sum_{n \in \mathbf{Z}} f(n)$ as $N \to \infty$. But this means that

$$\sum_n f(n) = \int_\gamma \frac{f(z)}{e^{2\pi i z} - 1}\, dz,$$

where γ is the two horizontal strips at b and $-b$. Now we use the expansion

$$\frac{1}{z - 1} = \sum_{n=1}^\infty z^{-n},$$

for $|z| > 1$, to conclude

$$\int_{-\infty}^{\infty} \frac{f(x-ib)}{e^{2\pi i(x-ib)}-1}\,dx = \int_{-\infty}^{\infty} \sum_{n=1}^{\infty} \frac{f(x-ib)}{e^{2\pi n i(x-ib)}}\,dx$$
$$= \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} f(x-ib)e^{-2\pi n i(x-ib)}\,dx = \sum_{n=1}^{\infty} \widehat{f}(n),$$

where we have performed a contour shift at the end. Similarily, we use the expansion

$$\frac{1}{z-1} = -\sum_{n=0}^{\infty} z^n,$$

to conclude that

$$-\int_{-\infty}^{\infty} \frac{f(x+ib)}{e^{2\pi i(x+ib)}-1} = \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} f(x+ib)e^{2\pi i(x+ib)}\,dx$$
$$= \sum_{n=0}^{\infty} \widehat{f}(-n).$$

Combining these two calculations completes the proof. □

8.3 The Laplace Transform

We now look at things from the dual perspective. Instead of looking at whether a function can be extended to a holomorphic function, we look at whether the Fourier transform can be extended to a holomorphic function. For a function $x : \mathbf{R} \to \mathbf{R}$, this gives rise to the *Laplace transform*

$$X(z) = \int_{-\infty}^{\infty} x(t)e^{-zt}\,dt,$$

also denoted by $(\mathcal{L}x)(z)$. For $\xi \in \mathbf{R}$, $X(i\xi) = \widehat{x}(\xi)$ operates as the usual Fourier transform (slightly rescaled from the version in our notes). But the Laplace transform can also be extended to not-necessarily integrable functions. Given x, we can define $X(z)$ for any $z = x + iy$ such that

$$\int e^{-xt}|f(t)|\,dt < \infty.$$

It is simple to see this forms a vertical tube in the complex plane, called the *region of convergence* for the Laplace transform. For a particular vertical tube $I \subset \mathbf{C}$, we let $\mathcal{E}(I)$ be the collection of all functions x whose region of convergence for the Dirichlet transform contains I.

Example. *Let*

$$H(x) = \begin{cases} 0 & : x < 0, \\ 1/2 & : x = 0, \\ 1 & : x > 0. \end{cases}$$

The function H is called the Heavyside Step Function. *It's region of convergence consists of the right-most half plane, i.e. all $\omega + i\xi$, where $\omega > 0$. And if $z = \omega + i\xi$, we calculate that*

$$\mathcal{L}(H)(z) = \int_0^\infty e^{-zt}\, dt = z^{-1}.$$

We note that even though the integral formula does not define the Laplace transform of H in the right-most half plane, we can *analytically continue $\mathcal{L}(H)$ to a meromorphic function on the entire complex plane.*

Example. *Similarily, an integration by parts shows that for $z = \omega + i\xi$ with $\omega > 0$, we have*

$$\mathcal{L}(tH)(z) = \int_0^\infty te^{-zt} = \int_0^\infty \frac{e^{-zt}}{z} = z^{-2}.$$

Against, $\mathcal{L}(tH)$ extends to a meromorphic function on the entire complex plane.

What distinguishes the Laplace transform from the Fourier transform is the ability to use techniques of complex analysis. If x has region of convergence I, then X is continuous on I, and analytic on I°. We can even calculate an explicit formula for the derivative As expected from the Fourier transform of the derivative, if $y(t) = tx(t)$, and Y is the Laplace transform of y, then $X'(z) = -Y(z)$. One can verify this quite simply by taking limits of the derivatives of the analytic integrals

$$\int_{-N}^{N} x(t)e^{-zt}\, dt,$$

as $N \to \infty$. Like the Fourier transform, the Laplace transform is symmetric under modulation, translation, and polynomial multiplication:

- If $w \in \mathbf{C}$, and x is a function, set $y(t) = e^{wt}x(t)$. Then if z is in the region of convergence for x, $z - w$ is in the region of convergence for y, and $X(z) = Y(z - w)$.
- If x has region of convergence I, then the region of convergence for $y(t) = tx(t)$ contains I°, and $Y(z) = -X(z)$.
- If x has region of convergence I, $t_0 \in \mathbf{R}$, and we set $y(t) = x(t + t_0)$, then y has region of convergence I, and $Y(z) = e^{zt_0}X(z)$.
- For a function x, define

  $$(\Delta_s x)(t) = \frac{x(t+s) - x(t)}{s}.$$

  If ω is fixed, if

  $$\lim_{s \to 0} \int |(\Delta_s x)(t) - x'(t)| e^{-\omega t}\, dt = 0,$$

  if $y(t) = x'(t)$, and if $z = \xi + i\omega$ for some $\xi \in \mathbf{R}$, then $Y(z) = zX(z)$.

  In particular, this is true if x is supported on $[-N, \infty)$ for some N, has a continuous derivative x', and there is $\omega_0 < \omega$ such that

  $$\lim_{t \to \infty} x(t)e^{-\omega_0 t} = \lim_{t \to \infty} x'(t)e^{-\omega_0 t} = 0.$$

Remark. It will be interesting for us to consider functions x supported on $[-N, \infty)$ which have a piecewise continuous derivative x' except at finitely many points $t_1, \dots, t_N$, such that the left and right-hand limits exist at each t_i. For each $i \in \{1, \dots, N\}$, we let

$$A_i = x(t_i+) - x(t_i-) \quad \text{and} \quad B_i = x'(t_i+) - x'(t_i-).$$

If $y(t) = x'(t)$, we calculate a relation between the Laplace transforms of X and Y at $z = \omega + i\xi$ such that there exists $\omega_0 < \omega$ such that

$$\lim_{t \to \infty} x(t)e^{-\omega_0 t} = \lim_{t \to \infty} x'(t)e^{-\omega_0 t} = 0.$$

We consider the function

$$x_1(t) = x(t) - \sum_{i=1}^N A_i H(t - t_i) - \sum_{i=1}^N B_i (t - t_i) H(t - t_i).$$

Then x_1 is continuous everywhere, and moreover, has a continuous derivative. We have

$$x_1'(t) = x'(t) - \sum_{i=1}^{N} B_i H(t - t_i).$$

Thus if $\omega > 0$, and $z = \omega + i\xi$, if $y_1(t) = x_1'(t)$, we find

$$Y_1(z) = zX_1(z).$$

Now

$$Y_1(z) = Y(z) - \sum_{i=1}^{N} \frac{B_i e^{-izt_i}}{iz}$$

and

$$X_1(z) = X(z) - \sum_{i=1}^{N} \frac{A_i e^{-izt_i}}{iz} + \sum_{i=1}^{N} \frac{B_i e^{-izt_i}}{z^2}.$$

Thus, rearranging, we conclude

$$Y(z) = zX(z) - \sum_{i=1}^{N} A_i e^{-izt_i}$$

We can carry this through recursively to higher order derivatives. For each k, we set $A_i^k = f^{(k)}(t_i+) - f^{(k)}(t_i-)$. Then if $y(t) = f^{(n)}(t)$, then

$$Y(z) = z^n X(z) - \sum_{k=0}^{n-1} \sum_{i=1}^{N} z^{n-1-k} A_i^k e^{-izt_i}.$$

This is very useful when wants to solve differential equations, provided the solutions to those differential equations do not grow faster than exponentially.

Example. *Suppose we wish to find a formula for the unique real-valued function $x : [0, \infty) \to \mathbf{R}$ such that $x''(t) - x'(t) - 6x(t) = 5e^{3t}$ for $t \geqslant 0$, such that $x(0) = 6$ and $x'(0) = 1$. Such a function increases at most exponentially, since it is linear, so we may take the Laplace transform of each sides to conclude that if X is the Laplace transform of x, then*

$$\mathcal{L}(x'')(z) = z^2 X(z) - 6z - 1 \quad \textit{and} \quad \mathcal{L}(x')(z) = zX(z) - 6.$$

Thus we conclude

$$[z^2 X(z) - 6z - 1] - [zX(z) - 6] - (6X) = \frac{5}{z-3}.$$

Thus

$$X(z) = \frac{(3z-4)(2z-5)}{(z-3)^2(z+2)} = \frac{3.6}{z+2} + \frac{2.4}{z-3} + \frac{1}{(z-3)^2}.$$

But this implies that for $t \geqslant 0$, $x(t) = 3.6e^{-2t} + 2.4e^{3t} + te^{3t}$. In particular, we note that the pole of X determines the large scale behaviour of X, i.e. for large t, and for any $\varepsilon > 0$,

$$e^{(3-\varepsilon)t} \lesssim_\varepsilon x(t) \lesssim_\varepsilon e^{(3+\varepsilon)t}.$$

In the next section, we generalize this situation to give asymptotics of functions whose Laplace transforms extend to meromorphic functions on the complex plane.

8.4 Asymptotics via the Laplace Transform

For simplicity, in this chapter we study integrable functions $x : [0,\infty) \to \mathbf{R}$, whose Laplace transform is thus well defined on the closed, right half-plane. If the Fourier transform of x is integrable, then we can apply the inversion formula to conclude that for each $t \in \mathbf{R}$,

$$x(t) = \int_{-\infty}^{\infty} X(i\xi) e^{i\xi t}\, d\xi.$$

Now suppose that X can be analytically continued to a holomorphic function $X(\omega + i\xi)$ for all $\omega \geqslant -\varepsilon$ which is continuous at the boundary, such that, uniformly for $\omega \in [-\varepsilon, 0]$,

$$\lim_{|\xi|\to\infty} X(\omega + i\xi) = 0.$$

Then a contour shift argument implies that for each t,

$$x(t) = \lim_{R\to\infty} \int_{-R}^{R} X(-\varepsilon + i\xi) e^{(-\varepsilon + i\xi)t}\, d\xi = e^{-\varepsilon t} \lim_{R\to\infty} \int_{-R}^{R} X(-\varepsilon + i\xi) e^{i\xi t}\, d\xi.$$

For simplicity, we study functions supported on $[0,\infty)$. The region of convergence for such functions then takes the form of a half plane. For a given $a \in \mathbf{R}$, we let $\mathcal{E}_a$ be the set of functions whose region of convergence contains $\omega + i\xi$ for all $\omega > a$.

Theorem 8.5. *Suppose $x : [0, \infty) \to \mathbf{R}$ is a continuous function such that some ω,*

$$\int |x(t)| e^{-\omega t}\, dt < \infty.$$

Proof. Since $|X(u + iv)| \to 0$ uniformly as $v \to \infty$, we can shift the Fourier inversion formula

$$x(t) = \lim_{R \to \infty} \frac{1}{2\pi} \int_{-R}^{R} X(\omega + i\xi) e^{(\omega + i\xi)t}\, d\xi$$

(where the 2π comes up from our rescaling of the Fourier transform) to conclude that

$$x(t) = \lim_{R \to \infty} \frac{1}{2\pi}$$

$$X(z) = \lim$$

$\square$

Part II

Distributional Methods

Chapter 9

The Theory of Distributions

Distribution theory is a tool which enables us to justify formal manipulations in harmonic analysis without having to worry about technical issues arising form interpreting these manipulations analytically. For instance, the Fourier transform of a function is only defined as an absolutely convergent integral for functions in $L^1(\mathbf{R}^d)$. On the other hand, the theory of tempered distributions enables us to define the Fourier transform of *any* locally integrable function, and even more general functions. Thus distributions are a cornerstone to the formulation of many problems in modern harmonic analysis.

The path of modern analysis has extended analysis from the study of continuous and differentiable functions to measurable functions. The power of this approach is that we can study a very general class of functions. On the other hand, the more general the class of functions we work with, the more restricted the analytical operations we can perform. Nonetheless, $C_c^\infty(\mathbf{R}^d)$ is dense in almost all the spaces of measurable functions we consider in basic analysis, and for such functions we can apply all the fundamental analytical operations in this region. One approach to studying the general class of measurable functions is to prove results for elements of $\mathcal{D}(\mathbf{R}^d)$, and then apply an approximation result to obtain the result for a wider class of measurable functions. The theory of distributions provides a complimentary approach, using *duality* to formally extend analytical operations on $C_c^\infty(\mathbf{R}^d)$ to larger sets.

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 perspective is often not the most useful. This is most clear in measure theory, where we are used to treating a function only defined 'up to a set of measure zero', and thus not defined at any particular point. In distribution theory, we view functions as 'integrands', whose properties are understood 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 (or 'testing') against elements of $L^p(\mathbf{R}^d)$, i.e. through the linear functional on $L^p(\mathbf{R}^d)$ given by

$$\phi \mapsto \int_{\mathbf{R}^d} f(x)\phi(x)\, dx.$$

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 on $C(K)$ through the map

$$\phi \mapsto \int_K \phi(x) d\mu(x).$$

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 $\mathcal{D}(\mathbf{R}^d)$. Since most operations in analysis can be applied to elements of $\mathcal{D}(\mathbf{R}^d)$, we can then use duality to extend these operations to distributions. Moreover, since $\mathcal{D}(\mathbf{R}^d)$ is a very 'tame' space of functions, distributions are a very general family of generalized functions. The class $\mathcal{D}(\mathbf{R}^d)$ has proven to be the most natural class of functions for most problems studied in harmonic analysis. But one can apply the ideas described in this chapter to many other classes of test functions. Provided that the test functions can be suitably localized, one will likely obtain similar results to that described in this chapter. On the other hand, if one deals with non localizable families of test functions, one is likely to obtain quite a different theory of generalized functions. This is encountered, for instance, if one takes the family of analytic functions as the test functions, which gives the theory of *hyperfunctions*.

Remark. From the perspective of experimental physics, viewing functions as integrands is more natural than viewing functions in the set-theoretic sense. Indeed, 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 the only physically meaningful properties of a 'function' are those obtained by testing that function against some family of test functions, obtained from some physical measurements.

9.1 The Space of Test Functions

We fix an open subset Ω of $\mathbf{R}^n$, and let $\mathcal{D}(\Omega) = \mathcal{D}(\Omega)$ denote the family of all smooth functions on Ω with compact support. Our goal is to equip $\mathcal{D}(\Omega)$ with a complete locally convex topology, so that we can consider the dual space $\mathcal{D}^*(\Omega)$ of *distributions* on Ω. We could equip $\mathcal{D}(\Omega)$ 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 $\mathcal{D}(\Omega)$ is not complete.

Example. *Let $\Omega = \mathbf{R}$, pick a bump function $\phi \in \mathcal{D}(\mathbf{R})$ 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 $\mathcal{D}(\mathbf{R})$, 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 instead assign $\mathcal{D}(\Omega)$ a stronger locally convex topology which prevents convergent functions from 'escaping open sets'; the cost, however, is that the topology is no longer metrizable. The process we perform here is quite general and can be viewed as a way to construct the 'categorical limit' of a family of complete, locally convex spaces. For each compact set $K \subset \Omega$, the subspace $C_c^\infty(K) \subset \mathcal{D}(\Omega)$ is a complete metric space under the family of seminorms $\|\cdot\|_{C^n(K)}$. We consider a convex topology on $\mathcal{D}(\Omega)$ by considering the family of sets $\{\phi + W\}$ as a basis, where ϕ ranges over all elements of $\mathcal{D}(\Omega)$, and W ranges over all convex, balanced subsets of $\mathcal{D}(\Omega)$ such that $W \cap C_c^\infty(K)$ is open in $C_c^\infty(K)$ for each $K \subset \Omega$.

Theorem 9.1. *This gives a basis of a Hausdorff topology on $\mathcal{D}(\Omega)$.*

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 $\mathcal{D}(\Omega)$ 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 $W_\varepsilon = \{f \in \mathcal{D}(\Omega) : \|f\|_{L^\infty(\Omega)} < \varepsilon\}$, and it is easy to see these sets are open. Since $\bigcap_{\varepsilon > 0} W_\varepsilon = \{0\}$, this means $\phi = 0$. $\square$

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. *$\mathcal{D}(\Omega)$ 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,

$$\begin{aligned}\gamma\left(\phi + \frac{W}{2(|\lambda| + \varepsilon)}\right) &= \lambda\phi + (\gamma - \lambda)\phi + \frac{\gamma}{2(|\lambda| + \varepsilon)} W \\ &\subset \lambda\phi + W/2 + W/2 \subset \lambda\phi + W\end{aligned}$$

so multiplication is continuous. ☐

Theorem 9.3. *For each compact set $K \subset \Omega$, the canonical embedding of $C_c^\infty(K)$ in $\mathcal{D}(\Omega)$ is continuous.*

Proof. We shall prove a convex, balanced neighbourhood V is open in $\mathcal{D}(\Omega)$ 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 $C_c^\infty(K) \to \mathcal{D}(\Omega)$ 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 $\mathcal{D}(\Omega)$ 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

$$\|f\| = \min_{\mathrm{supp}(f) \subset U_n} \alpha_n \|f\|_{C^{k_n}(U_n)}$$

is well defined on $\mathcal{D}(\Omega)$ and continuous. But if $\{f_i\}$ is a sequence such that $\lim_{i\to\infty} f_i = 0$, then $\lim_{i\to\infty} \|f_i\| = 0$ for any choice of constants α_n and k_n. This means that, asymptotically, as we approach the boundary of Ω, the sequence $\{f_i\}$ must converge arbitrarily rapidly to zero. 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. *Consider any $E \subset \mathcal{D}(\Omega)$. Then E is a bounded subset of $\mathcal{D}(\Omega)$ 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

$$W_n = \left\{ \psi \in \mathcal{D}(\mathbf{R}^d) : |\psi(x_n)| < n^{-1}|\phi_n(x_n)| \right\}.$$

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 $\mathcal{D}(\Omega)$. $\square$

Corollary 9.5. *$\mathcal{D}(\Omega)$ 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). $\square$

Corollary 9.6. *$\mathcal{D}(\Omega)$ is quasicomplete.*

Proof. If $\phi_1, \phi_2, \dots$ is a Cauchy sequence in $\mathcal{D}(\Omega)$, 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 $\mathcal{D}(\Omega)$. $\square$

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 $f \in \mathcal{D}(\mathbf{R}^d)$ is a *tensor function* if there are $f_1, \dots, f_n \in \mathcal{D}(\mathbf{R})$ 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 $\mathcal{D}(\mathbf{R}^d)$.*

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 $\phi \in \mathcal{D}(\mathbf{R}^d)$ 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 $\mathcal{D}(\Omega)$ is the limit of metrizable spaces, it's linear operators still have many of the same properties as metrizable spaces.

Theorem 9.8. *If $T : \mathcal{D}(\Omega) \to X$ is a map from $\mathcal{D}(\Omega)$ 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 $C_c^\infty(K)$. Then T is bounded as a map from $C_c^\infty(K)$ to X, hence $\{T\phi_n\} \to 0$. (3) implies (4) 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 $\mathcal{D}(\Omega)$, 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 $\mathcal{D}(\Omega)$. More generally, any linear differential operator with coefficients in $\mathcal{D}(\Omega)$ is a continuous operator from $\mathcal{D}(\Omega)$ to itself.

- The inclusion $\mathcal{D}(\Omega) \to L^p(\Omega)$ 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 $g \in \mathcal{D}(\mathbf{R}^d)$, $f * g \in \mathcal{D}(\mathbf{R}^d)$. 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 $\mathcal{D}(\mathbf{R}^n)$. 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 $\mathcal{D}(\mathbf{R}^n)$, we obtain continuity of the operator on $\mathcal{D}(\mathbf{R}^n)$.

Theorem 9.9. *If a map $T : C_c^\infty(K_0) \to \mathcal{D}(\mathbf{R}^n)$ is continuous, then the image of $C_c^\infty(K_0)$ is actually $C_c^\infty(K_1)$ for some compact set K_1.*

Proof. Suppose there is a sequence $\{x_i\}$ in $\mathbf{R}^d$ with no limit point and smooth functions $\{\phi_i\}$ compactly supported on $C_c^\infty(K_0)$ such that

$$(T\phi_i)(x_i) \neq 0.$$

Then for any sequence $\{\alpha_i\}$ of positive scalars, the sequence $\{\alpha_i T\phi_i\}$ does not converge to zero, since the union of the supports of $\alpha_i T\phi_i$ is unbounded. This means $\alpha_i \phi_i$ does not converge to zero. But this is clearly not true, for if we let

$$\alpha_i = \frac{1}{2^i \|\phi_i\|_{C^i(\mathbf{R}^d)}},$$

then for any fixed n, $\lim_{i\to\infty} \|\alpha_i \phi_i\|_{C^n(\mathbf{R}^d)} = 0$, so the sequence $\{\alpha_i \phi_i\}$ converges to zero. Thus there cannot exist a sequence $\{x_i\}$, and so the union of the supports of $T(C_c^\infty(K_0))$ is supported on some compact set K_1. $\square$

Thus the topology on the space $\mathcal{D}(\mathbf{R}^d)$ is as strict as can be. As a consequence, we shall see that the weak-$*$ topology on $\mathcal{D}^*(\mathbf{R}^d)$ is essentially the weakest topology available in analysis. This is surprising, because we are still able to obtain the continuity of many operators in the dual space to $\mathcal{D}(\mathbf{R}^d)$.

9.2 The Space of Distributions

We now have the tools to explain the idea of a distribution. If $\psi \in \mathcal{D}(\Omega)$, then the linear functional $\Lambda[\psi]$ on $\mathcal{D}(\Omega)$ defined for each $\phi \in \mathcal{D}(\Omega)$ by setting

$$\Lambda[\psi](\phi) = \int \psi(x)\phi(x)\, dx$$

is continuous. Moreover, $\Lambda[\psi]$ determines ψ uniquely, and so we can safely identify ψ with $\Lambda[\psi]$ (thus looking at ψ from a 'distributional viewpoint'). The idea of the theory of distributions is to treat any continuous linear functional Λ on $\mathcal{D}(\Omega)$ as if it were given by integration against a test function. Thus for such a linear functional Λ, we often denote $\Lambda(\phi)$ by

$$\int_{\mathbf{R}^d} \Lambda(x)\phi(x)\, dx,$$

even if Λ is not given by integration against some function. The space $\mathcal{D}^*(\Omega)$ will be called the space of distributions on Ω.

One huge advantage of this approach is that we can generalize many analytical operations defined on $\mathcal{D}(\Omega)$ *distributionally* to give an operation on $\mathcal{D}^*(\Omega)$, even if the original analytical operations required some degree of smoothness to define. If A is an operator on $\mathcal{D}(\mathbf{R}^d)$ with adjoint A^*, then for any $\phi, \psi \in \mathcal{D}(\mathbf{R}^d)$,

$$\int_{\mathbf{R}^d} (A\phi)(x)\psi(x)\, dx = \int_{\mathbf{R}^d} \phi(x)(A^*\psi)(x)\, dx.$$

Thus given *any* distribution Λ, we define $A\Lambda$ to be the distribution such that

$$\int_{\mathbf{R}^d} (A\Lambda)(x)\phi(x)\, dx = \int_{\mathbf{R}^d} \Lambda(x)(A^*\phi)(x)\, dx.$$

Thus we have obtained a formal definition of A which works for arbitrary distributions.

For instance, we can use this idea to define the derivative of an arbitrary distribution. For $\phi, \psi \in \mathcal{D}(\mathbf{R})$, integration by parts tells us that

$$\int_{-\infty}^{\infty} \phi'(x)\psi(x)\, dx = -\int_{-\infty}^{\infty} \phi(x)\psi'(x)\, dx.$$

Thus if $A\phi = \phi'$ is the derivative operator then it's adjoint is $A^*\psi = -\psi'$. Thus, for a distribution Λ on $\mathbf{R}$, we define it's derivative to be the distribution Λ' such that for $\phi \in \mathcal{D}(\mathbf{R})$,

$$\Lambda'(\phi) = \Lambda(A^*\phi) = -\Lambda(\phi').$$

More generally, for a distribution Λ on $\mathbf{R}^d$, and a multi-index α, we 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*

$$\int_{-\infty}^{\infty} H'(x)\phi(x)\, dx = -\int_{-\infty}^{\infty} H(x)\phi'(x) = -\int_0^{\infty} \phi'(x) = \phi(0)$$

Thus the distributional derivative *of the Heaviside step function is the Dirac delta function. 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 δ such that for any $\phi \in \mathcal{D}(\mathbf{R})$,*

$$\int_{-\infty}^{\infty} \delta(x)\phi(x)\, dx = \phi(0).$$

Then

$$\int_{-\infty}^{\infty} \delta'(x)\phi(x)\, dx = -\int_{\mathbf{R}^d} \delta(x)\phi'(x)\, dx = -\phi'(0).$$

This is a distribution that does not arise from integration with respect to a locally integrable function nor integration against a measure.

In general, we define a *distribution* to be a continuous linear functional on the space of test functions $\mathcal{D}(\Omega)$, i.e. an element of $\mathcal{D}(\Omega)^*$. In the last section, our exploration of continuous linear transformations on $\mathcal{D}(\Omega)$ guarantees that a linear functional Λ on $\mathcal{D}(\Omega)$ is continuous if and only if for every compact $K \subset X$ there is an integer n_k such that $|\Lambda\phi| \lesssim_K \|\phi\|_{C^{n_k}(K)}$ 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. Applying the Hahn-Banach theorem shows that if $\Lambda \in \mathcal{D}^*(\Omega)$ has order n, then Λ extends uniquely to a continuous functional on $C_c^n(\Omega)$.

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. This is because any non-negative distribution is actually given by integration against a Radon measure, and thus has order zero. Given a non-negative functional Λ (which is automatically continuous), we define Λf for a compactly supported continuous function $f \geqslant 0$ as

$$\Lambda f = \sup\{\Lambda g : g \in \mathcal{D}(\mathbf{R}^n), g \leqslant f\}$$

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.

Example. *If μ is a complex-valued Radon measure, then we can define a distribution $\Lambda[\mu]$ such that for each $\phi \in \mathcal{D}(\mathbf{R}^d)$.*

$$\Lambda[\mu](\phi) = \int_{\mathbf{R}^d} \phi(x) d\mu(x)$$

Thus $\Lambda[\mu]$ is a distribution, since if ϕ is supported on K, then

$$|\Lambda[\mu](\phi)| \leqslant \mu(K) \|\phi\|_{L^\infty(K)}.$$

The fact that this bound does not require information about the derivatives of ϕ implies that $\Lambda[\mu]$ is a distribution of order zero. In particular, the last pararaph, together with the Riesz-Markov-Kakutani representation theorem, shows that any *distribution of order zero is given by a complex-valued Radon measure.*

Example. *Consider a functional Λ defined for functions $\phi \in \mathcal{D}(\mathbf{R})$ vanishing in a neighbourhood of the origin by setting*

$$\Lambda(\phi) = \int_{-\infty}^{\infty} \frac{\phi(x)}{x}\, dx.$$

Such functions are not *dense in $\mathcal{D}(\mathbf{R})$. But we claim Λ is bounded on it's domain, and thus by the Hahn-Banach theorem, extends to at least one continuous functional on the entirety of $\mathcal{D}(\mathbf{R})$. To prove this, fix $\phi \in C_c^\infty[-N,N]$ vanishing on a neighbourhood $(-\varepsilon,\varepsilon)$ of the origin. Then*

$$|\Lambda\phi| = \left| \int_{-\infty}^{\infty} \frac{\phi(x)}{x}\, dx \right| = \left| \int_{\varepsilon \leqslant |x| \leqslant N} \frac{\phi(x) - \phi(0)}{x}\, dx \right|.$$

Applying the mean-value theorem, we find

$$|\Lambda\phi| \leqslant N \|\phi\|_{C^1[-N,N]}.$$

Since N was arbitrary, it follows that Λ is continuous in the topology induced by that of $\mathcal{D}(\mathbf{R})$, and thus by the Hahn-Banach theorem, extends uniquely to at least one distribution on the entirety of $\mathcal{D}(\mathbf{R})$.

One canonical choice of Λ *is the* principal value distribution $p.v.(1/x)$, *defined such that*

$$\int_{-\infty}^{\infty} p.v.(1/x)\phi(x)\, dx = \lim_{\delta \to 0} \int_{|x| \geqslant \delta} \phi(x)/x\, dx.$$

We essentially showed that this functional was continuous above. Another choice is the distribution $\lim_{\varepsilon \to 0} 1/(x + i\varepsilon)$, *defined such that*

$$\int_{-\infty}^{\infty} \lim_{\varepsilon \to 0} 1/(x + i\varepsilon) \cdot \phi(x)\, dx = \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \phi(x)/(x + i\varepsilon)\, dx.$$

If we pick $\delta = \varepsilon^{1/4}$, *then we can show using the fact that* $1/x$ *and* $1/(x+i\varepsilon)$ *are not too different for large x that*

$$\left| \int_{|x| \geqslant \delta} \frac{\phi(x)}{x} - \int_{-\infty}^{\infty} \frac{\phi(x)}{x + i\varepsilon} \right| \leqslant \|\phi\|_1 \cdot \varepsilon^{1/2}.$$

A contour integral shift shows that

$$\begin{aligned} \int_{-\delta}^{\delta} \frac{\phi(x)}{x + i\varepsilon} &= \int_{-\delta}^{\delta} \frac{\phi(0)}{x + i\varepsilon} + O(\delta) \\ &= -i\pi\phi(0) + O(\varepsilon/\delta) + O(\delta) \\ &= -i\pi\phi(0) + O(\varepsilon^{1/4}). \end{aligned}$$

Taking $\varepsilon \to 0$ *shows that*

$$p.v(1/x) = i\pi\delta + \lim_{\varepsilon \to 0} 1/(x + i\varepsilon),$$

where δ *is the Dirac delta distribution at the origin.*

More generally, if Λ_1 *and* Λ_2 *are two distributions which extend the functional* Λ, *then one can show that for any function* ϕ *vanishing away from the origin,* $\Lambda_1(\phi) - \Lambda_2(\phi) = 0$. *We will later define the* support *of a distribution, and so we have shown here that* $\Lambda_1 - \Lambda_2$ *is supported at* $\{0\}$. *It follows from later theorems in this chapter than* Λ_1 *and* Λ_2, *applied to a function* $\phi \in \mathcal{D}(\mathbf{R})$, *will differ by a finite linear combination of the values of* ϕ *and it's derivatives at the origin.*

The distribution $p.v(1/x)$ can also be described as the distributional derivative of the locally integrable function $\log|x|$*, since an integration by parts shows that for each $\phi \in \mathcal{D}(\mathbf{R}^d)$,*

$$\begin{aligned}
\int (\log|x|)' \, \phi(x) \, dx &= -\int \log|x| \phi'(x) \, dx \\
&= \lim_{\varepsilon \to 0} \int_{|x| \geqslant \varepsilon} \log|x| \phi'(x) \\
&= \lim_{\varepsilon \to 0} \left(\log(\varepsilon) \cdot (\phi(x) - \phi(-x)) + \int_{|x| \geqslant \varepsilon} \frac{\phi(x)}{x} \right) \\
&= p.v. \int \frac{\phi(x)}{x} \, dx.
\end{aligned}$$

An important elementary application of these distributions arises in the theory of the Hilbert transform.

Example. *One reason we could define a distribution agreeing with $1/x$ away from the origin is because there is a lot of cancellation at the origin from either side of the origin, since $1/x$ switches sign here. One has to rely on other tricks to make sense of a distribution extending $1/x^2$. Indeed, if we write*

$$\Lambda(\phi) = \int \frac{\phi(x)}{x^2}$$

for ϕ ranging over all functions vanishing in the neighborhood of the origin, then we can use the mean value theorem to obtain a bound $|\phi(x)| \leqslant x^2 \|\phi''\|_\infty$, from which it follows that

$$\Lambda(\phi) \lesssim \|\phi''\|_\infty,$$

and so Hahn-Banach extends Λ to a family of distributions. But in this case the principal value

$$\lim_{\delta \to 0} \int_{|x| \geqslant \delta} \frac{\phi(x)}{x^2}$$

rarely exists. Indeed, for any fixed $\phi \in \mathcal{D}(\mathbf{R})$ we have

$$\int_{|x| \geqslant \delta} \frac{\phi(x)}{x^2} = 2\phi(0)/\delta + O(\delta).$$

which will only converge if $\phi(0) = 0$. Thus, to get around this, we define the finite part distribution *(or* Hadamard regularization*) of $1/x^2$, i.e. the distribution f.p$(1/x^2)$ by setting*

$$\int f.p(1/x^2)\,\phi(x)\,dx = \lim_{\delta\to 0}\left(\int_{|x|\geqslant\delta}\phi(x)/x^2 - \frac{2\phi(0)}{\delta}\right),$$

which gets around the result that the distribution might explode near the origin if $\phi(0) \neq 0$ ($\phi'(0)$ does not cause a problem because of cancellation on both sides of the integral). Another approach is to consider the derivative of the distribution $-p.v(1/x)$, since the derivative of this distribution agrees with integration against $1/x^2$ away from the origin. In fact, the derivative of $-p.v(1/x)$ is precisely f.p$(1/x^2)$. We leave it to the reader to use similar tricks to define the finite parts of higher order singularities, such as $1/x^3$.

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 $\phi \in \mathcal{D}(\mathbf{R})$,*

$$\begin{aligned}\int_{-\infty}^{\infty}\phi(x)d\mu(x) &= -\int_{-\infty}^{\infty}\int_x^{\infty}\phi'(y)\,dy\,d\mu(x)\\ &= -\int_{-\infty}^{\infty}\phi'(y)\int_{-\infty}^{y}d\mu(x)\,dy\\ &= -\int_{-\infty}^{\infty}\phi'(y)f(y)dy\end{aligned}$$

Thus if Λ is the distribution corresponding to integration with respect to $f(x)\,dx$, then Λ' is given by integration with respect to μ. In particular, Λ' is given by integration with respect to $f'(x)\,dx$ precisely when f is absolutely continuous.

Example. *If $f \in C^1(\mathbf{R} - \{0\})$, and if the function $v(x)$ defined to be $f'(x)$ for $x \neq 0$ is integrable, then the limits $f(0-)$ and $f(0+)$ both exist (a simple argument using the fundamental theorem of calculus), and the distributional derivative of f is equal to*

$$f' = v + (f(0+) - f(0-))\delta_0.$$

To see this, we calculate that for $\phi \in C_c^\infty(\mathbf{R})$,

$$\begin{aligned}
\int f'(x)\phi(x)\,dx &= -\int f(x)\phi'(x)\,dx \\
&= \lim_{\varepsilon\to 0} -\int_{-\infty}^{-\varepsilon} f(x)\phi'(x)\,dx - \int_{\varepsilon}^{\infty} f(x)\phi'(x)\,dx \\
&= \lim_{\varepsilon\to 0} f(\varepsilon)\phi(\varepsilon) - f(-\varepsilon)\phi(-\varepsilon) + \int_{-\infty}^{-\varepsilon} v(x)\phi(x)\,dx + \int_{-\infty}^{-\varepsilon} v(x)\phi(x)\,dx \\
&= \int v(x)\phi(x)\,dx + [f(0+) - f(0-)]\phi(0).
\end{aligned}$$

As a particular example of this, the distributional derivative of $|x|$ is $sgn(x)$, and the distributional derivative of the Heaviside step function H given above is evaluated to be δ_0.

There are many other important operations one can apply to distributions. If Ω is a conic subset of $\mathbf{R}^d$, and $\phi, \psi \in \mathcal{D}(\Omega)$, we find

$$\int_\Omega \mathrm{Dil}_\lambda \phi(x)\psi(x)\,dx = \lambda^{-d} \int_\Omega \phi(x)\cdot \mathrm{Dil}_{1/\lambda}\psi(x)\,dx,$$

Thus if Λ is a distribution on Ω, then we define $\mathrm{Dil}_\lambda \Lambda$ by setting

$$\mathrm{Dil}_\lambda \Lambda(\phi) = \lambda^{-d}\Lambda(\mathrm{Dil}_{1/\lambda}\phi).$$

For $f \in C^\infty(\Omega)$, we have an operator $\phi \mapsto f\phi$ on $\mathcal{D}(\Omega)$. The adjoint is clearly $\psi \mapsto f\psi$, so for a distribution Λ on Ω, we define $f\Lambda$ by setting $(f\Lambda)(\phi) = \Lambda(f\phi)$. Thus $\mathcal{D}^*(\Omega)$ is naturally a $C^\infty(\Omega)$ module. Similarily, the family $\mathcal{D}^*(\Omega)_k$ consisting of distributions of order k form a $C^k(\Omega)$ module.

Remark. Hörmander developed a sophisticated theory that enables us to define the product of two *distributions* using the Fourier transform. In many basic situations, one can perform a spatial decomposition to define the product. Given a distribution Λ, we define it's *singular support* $\mathrm{supp}_{\mathrm{sing}}(\Lambda)$ to be the *complement* of the set of all points x which have a neighborhood U such that $\Lambda|_U \in C^\infty(U)$. For any two distributions Λ and Ψ whose singular supports are disjoint, a decomposition argument enables us to define the product $\Lambda \cdot \Psi$ in a natural way.

Since $\mathcal{D}^*(\Omega)$ is the dual space of a topological vector space, we can give it a natural topology, the weak $*$ topology. Thus a net of distributions $\{\Lambda_\alpha\}$ 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 distributional sense' if we have a convergence $\Lambda[f_\alpha] \to \Lambda[f]$ for the corresponding distributions.

Example. *The distribution* $\lim_{\varepsilon \to 0} 1/(x + i\varepsilon)$ *defined above is precisely the limit of the distributions* $1/(x+i\varepsilon)$ *in the weak* $*$ *topology. Similarily,* $p.v(1/x)$ *is the weak* $*$ *limit of the functions* $\mathbf{I}_{|x| \geqslant \delta}(x) \cdot (1/x)$*. The distribution* $f.p(1/x^2)$ *is the distributional limit of* $\mathbf{I}_{|x|\geqslant\delta}(x) \cdot (1/x^2) - 2\delta_0/\delta$*, where* δ_0 *is the Dirac delta function at the origin.*

Example. *If* n *is a positive integer, then integration by parts shows that for any* $\phi \in \mathcal{D}(\mathbf{R})$*,*

$$\int_{-\infty}^{\infty} t^n e^{2\pi i t x} \phi(x)\, dx = i^{n+1} t^{-1} \int_{-\infty}^{\infty} e^{itx} \phi^{(n+1)}(x)\, dx,$$

which converges to zero as $t \to \infty$*. Thus* $t^n e^{2\pi i t x}$ *converges distributionally to zero as* $t \to \infty$*. Another way to see this is to note that the distribution* Λ_t *given by integration against* $t^n e^{itx}$ *can be written as* $\Lambda_t(\phi) = t^n \widehat{\phi}(-t)$*, and the Fourier transform of* ϕ *decays rapidly. If we tested against functions that were less smooth, then this statement would no longer be true.*

Example. *Let* $u_t(x) = t^{1/k} e^{itx^k}$*, where* k *is an integer bigger than one. Let*

$$F(x) = \int_0^x e^{iy^k}\, dy.$$

When $x > 0$*, a contour integration shift shows that*

$$F(x) = e^{i\pi/2k} \int_0^x e^{-y^k}\, dy + O(|x|^{-(k-1)}).$$

If k *is even, then for* $x < 0$*,*

$$F(x) = -e^{i\pi/2k} \int_0^x e^{-y^k}\, dy + O(|x|^{-(k-1)})$$

and for k odd,

$$F(x) = -e^{-i\pi/2k} \int_0^x e^{-y^k}\, dy + O(|x|^{-(k-1)}).$$

Thus given $\phi \in \mathcal{D}(\mathbf{R})$, we can apply an integration by parts to write

$$\begin{aligned}
\int_{-\infty}^{\infty} u_t(x)\phi(x)\, dx &= \int_{-\infty}^{\infty} t^{1/k} e^{itx^k} \phi(x)\, dx \\
&= \int_{-\infty}^{\infty} t^{1/k} F'(t^{1/k}x)\phi(x)\, dx \\
&= -\int_{-\infty}^{\infty} F(t^{1/k}x)\phi'(x)\, dx,
\end{aligned}$$

By decomposing this integral into the region where $|x| \geqslant t^{1/k}$ and $|x| \leqslant t^{1/k}$ shows that this quantity converges to

$$-F(\infty)\int_0^{\infty} \phi'(x)\, dx - F(-\infty)\int_{-\infty}^{0} \phi'(x)\, dx = (F(\infty) - F(-\infty))\phi(0).$$

Thus u_t converges distributionally to $\left(2e^{i\pi/2k}\int_0^\infty e^{-y^k}\, dy\right)\cdot\delta_0$ if k is even, and to $\left(2\cos(\pi/2k)\int_0^\infty e^{-y^k}\, dy\right)\cdot\delta_0$ if k is odd.

Since convergence in $\mathcal{D}(\Omega)$ is incredibly strict, a sequence of distributions can very easily converge in the weak $*$ topology. The following is thus quite a surprising result.

Theorem 9.10. *Suppose that $\{\Lambda_\alpha\}$ converges weakly to a distribution Λ. Then $D^\alpha \Lambda_i$ converges weakly to $D^\alpha\Lambda$ for any multi-index α.*

Proof. For each $\phi \in \mathcal{D}(\Omega)$, $D^\alpha\phi \in \mathcal{D}(\Omega)$, so

$$\begin{aligned}
\lim_\alpha (D^\alpha\Lambda_i)(\phi) &= (-1)^{|\alpha|}\lim_\alpha \Lambda_i(D^\alpha\phi) \\
&= (-1)^{|\alpha|}\Lambda(D^\alpha\phi) \\
&= (D^\alpha\Lambda)(\phi).
\end{aligned}$$

□

Thus differentiation is continuous in the space of distributions.

Theorem 9.11. *Fix $\{g_\alpha\}$ in $C^\infty(\Omega)$ and $\{\Lambda_\alpha\}$ in $\mathcal{D}^*(\Omega)$ such that $g_\alpha \to g$ in $C^\infty(\mathbf{R}^d)$ and $\Lambda_\alpha \to \Lambda$ weakly. Then $g_\alpha \Lambda_\alpha$ converges weakly to $g\Lambda$.*

Proof. For each $\phi \in \mathcal{D}(\mathbf{R}^d)$, the map

$$(\Lambda_\alpha, g_\alpha) \mapsto (\Lambda_\alpha(x) g_\alpha)(\phi)$$

is bilinear, and continuous in each variable. The result then follows from a variant of Banach-Steinhaus. □

There is also an often useful result resulting from bounded countable families of distributions.

Theorem 9.12. *Suppose $\mathcal{U} \subset \mathcal{D}^*(\Omega)$ is a family of distributions such that for each $\phi \in \mathcal{D}(\Omega)$, $\sup_{u \in \mathcal{U}} |u(\phi)| < \infty$. Then for every compact set K, there exists m such that for an $u \in \mathcal{U}$ and $\phi \in C_c^\infty(K)$,*

$$|u(\phi)| \lesssim \|\phi\|_{C^m(K)}.$$

If $\{u_n\}$ is a sequence of distributions for which $\lim_n u_n(\phi)$ exists for every $\phi \in \mathcal{D}(\Omega)$, then $u(\phi) = \lim_n u_n(\phi)$ defines a distribution, and for every compact set K there is an integer m such that for each $\phi \in C_c^\infty(K)$,

$$|u_n(\phi)| \lesssim \|\phi\|_{C^m(K)}$$

and

$$\lim_{n \to \infty} \sup_{\phi \in C_c^\infty(K)} \frac{|u(\phi) - u_n(\phi)|}{\|\phi\|_{C^m(K)}} = 0.$$

Proof. Each distribution in $\mathcal{U}$ acts as a continuous operator on the Frechét space $C_c^\infty(K)$, and this family satisfies the uniform boundedness principle, and the existence of an n as above follows as a result of the uniform boundedness principle, i.e. it shows that restricted to K, the distributions in $\mathcal{U}$ are equicontinuous.

Now assume the second condition. This clearly implies the first, hence we get the uniform boundedness property above. Now a ball of finite radius in $C^{m+1}(K)$ is precompact in $C^m(K)$, by the Arzela-Ascoli theorem. Thus we can find $\phi_1, \dots, \phi_N \in C_c^\infty(K)$ such that if $\phi \in C_c^\infty(K)$ and $\|\phi\|_{C^{n+1}(K)} \leqslant 2$, then there exists i such that $\|\phi - \phi_i\|_{C^n(K)} \leqslant \varepsilon$. Pick n_0

such that for any $n \geqslant n_0$ and $1 \leqslant i \leqslant N$, $|(u - u_n)(\phi_i)| \leqslant \varepsilon$. Then given any $\phi \in C_c^\infty(K)$ with $\|\phi\|_{C^{n+1}(K)} \leqslant 1$, we can find i as above, and then

$$\begin{aligned} |(u - u_n)(\phi)| &\leqslant |(u - u_n)(\phi - \phi_i)| + |(u - u_n)(\phi_i)| \\ &\lesssim \|\phi - \phi_i\|_{C^n(K)} + \varepsilon \\ &\lesssim \varepsilon. \end{aligned}$$

Thus we have proven the required limiting statement. $\square$

9.3 Homogeneous Distributions

An important family of distributions are the *homogenous distributions*, which are those distributions Λ on $\mathbf{R}^d - \{0\}$ such that for each $\lambda > 0$, $\text{Dil}_\lambda \Lambda = \lambda^\alpha \Lambda$, where α is the *order* of the homogenous distribution Λ.

Example. *If $f \in L_1^{loc}(\mathbf{R}^d)$ and $f(\lambda x) = \lambda^\alpha f(x)$ for all $x \in \mathbf{R}^d - \{0\}$ then integration against $f(x)\,dx$ defines a homogenous distribution of order α.*

Example. *For any complex number a with $Re(a) > -1$, if we define a distribution on $\mathbf{R} - \{0\}$ by setting*

$$x_+^a = \begin{cases} x^a & x > 0 \\ 0 & : x <= 0 \end{cases},$$

then x_+^a is a homogeneous distribution of order α, and $x \cdot x_+^a = x_+^{a+1}$, and if $Re(a) > 0$,

$$\frac{d}{dx}\left(x_+^a\right) = a x_+^{a-1}.$$

Our goal is to extend this distribution to a larger range of values $a \in \mathbf{C}$, such that the association $a \mapsto x_+^a$ is continuous. For any $\phi \in \mathcal{D}(\mathbf{R})$, the function

$$a \mapsto \langle x_+^a, \phi \rangle$$

is analytic in a for $Re(a) > -1$. Integration by parts shows that

$$\langle x_+^{a+1}, \phi' \rangle = -(a+1)\langle x_+^a, \phi \rangle.$$

The formula $\langle x_+^a, \phi \rangle = -(a+1)^{-1}\langle x_+^{a+1}, \phi' \rangle$ allows us to extend the definition of x_+^a to all $a \in \mathbf{C}$ with $Re(a) > -2$, except that we have a pole of order one *when $a = -1$. Iterating this allows us to uniquely extend the definition of x_+^a for all $a \in \mathbf{C}$, except when a is a negative integer, and these distributions will all be homogeneous.*

Marcel Riesz also used some other complex analytic tricks to define x_+^{-k} for all integers k, but then we lose some of the homogeneity. For any $\phi \in \mathcal{D}(\mathbf{R})$, the function $a \mapsto \langle x_+^a, \phi \rangle$ is meromorphic, with simple poles at each negative integer $-k$, and the residue at $-k$ is equal to $(-1)^k D^{k-1}\phi(0)/(k-1)!$, then we conclude that for any $\phi \in \mathcal{S}(\mathbf{R})$, as $a \to -k$,

$$\langle (a+k)x_+^a, \phi \rangle \to (-1)^{k-1} D^{k-1}\phi(0)/(k-1)!$$

In fact, expanding things out gives a constant $C_{-k}(\phi)$ such that as $a \to k$,

$$\langle x_+^a, \phi \rangle = \frac{(-1)^{k-1} D^{k-1}\phi(0)}{(k-1)!} \cdot \frac{1}{a+k} + C_{-k}(\phi) + O(a+k).$$

We define $\langle x_+^{-k}, \phi \rangle = C_{-k}(\phi)$, i.e. by keeping only the finite part *of the integral. Since, for a close to k, we have*

$$\begin{aligned}
\langle x_+^a, \phi \rangle &- \frac{(-1)^{k-1} D^{k-1}\phi(0)}{(k-1)!} \cdot (a+k)^{-1} \\
&= (-1)^k (a+1)^{-1} \cdots (a+k-1)^{-1} \int_0^\infty \frac{x^{a+k}-1}{a+k} D^k \phi(x)\, dx \\
&\quad + \frac{(-1)^{k-1}}{a+k} \left((-1-a)^{-1} \dots (1-a-k) - \frac{1}{(k-1)!} \right) D^{k-1}\phi(0) \\
&\to \frac{-1}{(k-1)!} \int_0^\infty \log(x) D^k \phi(x)\, dx + \frac{1}{(k-1)!} \left(\sum_{i=1}^k 1/i \right) D^{k-1}\phi(0).
\end{aligned}$$

and thus

$$\langle x_+^{-k}, \phi \rangle = \frac{-1}{(k-1)!} \int_0^\infty \log(x) D^k \phi(x)\, dx + \frac{1}{(k-1)!} \left(\sum_{i=1}^k 1/i \right) D^{k-1}\phi(0).$$

Our extension of x_+^a for all $a \in \mathbf{C}$ continues to satisfy $x \cdot x_+^a = x_+^{a+1}$ for all $a \in \mathbf{C}$. The derivative formula is not *maintained, namely,*

$$D x_+^{-k} = -k x_+^{-k-1} + (-1)^k D^k \delta / k!$$

Moreover, x_+^{-k} is no longer homogeneous of degree $-k$. Plugging into the formula shows that

$$(tx)_+^{-k} = t^{-k}x_+^{-k} + \frac{\log t}{(k-1)!} \cdot D^{k-1}\delta.$$

One can also define x_+^a by first removing the singularity, considering the distributions

$$\langle x_\varepsilon^a, \phi \rangle = \int_\varepsilon^\infty x^a \phi(x)\, dx.$$

If k is the smallest non-negative integer such that $k + Re(a) > -1$, then we can integrate by parts to conclude that there are constants C_k such that

$$\langle x_\varepsilon^a, \phi \rangle = \sum_{i=0}^{k-1} C_i \varepsilon^{-i} + (-1)^k \int_0^\infty \frac{1}{(a+1)\dots(a+k)} x^{a+k} D^k \phi(x)\, dx + o(1).$$

Discarding the singular terms, and letting $\varepsilon \to 0$ gives the distributions x_+^a above. One can analogously define the distributions x_-^a, and the distributions $|x|^a$, by reflecting and symmetrizing the distributions about the origin.

Example. *One can fix the singularities at the integers by normalizing. The appearance of the singularities in the extension of x_+^a occured because of the a term in the formula $D(x_+^a) = ax_+^{a-1}$. If we consider the normalization $\chi_+^a = x_+^a/\Gamma(a)$ when $Re(a) > -1$, then $D(\chi_+^a) = -\chi_+^{a-1}$, and because the Gamma function has no zeroes, this allows us to extend the definition of χ_+^a to an analytic function for all $a \in \mathbf{C}$, each a homogeneous distribution of order a. But since χ_+^0 is the Heaviside step function, we conclude that $\chi_+^{-k} = D^{k-1}\delta$.*

Example. *Let δ be the Dirac-delta distribution at the origin in $\mathbf{R}^d$. Then for $\phi \in \mathcal{D}(\mathbf{R}^d)$,*

$$\int_{\mathbf{R}^d} (Dil_\lambda \delta)(x)\phi(x)\, dx = \lambda^{-d} \int_{\mathbf{R}^d} \delta(x) Dil_{1/\lambda}\phi(x)\, dx = \lambda^{-d}\phi(0).$$

Thus $Dil_\lambda \delta = \lambda^{-d}\delta$, which implies δ is a homogenous distribution of order $-d$.

A homogeneous distribution is apriori defined by testing against a distribution in $\mathcal{D}(\mathbf{R}^d - \{0\})$. But in most cases the distribution can be extended so that it can be tested against an arbitrary distribution in $\mathcal{D}(\mathbf{R}^d)$.

Before we prove this, we begin with some simple observations. First is a formula due to Euler, which in the distributional setting states that for any distribution u of degree a,

$$\sum_{i=1}^{d} x_i \partial_i u = (a+d)u.$$

When $d = 1$, this actually implies that u is a multiple of $|x|^a$ on each coordinate axis. The identity also implies that for any $\psi \in C_c^\infty(\mathbf{R}^d - \{0\})$ with $\int_0^\infty r^{a+d-1}\psi(rx)\, dr = 0$ for all x, $\langle u, \phi \rangle = 0$ by rewriting the formula in polar coordinates.

Theorem 9.13. *Let u be a homogeneous distribution of order a, which is not a negative integer smaller than or equal to $-n$. Then u has a unique extension to a distribution $E(u)$ on $\mathbf{R}^d$, such that for any homogeneous polynomial P, $E(Pu) = PE(u)$, and if u is not a distribution of order $1-n$, $E(\partial_i u) = \partial_i E(u)$. Moreover, the map $u \mapsto E(u)$ is continuous between the two spaces of distributions.*

Proof. The uniqueness is obvious, because any distribution supported at the origin is a linear combination of derivatives of the Dirac delta function, which are all homogeneous of integer order $\leqslant -n$. To show existence, we note that if u was locally integrable, then polar coordinates gives

$$\int u(x)\phi(x) = \int_0^\infty \int_{|w|=1} u(w) t^{a+d-1} \phi(tw)\, d\sigma(w)\, dt.$$

Thus we need only study the behaviour of u near the unit sphere, which is supported away from the origin. Doing this more formally yields the extension map E. For any $\phi \in \mathcal{D}(\mathbf{R}^d)$, define $R_a\phi(x) = \langle t_+^{a+d-1}, \phi(tx) \rangle$. Then $R_a\phi$ is homogeneous of degree $-n-a$, and is continuous from $C_c^\infty(K)$ to $C^\infty(\mathbf{R}^d - \{0\})$ for any compact set K. If $\psi \in C_c(\mathbf{R}^d - \{0\})$ and $\int_0^\infty \psi(tx)/t\, dt = 1$ for all $x \neq 0$, then $\psi R_a \phi \in C_c^\infty(\mathbf{R}^d - \{0\})$, and $R_a(\psi R_a \phi) = R_a \phi$, so that (because of our observations before the proof) $E(u) = \langle u, \psi R_a \phi \rangle$ is independent of the choice of ψ. Moreover, $\langle u, \psi R_a \phi \rangle = \phi$ for each $\phi \in C_c^\infty(\mathbf{R}^d - \{0\})$. The continuity of the map $u \mapsto E(u)$ is not too difficult to see, completing the proof. ∎

When a is an integer smaller than or equal to $-n$, one can still use the construction in the proof to define an operator $E(u)$, but then it can depend on ψ, may fail to be homogeneous near the origin, and may fail to satisfy the identities $E(Pu) = PE(u)$ and $E(\partial_i u) = \partial_i E(u)$ (See Hörmander, 3.2).

9.4 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 9.14. *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. $\square$

In the language of commutative algebra, the association of $\mathcal{D}^*(U)$ to each open subset U of Ω gives the structure of a sheaf of modules on 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. This agrees with the sheaf theoretic definition.

Theorem 9.15. *If a distribution $\Lambda \in \mathcal{D}^*(\Omega)$ has compact support, then Λ has some finite order n, and extends uniquely to a continuous linear functional on $C^n(\Omega)$.*

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 some compact set K, then there is n such that for any $\phi \in C_c^\infty(K)$,

$$|\Lambda(\phi)| \lesssim \|\phi\|_{C^n(K)},$$

and so for any other compact set K,

$$|\Lambda(\phi)| = |\Lambda(\phi\psi)| \lesssim \|\phi\psi\|_{C^n(K)} \lesssim \|\psi\|_{C^n(K)} \|\phi\|_{C^n(K)}.$$

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 $C^\infty(X)$. $\square$

If $\mathcal{E}(\Omega)$ denotes $C^\infty(\Omega)$, equipped with the topology such that $f_n \to f$ if $D^\alpha f_n$ converges to $D^\alpha f$ locally uniformly for all α, then the last theorem implies the family of compactly supported distributions embeds itself in $\mathcal{E}(\Omega)^*$. Conversely, *every* element of $\mathcal{E}(\Omega)^*$ is a compactly supported distribution. Indeed, since $\mathcal{E}(\Omega)$ is a Frechét space, if Λ is a continuous linear functional on $\mathcal{E}(\Omega)$, then there exists a compact set K and some $n > 0$ such that

$$|\Lambda(\phi)| \lesssim \|\phi\|_{C^n(K)}.$$

It follows from this that Λ is a distribution with support contained in K.

Remark. For general compact sets K, it is *not* true that if Λ is a distribution supported on a set K, then there exists $n > 0$ such that

$$|\Lambda(\phi)| \lesssim \|\phi\|_{C^n(K)}.$$

Suppose K is not the union of finitely many compact connected sets. Then we can find a family of disjoint compact sets $\{K_i\}$ in K such that $K - (K_1 \cup \cdots \cup K_n)$ is compact for any $n > 0$. Fix $x_i \in K_i$, let $x \in K$ be a limit point of this sequence, consider a sequence of numbers $\{a_i\}$ such that $\sum a_i |x_i - x| = 1$, and $\sum a_i = \infty$, and let Λ be the distribution

$$\Lambda(\phi) = \sum_i a_i (\phi(x_i) - \phi(x)).$$

Then

$$|\Lambda(\phi)| \leqslant \|\phi'\|_{L^\infty(\mathbf{R}^d)},$$

so Λ is a distribution of order at most 1. On the other hand, if we choose a function $\phi \in \mathcal{D}(\Omega)$ which is equal to one on a neighborhood of $K_1 \cup \cdots \cup K_n$, and zero on a neighborhood of $K - (K_1 \cup \cdots \cup K_n)$, then

$$\Lambda(\phi) = \sum_{i=1}^n a_i,$$

so we cannot have a bound of the form

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant k} \|D^\alpha \phi\|_{L^\infty(K)} \lesssim 1.$$

On the other hand, for any precompact neighborhood U of K, we have a bound

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant 1} \|D^\alpha \phi\|_{L^\infty(U)}.$$

which is almost as good as the bound above.

If Λ is a distribution of order k supported on K, though we do not have a uniform bound $|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant k} \|D^\alpha \phi\|_{L^\infty(K)}$, if the right hand side vanishes, so does the left hand side.

Lemma 9.16. *Suppose Λ is a distribution of order k supported on K, and $\phi \in C^k(\Omega)$ satisfies $D^\alpha \phi(x) = 0$ for all $|\alpha| \leqslant k$ and $x \in K$, then $\Lambda(\phi) = 0$.*

Proof. By a density argument, we may assume that $\phi \in C^\infty(\Omega)$ without loss of generality. Find $\chi_\varepsilon \in \mathcal{D}(\Omega)$ such that $\chi_\varepsilon(x) = 1$ for $x \in K$, $\chi_\varepsilon(x) = 0$ if $d(x, K) \geqslant \varepsilon$, and $\|D^\alpha \chi_\varepsilon\|_{L^\infty} \lesssim \varepsilon^{-|\alpha|}$ for all $|\alpha| \leqslant k$. Then for any $\phi \in \mathcal{D}(\Omega)$,

$$|\Lambda(\phi)| = |\Lambda(\phi \chi_\varepsilon)| \lesssim \sum_{|\alpha| \leqslant k} \|D^\alpha (\phi \chi_\varepsilon)\|_{L^\infty} \lesssim \sum_{|\alpha| \leqslant k} \varepsilon^{|\alpha| - k} \|D^\alpha \phi\|_{L^\infty(K_\varepsilon)}.$$

For any $y \in K_\varepsilon$, we can pick $x \in K$ such that $|x - y| \leqslant \varepsilon$. Taylor's formula at x, together with the fact that all the derivatives of ϕ up to order k vanish at x, implies that

$$|(D^\alpha \phi)(y)| \lesssim \varepsilon^{k+1-|\alpha|}.$$

Thus we conclude that $|\Lambda(\phi)| \lesssim \varepsilon$, and we can then take $\varepsilon \to 0$. □

The last lemma implies that the value of *any distribution* Λ of order k supported on a point x_0 depends solely on the values $D^\alpha \phi(x_0)$ for $|\alpha| \leqslant k$. Thus there exists constants λ_α such that

$$\Lambda(\phi) = \sum_{|\alpha| \leqslant k} \lambda_\alpha D^\alpha \phi(x_0).$$

If we work harder, we can obtain a similar process for more general supports.

Using the Whitney extension theorem as a black box, we can also obtain better bounds for distributions supported on suitably regular compact sets.

Theorem 9.17 (Whitney). *Let K be a compact set in $\mathbf{R}^d$, and for each $|\alpha| \leqslant k$, a function $u_\alpha \in C(K)$. If*

$$U_\alpha(x,y) = \sum_{|\alpha| \leqslant k} \sup_{x,y \in K} \left| u_\alpha(x) - \sum_{|\beta| \leqslant k - |\alpha|} u_{\alpha+\beta}(y) \cdot (x-y)^\beta / \beta! \right| \cdot |x-y|^{|\alpha|-k}$$

for $x \neq y$, and $U_\alpha(x,x) = 0$, then provided U_α is continuous on $K \times K$, we can find a function $v \in C^k(\mathbf{R}^d)$ such that $D^\alpha v = u_\alpha$ on K for $|\alpha| \leqslant k$, and

$$\sum_{|\alpha| \leqslant k} \|D^\alpha v\|_{L^\infty} \lesssim \sum_{|\alpha| \leqslant k} \|U_\alpha\|_{L^\infty(K \times K)} + \sum_{|\alpha| \leqslant k} \|u_\alpha\|_{L^\infty(K)}.$$

A consequence is the following strengthening of the last lemma.

Lemma 9.18. *For* any *compact set K, and any distribution Λ of order k supported on K, we have*

$$\begin{aligned} |\Lambda(\phi)| \lesssim &\sum_{|\alpha| \leqslant k} \sup_{x,y \in K} \left| D^\alpha \phi(x) - \sum_{|\beta| \leqslant k - |\alpha|} D^{\alpha+\beta} \phi(y) \cdot (x-y)^\beta / \beta! \right| \cdot |x-y|^{|\alpha|-k} \\ &+ \sum_{|\alpha| \leqslant k} \|D^\alpha \phi\|_{L^\infty(K)}. \end{aligned}$$

Proof. To do this, we apply the Whitney extension theorem, setting $u_\alpha = D^\alpha \phi|_K$. We then apply the Whitney extension theorem to find $\psi \in C^k(\mathbf{R}^n)$ extending u_α with the required bounds above. Then $D^\alpha(\phi - \psi) = 0$ on K for all $|\alpha| \leqslant K$, from which it follows that $\Lambda(\phi) = \Lambda(\psi)$. The bound

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant k} \|D^\alpha \psi\|_{L^\infty},$$

which gives the required bound above. □

Recall that a compact set K is *Whitney regular*, which means that K is a finite union of compact, connected components, and for any two points $x, y \in K$ contained in a common component, there exists a rectifiable curve γ from x to y with length $O(|x - y|)$.

Lemma 9.19. *If K is Whitney regular, then for any distribution Λ supported on K, there exists k such that we have a bound*

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant k} \|D^\alpha \phi\|_{L^\infty(K)}.$$

Proof. Fix a rectifiable unit velocity curve $\gamma : [0, L] \to K$ between two points x and y in K, and let

$$F_\alpha(s) = D^\alpha \phi(\gamma(s)) - \sum_{|\beta| \leqslant k - |\alpha|} D^{\alpha+\beta} \phi(y)(\gamma(s) - y)^\beta / \beta!$$

Then $|F_\alpha(s)| \lesssim s^{k-|\alpha|} \sum_{|\beta|=k} \|D^\beta \phi\|_{L^\infty(K)}$. This is immediate if $|\alpha| = k$. For $|\alpha| < k$ we prove this result by induction, noting that the case for higher order k implies that

$$\begin{aligned} |dF_\alpha/ds| &\leqslant \sum_{i=1}^d \left| \left(D^{\alpha+i} \phi(\gamma(s)) - \sum_{|\beta| \leqslant k - |\alpha|} D^{(\alpha+i)+(\beta-i)} \phi(y)(\gamma(s) - y)^{\beta-i} / (\beta - 1)! \right) \cdot \gamma_i'(s) \right| \\ &\lesssim s^{k-|\alpha|-1} \sum_{|\beta|=k} \|D^\beta \phi\|_{L^\infty(K)} \end{aligned}$$

Integrating this inequality in s together with the fact that $F_\alpha(0) = 0$ gives the higher order bound. But this means that

$$\left|D^\alpha \phi(\gamma(s)) - \sum_{|\beta| \leqslant k-|\alpha|} D^{\alpha+\beta}\phi(y)(\gamma(s)-y)^\beta/\beta!\right| = |F_\alpha(L)|$$

$$\lesssim L^{k-|\alpha|} \sum_{|\beta|=k} \|D^\beta \phi\|_{L^\infty(K)}.$$

Choosing γ optimally gives

$$\left|D^\alpha \phi(\gamma(s)) - \sum_{|\beta| \leqslant k-|\alpha|} D^{\alpha+\beta}\phi(y)(\gamma(s)-y)^\beta/\beta!\right| \lesssim |x-y|^{k-|\alpha|} \sum_{|\beta|=k} \|D^\beta \phi\|_{L^\infty(K)}.$$

The last Lemma, together with this bound, completes the proof. □

Remark. Similar arguments can be used to show that if Λ is a distribution of order k supported on a compact set K, and there exists $\gamma \leqslant 1$ such that K is a finite union of connected components, such that for any pair of points x, y in that component, there exists a rectifiable path from x to y with length $O(|x-y|^\gamma)$, and $m \geqslant k/\gamma$, then

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant m} \|D^\alpha \phi\|_{L^\infty(K)}.$$

Let us finish by considering a consequence of these results, applied to distributions supported on hyperplanes. For simplicity in notation, we assume this hyperplane is axis oriented.

Theorem 9.20. *Let $x = (x_0, x_1)$, where $x_0 \in \mathbf{R}^{d_1}$, $x_1 \in \mathbf{R}^{d_2}$, and $d = d_1 + d_2$. Let $H = \{(x_0, x_1) \in \mathbf{R}^d : x_1 = 0\}$. If Λ is a distribution of order k compactly supported on H, then there exists distributions Λ_α of order $k - |\alpha|$ on $\mathbf{R}^{d_1}$ for each $|\alpha| \leqslant k$, where α is a multi-index in the $\mathbf{R}^{d_2}$ variables, and constants γ_α such that*

$$\Lambda(\phi) = \sum \Lambda_\alpha (D^\alpha \phi|_H).$$

Proof. Fix a function $\psi \in C_c^\infty(\mathbf{R}^{d_1})$ equal to one in a neighborhood of the origin. Given $\phi \in \mathcal{D}(\mathbf{R}^{d_1})$, all derivatives of the function

$$\sum_{|\alpha| \leqslant k} D^\alpha \phi(x_0, 0) \cdot (x_1^\alpha/\alpha!) \cdot \psi(x_1) = \sum_{|\alpha| \leqslant k} D^\alpha \phi|_H(x_0) \cdot (x_1^\alpha/\alpha!)$$

agree with ϕ on H, where α ranges over all derivatives in the x_1 direction. It follows that if we define a distribution Λ_α on $\mathbf{R}^{d_1}$ such that for $\psi \in \mathcal{D}(\mathbf{R}^{d_1})$,

$$\Lambda_\alpha(\psi) = \Lambda(\psi \otimes (x_1^\alpha/\alpha!)),$$

then

$$\Lambda(\phi) = \sum_{|\alpha| \leqslant k} \Lambda_\alpha(D^\alpha \phi|_H).$$

The hard part is showing that Λ_α has order $k - |\alpha|$. If the support of Λ in the x_0 variable is contained in a compact ball B, then, because B is Whitney regular,

$$\begin{aligned} |\Lambda_\alpha(\psi)| &\lesssim \sum_{|\beta_1|+|\beta_2| \leqslant k} \|D^{\beta_1+\beta_2} \{\psi \otimes (x_1^\alpha/\alpha!)\}\|_{L^\infty(B\times\{0\})} \\ &= \sum_{|\beta| \leqslant k-|\alpha|} \|D^\beta \psi\|_{L^\infty(B)}. \end{aligned}$$

$\square$

Remark. This argument does not really need the power of the full extension theorem machinery, since the Whitney extension theorem is relatively trivial in the application we give (we can consider a simple convolution argument to extend a function on a hyperplane to the full space). But the more developed machinery can be applied to characterize distributions on more general sets, which we leave to the reader to experiment with.

9.5 Distributional Solutions to ODEs

Distribution theory was originally invented to provide a more amenable setting to the theory of existence for linear partial differential equations. Let us build up some basic results for solving *ordinary differential equations* using distributions

Theorem 9.21. *If $u \in \mathcal{D}^*(\mathbf{R}^d)$ and there exists an index i such that $D^i u = 0$, then there exists $v \in \mathcal{D}^*(\mathbf{R}^{d-1})$ such that*

$$\int_{\mathbf{R}^d} u(x)\phi(x)\,dx = \int_{\mathbf{R}^{d-1}} v(x) \left(\int_{-\infty}^{\infty} \phi(x)\,dx^i \right)\,dx,$$

i.e. u is 'constant' in the direction i. In particular, if $d = 1$, and $Du = 0$, then $u = C$ for some constant C.

Proof. Suppose without loss of generality that $i = d$. Suppose $\phi \in \mathcal{D}(\mathbf{R}^d)$ and for each $x \in \mathbf{R}^{d-1}$,

$$\int_{-\infty}^{\infty} \phi(x,t)\,dt = 0.$$

Then the function

$$\psi(x,t) = \int_{-\infty}^{t} \phi(x,s)\,ds = 0$$

has compact support and $D^i\psi = \phi$. Thus

$$\begin{aligned}\int_{-\infty}^{\infty} u(x,t)\phi(x,t)\,dx\,dt &= \int_{-\infty}^{\infty} u(x,t)D^i\psi(x,t)\,dx\,dt\\ &= -\int_{-\infty}^{\infty} D^i u(x,t)\psi(x,t)\,dx\,dt = 0.\end{aligned}$$

Now fix $\phi_0 \in \mathcal{D}(\mathbf{R})$ with $\int_{-\infty}^{\infty} \phi_0(x) = 1$. Then given any $\phi \in \mathcal{D}(\mathbf{R}^d)$,

$$\int_{-\infty}^{\infty} u(x,t)\phi(x,t)\,dx\,dt = \int_{-\infty}^{\infty} u(x,t)\phi_0(t)\left(\int_{-\infty}^{\infty} \phi(x,s)\,ds\right)\,dx\,dt.$$

Thus it suffices to set

$$v(x) = \int_{-\infty}^{\infty} u(x,t)\phi_0(t)\,dt. \qquad \square$$

A change of variables shows we have found all distributional solutions to the transport equation $w \cdot \nabla u = 0$ for a fixed vector $w \in \mathbf{R}^d$. A corollary is a regularity result for distributional solutions to the equation $w \cdot \nabla u(x) + a(x)u(x) = f(x)$.

Lemma 9.22. *If $u \in \mathcal{D}(\mathbf{R})^*$ is a distributional solution to the equation $Du + a \cdot u = f(x)$, where $f \in C(\mathbf{R})$, and $a \in C^\infty(\mathbf{R})$, then $u \in C^1(\mathbf{R})$, and so u is a classical solution to the equation.*

Proof. We just apply classical techniques distributionally. First assume $a = 0$. If F is an antiderivative of f in the ith direction, then $F \in C^1(\mathbf{R})$, and $D(u - F) = 0$, so u differs from F by a constant, and is therefore also in $C^1(\mathbf{R})$. For $a \neq 0$, let A be an antiderivative of a, and set

$$E(x) = e^{A(x)}.$$

Then $E \in C^\infty(\mathbf{R})$. Thus if u is a distribution solving $Du + a \cdot u = f(x)$, and if $v = E \cdot u$ then the product rule shows that

$$D(v) = E \cdot f$$

The $a = 0$ case implies that $v \in C^1(\mathbf{R})$, and so $u \in C^1(\mathbf{R})$. □

Remark. The idea of this result generalizes to a system of differential equations given by a matrix a with C^∞ entries, and where f is a vector with continuous entries, by finding an invertible matrix $E(x)$ such that $E'(x) = E(x) \cdot a(x)$. In particular, since higher order ordinary differential equations can be reduced to one dimensional systems of ordinary differential equations, we conclude that any distributional solution to a linear ordinary differential equation of the form

$$D^m u + a_{m-1} D^{m-1} u + \dots + au = f,$$

for $a_i \in C^\infty(\mathbf{R})$, and $f \in C(\mathbf{R})$, lies in $C^m(\mathbf{R})$, and is a classical solution to the equation.

Higher dimensional analogues of these results are not as strong. Indeed, we have already seen that distributional solutions to $D^i u = 0$ are not necessarily classical solutions. On the other hand, we can 'almost' upgrade continuous distributional solutions to this equation

Lemma 9.23. *Suppose $u, f \in C(\mathbf{R}^d)$, and u as a distribution satisfies $D^i u = f$. Then $D^i u(x)$ exists in a classical sense for all $x \in \mathbf{R}^d$, and $D^i u(x) = f(x)$ for all x.*

Proof. Assume $i = d$ without loss of generality, and write $x = (x_0, t)$, for $x_0 \in \mathbf{R}^{d-1}$ and $t \in \mathbf{R}$. Set

$$v(x_0, t) = \int_0^t f(x_0, s)\, ds.$$

Then v is a distributional solution to the equation $D^i v = f$, and so $D^i(u - v) = 0$. It follows that there exists a distribution $w \in \mathcal{D}(\mathbf{R}^{d-1})^*$ such that

$u(x,t) - v(x,t) = w(x)$. The proof of the existence of w actually implies that w is continuous, since u and v are continuous. But then $u(x,t) = v(x,t)+w(x)$ is differentiable in the t-variable by the fundamental theorem of calculus. $\square$

9.6 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 9.24. *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 9.25. *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 9.26. *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.*

9.7 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 9.27. *$\Lambda * \phi$ 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 $k \in \{1, \dots, d\}$ and $h \in \mathbf{R}$, we set

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

then $\Delta_h\phi$ converges to $D^k\phi$ in $\mathcal{D}(\mathbf{R}^d)$, and as such

$$\begin{aligned}\Delta_h(\Lambda * \phi)(x) &= \frac{(\Lambda * \phi)(x+he_k) - (\Lambda * \phi)(x)}{h}\\ &= \Lambda\left(\frac{T_{-x-he_k}\phi^* - T_{-x}\phi^*}{h}\right)\end{aligned}$$

As $h \to 0$, in $\mathcal{D}(\mathbf{R}^d)$ we have

$$\frac{T_{-x-he_k}\phi^* - T_{-x}\phi^*}{h} \to -T_{-x}D_k\phi^* = T_{-x}(D_k\phi)^*.$$

Thus, by continuity,

$$\lim_{h\to 0} \Delta_h(\Lambda * \phi)(x) = \Lambda(T_{-x}(D_k\phi)^*) = (\Lambda * D_k\phi)(x)$$

Iteration gives the general result that $\Lambda * \phi \in C^\infty(\mathbf{R}^d)$. An easy calculation then shows that for each $x \in \mathbf{R}^d$,

$$\begin{aligned}
[(D^\alpha\Lambda) * \phi](x) &= (D^\alpha\Lambda)(T_{-x}\phi^*) \\
&= (-1)^{|\alpha|}\Lambda(T_{-x}D^\alpha\phi^*) \\
&= \Lambda(T_{-x}(D^\alpha\phi)^*) \\
&= (\Lambda * D^\alpha\phi)(x).
\end{aligned}$$

□

There is a certain duality going on here. Distributions can be viewed as linear functionals on $\mathcal{D}(\mathbf{R}^d)$, but one can also view them as a certain family of linear operators from $\mathcal{D}(\mathbf{R}^d) \to C^\infty(\mathbf{R}^d)$, and the convolution operator uniquely represents the distribuition. In fact, any such operator that is translation invariant and continuous can be represented as convolution by a distribution.

Theorem 9.28. *Let $T : \mathcal{D}(\mathbf{R}^d) \to C^\infty(\mathbf{R}^d)$ be a translation invariant continuous operator. Then there exists a distribution Λ such that $T\phi = \Lambda * \phi$ for all $\phi \in \mathcal{D}(\mathbf{R}^d)$.*

Proof. If we knew $T\phi = \Lambda * \phi$ for some Λ, then we could recover Λ since

$$\int \Lambda(x)\phi(x)\,dx = T\tilde{\phi}(0).$$

Since T is a continuous operator, the right hand side defines a distribution Λ, and translation invariance allows us to conclude that $T\phi = \Lambda * \phi$ for all $\phi \in \mathcal{D}(\mathbf{R}^d)$. □

Example. *A linear differential operator $P : \mathcal{D}(\mathbf{R}^d) \to \mathcal{D}(\mathbf{R}^d)$ is translation invariant, from which it follows that there exists a distribution Λ such that $P\phi = \Lambda * \phi$. Of course, $\Lambda(\phi) = P\phi(0)$ is just given by applying the differential operator at the origin.*

For more general operators that are translation invariant, we cannot represent all operators via convolution by distributions. A significantly more general family of operators can be found if, instead of considering operators of the form

$$T\phi(y) = \int \Lambda(y-x)\phi(x)\,dx$$

we instead study *kernel* operators

$$T\phi(y) = \int K(x,y)\phi(x)\,dx$$

where K is a distribution on $\Omega_1 \times \Omega_2$ and $\phi \in \mathcal{D}(\Omega_1)$. To formally interpret the output of this operator, we need to test it against another bump function, i.e. for $\psi \in \mathcal{D}(\Omega_2)$ we consider

$$\int T\phi(y)\psi(y)\,dy = \int K(x,y)\phi(x)\psi(y)\,dx\,dy,$$

which is K tested against $\phi \otimes \psi$. Thus $T\phi$ is naturally a distribution on $\mathbf{R}^m$, and this definition naturally gives a continuous map from $\mathcal{D}(\mathbf{R}^n)$ to $\mathcal{D}^*(\mathbf{R}^m)$. In 1953, Schwartz showed that essentially every linear operator encountered in Euclidean analysis is of this form, which explains the prevalence of kernel operators in analysis.

Theorem 9.29. *Let $T : \mathcal{D}(\Omega_1) \to \mathcal{D}^*(\Omega_2)$ be a continuous linear operator. Then there exists a unique distribution $K \in \mathcal{D}^*(\Omega_1 \times \Omega_2$ such that for $\phi \in \mathcal{D}(\mathbf{R}^n)$ and $\psi \in \mathcal{D}(\mathbf{R}^m)$,*

$$\int T\phi(y)\psi(y)\,dy = \int K(x,y)\phi(x)\psi(y)\,dx\,dy.$$

Looking at the properties of kernels defining an operator is often a useful technique to gain insight in how an operator behaves, since one can study the singularities and smoothness of the kernel separately from the operator itself.

Theorem 9.30. *If $\phi, \psi \in \mathcal{D}(\mathbf{R}^n)$, then $\Lambda * (\phi * \psi) = (\Lambda * \phi) * \psi$.*

Proof. Let K be a compact set containing the supports of ϕ and ψ. It is simple to verify that for each $x \in \mathbf{R}^d$,

$$(\phi * \psi)^*(x) = \int \phi^*(x+y)\psi(y)\,dy = \int (T_y\phi^*)(x)\psi(y)\,dy$$

since the map $y \mapsto (T_y \phi)^* \psi(y)$ is continuous, and vanishes out of the compact set K, we can consider the $C_c^\infty(K)$ 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 $\{\phi_\alpha\}$ is known as an *approximate identity* in the space of distributions if $\Lambda * \phi_\alpha \to \Lambda$ weakly as $\alpha \to \infty$, for every distribution Λ, and an approximate identity in the space of test functions if $\psi * \phi_\alpha \to \psi$ in $\mathcal{D}(\mathbf{R}^n)$.

Theorem 9.31. *If ϕ_α is a family of non-negative functions in $\mathcal{D}(\mathbf{R}^n)$ 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 $f \in \mathcal{D}(\mathbf{R}^n)$, 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 $\mathcal{D}(\mathbf{R}^n)$. 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 $\mathcal{D}(\mathbf{R}^n)$ 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 that for each $x \in \mathbf{R}^d$,

$$f(x) = \lim (\Lambda * \phi_n)(x) = \lim \Lambda(T_x \phi_n^*) = \Lambda(\lim T_x \phi_n^*) = \Lambda(T_x \phi^*) = (\Lambda * \phi)(x).$$

Here we have used the fact that $T_x \phi_n^*$ converges to $T_x \phi^*$ in $\mathcal{D}(\mathbf{R}^n)$. Suprisingly, the converse is also true.

Theorem 9.32. *If $L : \mathcal{D}(\mathbf{R}^n) \to C^\infty(\mathbf{R}^n)$ 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 for an arbitrary operator L. Indeed, it then follows that 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 9.33. *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 9.34. *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. $\square$

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 9.35. *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 9.36. *If Λ and Ψ are distributions, one of which having compact support, 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. □

9.8 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 closed under multiplication by polynomials, and closed under applications of arbitrary constant-coefficient differential operators. A natural choice is then the family of functions which *decays rapidly*, as well as all of it's derivatives; i.e. we let $\mathcal{S}(\mathbf{R}^d)$ be the space of all functions $f \in C^\infty(\mathbf{R}^d)$ such that for any integer n and multi-index α, $|x|^n D^\alpha f \in L^\infty(\mathbf{R}^d)$. The space $\mathcal{S}(\mathbf{R}^d)$ is then locally convex if we consider the family of seminorms

$$\|f\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)} = \|\langle x\rangle^m \nabla^n f\|_{L^\infty(\mathbf{R}^d)}.$$

Elements of $\mathcal{S}(\mathbf{R}^d)$ are known as *Schwartz functions*, and $\mathcal{S}(\mathbf{R}^d)$ is often 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 $\|f\|_{\mathcal{S}^{\alpha,\beta}(\mathbf{R}^d)} = \|x^\alpha D^\beta f\|_{L^\infty(\mathbf{R}^d)}$, because x^α often behaves more nicely under various Fourier analytic operations. It is obvious that $\mathcal{S}(\mathbf{R}^d)$ is separated by the seminorms defined on it, because $\|\cdot\|_{L^\infty(\mathbf{R}^d)} = \|\cdot\|_{\mathcal{S}^{0,0}(\mathbf{R}^d)}$ is a norm used to define the space. We now show the choice of seminorms make the space complete.

Theorem 9.37. *$\mathcal{S}(\mathbf{R}^d)$ is a complete metric space.*

Proof. Let $\{f_i\}$ be a Cauchy sequence with respect to the seminorms $\|\cdot\|_{\mathcal{S}^{n,\alpha}(\mathbf{R}^d)}$. This implies that for each integer m, and multi-index α, the sequence of functions $\langle x\rangle^m D^\alpha f_k$ is Cauchy in $L^\infty(\mathbf{R}^d)$. Since $L^\infty(\mathbf{R}^d)$ is complete, there are functions $g_{m,\alpha}$ such that $\langle x\rangle^m D^\alpha f_k$ 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 $\langle x\rangle^m D^\alpha f = g_{m,\alpha}$. It is then easy to show that f_i converges to f in $\mathcal{S}(\mathbf{R}^d)$. ∎

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 $D^\alpha \phi = P_\alpha \phi$; this can be established by a simple induction. But this means that for each fixed α, $|P_\alpha(x)| \lesssim_\alpha 1 + |x|^{|\alpha|}$. Since $e^{-|x|^2} \lesssim_{m,\alpha} \langle x\rangle^{-m-|\alpha|}$ for any fixed m and α, we find that for any $x \in \mathbf{R}^d$,*

$$|(1 + |x|^m) D^\alpha \phi| \lesssim_{\alpha,m} 1.$$

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_0 and m_0, there is n_1, m_1 such that

$$\|Tf\|_{\mathcal{S}^{n_0,m_0}(\mathbf{R}^d)} \lesssim_{n_0,m_0} \|f\|_{\mathcal{S}^{n_1,m_1}(\mathbf{R}^d)}.$$

For a functional $\Lambda : \mathcal{S}(\mathbf{R}^d) \to \mathbf{R}$, it suffices to show that there exists n and m such that $|\Lambda f| \lesssim \|f\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)}$. The minimal such choice of n is known as the *order* of the functional 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 functions. We really just care about proving the existence of such a constant.

Lemma 9.38. *The map $(f,g) \mapsto fg$ for $f,g \in \mathcal{S}(\mathbf{R}^d)$ gives a bounded bilinear map from $\mathcal{S}(\mathbf{R}^d) \times \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$.*

Proof. A simple application of the Leibnitz formula shows that for any multi-index α with $|\alpha| = m$, and two non-negative integers n_1 and n_2 with $n_1 + n_2 = n$,

$$\|fg\|_{\mathcal{S}^{n,\alpha}(\mathbf{R}^d)} \lesssim_n \|f\|_{\mathcal{S}^{n_1,m}(\mathbf{R}^d)} \|g\|_{\mathcal{S}^{n_2,m}(\mathbf{R}^d)}.$$

More generally, this argument shows that the analogoue bilinear map from $C^\infty(\mathbf{R}^d) \times \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ is bounded. □

Theorem 9.39. *The following 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 from $\mathcal{S}(\mathbf{R}^d)$ to $\mathcal{S}(\mathbf{R}^d)$.*

Furthermore, the Fourier transform is an isomorphism of $\mathcal{S}(\mathbf{R}^d)$.

Proof. We leave all but the last point as exercises. Here it will be convenient to use the norms $\|\cdot\|_{\mathcal{S}^{\alpha,\beta}(\mathbf{R}^d)}$ as well as the norms $\|\cdot\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)}$. If $|\alpha| \leqslant m$, $|\beta| \leqslant n$, then we can use the Leibnitz formula to conclude that

$$\begin{aligned}
\|\xi^\alpha D^\beta \mathcal{F}(f)\|_{L^\infty(\mathbf{R}^d)} &\lesssim_{\alpha,\beta} \|\mathcal{F}(D^\alpha(x^\beta f))\|_{L^\infty(\mathbf{R}^d)} \\
&\lesssim_{\alpha,\beta} \max_{\gamma \leqslant \alpha \wedge \beta} \|\mathcal{F}(x^\gamma D^\gamma f)\|_{L^\infty(\mathbf{R}^d)} \\
&\leqslant \max_{\gamma \leqslant \alpha \wedge \beta} \|x^\gamma D^\gamma f\|_{L^1(\mathbf{R}^d)} \\
&\lesssim \|f\|_{\mathcal{S}^{\gamma,|\gamma|+d+1}(\mathbf{R}^d)}.
\end{aligned}$$

Thus $\mathcal{F}$ is a bounded linear operator on $\mathcal{S}(\mathbf{R}^d)$. Since all Schwartz functions are arbitrarily smooth, the Fourier inversion formula applies to all Schwartz functions, and so $\mathcal{F}$ is a bijective bounded linear operator with inverse $\mathcal{F}^{-1}$. The open mapping theorem then immediately implies that $\mathcal{F}^{-1}$ is bounded. $\square$

Corollary 9.40. *If f and g are Schwartz, then $f * g$ is Schwartz.*

Proof. Since
$$f * g = \mathcal{F}^{-1}(\mathcal{F}(f)\mathcal{F}(g)),$$
the result follows from the previous results we have shown. $\square$

Now we get to the interesting part of the theory. We have defined a homeomorphic linear transform from $\mathcal{S}(\mathbf{R}^d)$ to itself. The theory of functional analysis then says that we can define a dual map, which is a homeomorphism from the dual space $\mathcal{S}(\mathbf{R}^d)^*$ to itself. Note the inclusion map $\mathcal{D}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ is continuous, and $\mathcal{D}(\mathbf{R}^d)$ 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 $\mathcal{D}(\mathbf{R}^d)$ which have a continuous extension to $\mathcal{S}(\mathbf{R}^d)$. Intuitively, this corresponds to an asymptotic decay condition.

Example. *Recall that for any $f \in L^1_{loc}(\mathbf{R}^d)$, we can consider the distribution $\Lambda[f]$ defined by setting*

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

However, 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 for some n,

$$\int_{\mathbf{R}^d} \frac{d|\mu|(x)}{1+|x|^n}\, dx < \infty$$

then μ is known as a tempered measure*, and acts as a tempered distribution since*

$$\begin{aligned} |\Lambda[\mu](\phi)| &\leqslant \int_{\mathbf{R}^d} |\phi(x)|\, d|\mu|(x) \\ &\leqslant \left(\int_{\mathbf{R}^d} \frac{d|\mu|(x)}{1+|x|^n}\, dx \right) \cdot \|\phi\|_{\mathcal{S}^{0,n}(\mathbf{R}^d)}. \end{aligned}$$

Example. *Any compactly supported distribution is tempered. Indeed, if Λ is a distribution supported on a compact set K, then it has finite order n for some integer n, and extends to an operator on $C^\infty(\mathbf{R}^d)$. We then find*

$$|\Lambda(\phi)| \lesssim \|\phi\|_{C^n(\mathbf{R}^d)} \leqslant \|\phi\|_{\mathcal{S}^{0,n}(\mathbf{R}^d)}.$$

Example. *The distribution Λ on $\mathbf{R}$ given by*

$$\Lambda(\phi) = p.v. \int_{-\infty}^{\infty} \frac{\phi(x)}{x}\, dx$$

is tempered, since

$$\int_{|x| \geqslant 1} \frac{\phi(x)}{x} \lesssim \|\phi\|_{\mathcal{S}^{1,0}(\mathbf{R}^d)}$$

and

$$p.v. \int_{-\infty}^{\infty} \frac{\phi(x)}{x}\, dx \lesssim \|\phi\|_{C^1(\mathbf{R}^d)} = \|\phi\|_{\mathcal{S}^{0,1}(\mathbf{R}^d)}$$

and so Λ is tempered of order 1. The p.v. is called the principal value *of $1/x$.*

Using the same techniques as for distributions, the derivative $D^\alpha \Lambda$ of a tempered distribution Λ is tempered, as is $\phi\Lambda$, whenever ϕ is a Schwartz function, or $f\Lambda$, where f is a polynomial. Of course, we can multiply by polynomially increasing smooth functions as well.

Let us now apply the distributional method to define the Fourier transform of a tempered distribution. Recall that we heuristically think of Λ as formally corresponding to a regular function f such that

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

The multiplication formula

$$\int_{\mathbf{R}^d} \widehat{f}(\xi) g(\xi)\, d\xi = \int_{\mathbf{R}^d} f(x)\widehat{g}(x)\, dx$$

gives us the perfect opportunity to move the analytical operations on f to analytical operations on g. Thus if Λ is the distribution corresponding to a Schwartz $f \in \mathcal{S}(\mathbf{R}^d)$, the distribution $\widehat{\Lambda}$ corresponding to $\widehat{f}$, then for any Schwartz $\phi \in \mathcal{S}(\mathbf{R}^d)$,

$$\widehat{\Lambda}(\phi) = \Lambda(\widehat{g}).$$

In particular, this motivates us to define the Fourier transform of *any* tempered distribution Λ to be the unique tempered distribution $\widehat{\Lambda}$ such that the equation above holds for all Schwartz ϕ. This distribution exists because the Fourier transform is an isomorphism on the space of Schwartz functions. Clearly, the Fourier transform is a homeomorphism on the space of tempered distributions under the weak topology, and moreover, satisfies all the symmetry properties that the ordinary Fourier transform does, once we interpret scalar, rotation, translation, differentiation, etc, in a natural way on the space of distributions.

Example. *Consider the constant function* 1, *interpreted as a tempered distribution on* $\mathbf{R}^d$. *Then for any* $\phi \in \mathcal{S}(\mathbf{R}^d)$,

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

Thus for any $\phi \in \mathcal{S}(\mathbf{R}^d)$,

$$\widehat{1}\left(\widehat{\phi}\right) = 1(\phi) = \int \phi(\xi)\, d\xi = \widehat{\phi}(0).$$

Thus $\widehat{1}$ is the Dirac delta function at the origin. Similarily, the Fourier inversion formula implies that

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

so the Fourier transform of the Dirac delta function is the constant 1 function.

Example. *Let u denote a compactly supported distribution. We claim that $\widehat{u} \in C^\infty(\mathbf{R}^n)$ is a smooth function, such that*

$$\widehat{u}(\xi) = \langle u, e^{-2\pi i \xi \cdot x}\rangle.$$

Indeed, formally speaking,

$$\langle \widehat{u}, \phi\rangle = \langle u, \widehat{\phi}\rangle = \langle u, \int \phi(x) e^{-2\pi i \xi \cdot x}\, dx\rangle = \int \phi(x)\langle u, e^{-2\pi i \xi \cdot x}\rangle\, dx.$$

The proof that $\widehat{u}$ is smooth follows because we have control of the derivatives of $e^{-2\pi i \xi \cdot x}$ on the support of u.

Example. *The theory of tempered distributions enables us to take the Fourier transform of $f \in L^p(\mathbf{R}^d)$, when $p > 2$ or when $p < 1$. The introduction of distributions is in some sense, essential to this process, because for each $p \notin [1,2]$, there is $f \in L^p(\mathbf{R}^d)$ such that $\widehat{f}$ is* not *a locally integrable function. Otherwise, we could define an operator $T : L^p(\mathbf{R}^d) \to L^1(\mathbf{R}^d)$ given by*

$$Tf = \widehat{f}\,\mathbf{I}_{|\xi| \leqslant 1}.$$

If a sequence of functions $\{f_n\}$ converges to f in $L^p(\mathbf{R}^d)$, and Tf_n converges to g in $L^1(\mathbf{R}^d)$, then Tf_n converges distributionally to g, which implies $Tf = g$. The closed graph theorem thus implies that T is a continuous operator from $L^p(\mathbf{R}^d)$ to $L^1(\mathbf{R}^d)$, so there exists $M > 0$ such that

$$\int_{|\xi| \leqslant 1} |\widehat{f}(\xi)| \leqslant M \|f\|_{L^p(\mathbf{R}^d)}.$$

If $f_\alpha(x) = e^{-\pi\alpha|x|^2}$, then $\widehat{f_\alpha}(\xi) = \alpha^{-d/2} e^{-\pi|x|^2/\alpha}$. We have

$$\begin{aligned}\|f_\alpha\|_{L^p(\mathbf{R}^d)} &= \left(\int_{\mathbf{R}^d} e^{-\pi\alpha p|x|^2}\, dx\right)^{1/p}\\ &= (\alpha p)^{-d/2p}\left(\int_{\mathbf{R}^d} e^{-\pi|x|^2}\, dx\right)^{1/p} \lesssim_d (\alpha p)^{-1/2p}.\end{aligned}$$

On the other hand, for $|\xi| \leqslant 1$, $|\widehat{f_\alpha}(\xi)| \geqslant \alpha^{-d/2} e^{-\pi/\alpha}$, so

$$\int_{|\xi|\leqslant 1} |\widehat{f_\alpha}(\xi)| \gtrsim_d \alpha^{-d/2} e^{-\pi/\alpha}.$$

Thus we conclude that $\alpha^{-d/2} e^{-\pi/\alpha} \lesssim_d M(\alpha p)^{-d/2p}$, or equivalently,

$$\alpha^{d/2(1/p-1)} e^{-\pi/\alpha} \lesssim_d M p^{-d/2p}.$$

Taking $\alpha \to \infty$ gives a contradiction if $p < 1$. For $p > 2$, we give the Gaussian an oscillatory factor that does not affect the L^p norm but boosts the L^1 norm of the Fourier transform. We set

$$g_\delta(x) = \prod_{k=1}^d \frac{e^{-\pi x_k^2/(1+i\delta)}}{(1+i\delta)^{1/2}}.$$

The Fourier transform formula of the Gaussian, when applied using the theory of analytic continuation, shows that

$$\widehat{g_\delta}(\xi) = \prod_{k=1}^d e^{-\pi(1+i\delta)\xi_k^2}.$$

We have

$$\int_{|\xi|\leqslant 1} |\widehat{g_\delta}(\xi)| = \int_{|\xi|\leqslant 1} e^{-\pi|\xi|^2} \gtrsim 1.$$

On the other hand, for $\delta \geqslant 1$,

$$\begin{aligned}
\|g_\delta\|_{L^p(\mathbf{R}^d)} &= \left(\int |g_\delta(x)|^p \, dx \right)^{1/p} \\
&= |1+i\delta|^{-d/2} \left(\int_{-\infty}^{\infty} e^{-p\pi x^2/(1+\delta^2)} \, dx \right)^{d/p} \\
&\lesssim_d \delta^{-d/2} \delta^{d/p} p^{-d/p} = \delta^{d(1/p-1/2)} p^{-d/p}.
\end{aligned}$$

Thus we conclude $1 \lesssim_d M\delta^{d(1/p-1/2)} p^{d/p}$, which gives a contradiction as $\delta \to \infty$ if $p > 2$.

Example. *Consider the Riesz Kernel on $\mathbf{R}^d$, for each $\alpha \in \mathbf{C}$ with positive real part, as the function*

$$K_\alpha(x) = \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}}|x|^{-\alpha}.$$

Then for $0 < Re(\alpha) < d$, $\widehat{K_\alpha} = K_{d-\alpha}$. We recall that Γ is defined by the integral formula

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, ds,$$

where $Re(s) > 0$. We note that if $p = d/Re(\alpha)$, $K_\alpha \in L^{p,\infty}(\mathbf{R}^d)$. The Marcinkiewicz interpolation theorem implies that if $d/2 < Re(\alpha) < d$, then K_α can be decomposed as the sum of a $L^1(\mathbf{R}^d)$ function and a $L^2(\mathbf{R}^d)$ function, and so we can intepret the Fourier transform of $\widehat{K_\alpha}$ using techniques in $L^1(\mathbf{R}^d)$ and $L^2(\mathbf{R}^d)$, and moreover, the Marcinkiewicz interpolation theorem implies that

$$\|\widehat{K_\alpha}\|_{L^{q,\infty}(\mathbf{R}^d)} \leqslant \|K_\alpha\|_{L^{p,\infty}(\mathbf{R}^d)}.$$

where q is the dual of p. In particualr, the Fourier transform of K_α is a function. We note that K_α obeys multiple symmetries. First of all, K_α is radial, so $\widehat{K_\alpha}$ is also radial. Moreover, K_α is homogenous of degree $-\alpha$, i.e. for each $x \in \mathbf{R}^d$, $K_\alpha(\varepsilon x) = \varepsilon^{-\alpha} K_\alpha(x)$. This actually uniquely characterizes K_α among all locally integrable functions. Taking the Fourier transform of both sides of the equation for homogeneity, we find

$$\varepsilon^{-d}\widehat{K_\alpha}(\xi/\varepsilon) = \varepsilon^{-\alpha}\widehat{K_\alpha}(x).$$

Thus $\widehat{K_\alpha}$ is homogenous of degree $\alpha - d$. But this uniquely characterizes $\widehat{K_{d-\alpha}}$ out of any distribution, up to multiplicity, so we conclude that for $d/2 < Re(\alpha) < d$, that $\widehat{K_\alpha}$ is a scalar multiple of $K_{d-\alpha}$. But we know that by a change into polar coordinates, if A_d is the surface area of a unit sphere in $\mathbf{R}^d$, then

$$\begin{aligned}
\int_{\mathbf{R}^d} K_\alpha(x) e^{-\pi|x|^2}\, dx &= \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}} \int_{\mathbf{R}^d} |x|^{-\alpha} e^{-\pi|x|^2}\, dx \\
&= A_d \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}} \int_0^\infty r^{d-1-\alpha} e^{-\pi r^2}\, dr \\
&= A_d \frac{\Gamma(\alpha/2)}{2\pi^{d/2}} \int_0^\infty s^{(d-\alpha)/2-1} e^{-s}\, ds \\
&= A_d \frac{\Gamma(\alpha/2)\Gamma((d-\alpha)/2)}{\pi^{d/2}}.
\end{aligned}$$

But this is also the value of

$$\int_{\mathbf{R}^d} K_{d-\alpha}(x) e^{-\pi|x|^2},$$

so we conclude $\widehat{K_\alpha} = K_{d-\alpha}$ *if* $d/2 < Re(\alpha) < d$. *We could apply Fourier inversion to obtain the result for* $0 < Re(\alpha) < d/2$, *but to obtain the case* $Re(\alpha) = d/2$, *we must apply something different. For each* $s \in \mathbf{C}$ *with* $0 < Re(s) < d$, *and for each Schwartz* $\phi \in \mathcal{S}(\mathbf{R}^d)$ *we define*

$$A(s) = \int K_s(\xi)\widehat{\phi}(\xi)\, d\xi = \frac{\Gamma(s/2)}{\pi^{s/2}} \int |\xi|^{-s/2} \widehat{\phi}(\xi)\, d\xi.$$

and

$$B(s) = \int K_{d-s}(\xi)\widehat{\phi}(\xi)\, d\xi = \frac{\Gamma((d-s)/2)}{\pi^{(d-s)/2}} \int |\xi|^{(d-s)/2} \widehat{\phi}(\xi)\, d\xi.$$

The integrals above converge absolutely for $0 < Re(s) < d$, *and the dominated convergence theorem implies that* A *and* B *are both complex differentiable. Since* $A(s) = B(s)$ *for* $d/2 < Re(s) < d$, *analytic continuation implies* $A(s) = B(s)$ *for all* $0 < Re(s) < d$, *completing the proof. For* $Re(\alpha) \geqslant d$, K_α *is no longer locally integrable, and so we must interpret the distribution given by integration by* K_α *in terms of principal values. The fourier transform of these functions then becomes harder to define.*

Example. *Let us consider the complex Gaussian defined, for a given invertible symmetric matrix* $T : \mathbf{R}^d \to \mathbf{R}^d$, *as* $G_T(x) = e^{-i\pi(Tx\cdot x)}$. *Then*

$$\widehat{G_T} = e^{-i\pi\sigma/4} |\det(T)|^{-1/2} G_{-T^{-1}},$$

where σ *is the* signature *of* T, *i.e. the number of positive eigenvalues, minus the number of negative eigenvalues, counted up to multiplicity. Thus we need to show that for any Schwartz* $\phi \in \mathcal{S}(\mathbf{R}^d)$,

$$e^{-i\pi\sigma/4} |\det(T)|^{-1/2} \int_{\mathbf{R}^d} e^{i\pi(T^{-1}\xi\cdot\xi)} \widehat{\phi}(\xi)\, d\xi = \int_{\mathbf{R}^d} e^{-i\pi(Tx\cdot x)} \phi(x)\, dx.$$

Let us begin with the case $d = 1$, *in which case we also prove the theorem when* T *is a complex symmetric matrix. If* T *is given by multiplication by* $-iz$, *and if*

$\sqrt{\cdot}$ denotes the branch of the square root defined for all non-negative numbers and positive on the real-axis, then we note that when $z = \lambda i$,

$$e^{-i\pi\sigma/4}|\det(T)|^{-1/2} = e^{-i\pi sgn(\lambda)/4}|\lambda|^{-1/2} = \sqrt{z}.$$

Thus it suffices to prove the analytic family of identities

$$\int_{-\infty}^{\infty} e^{-(\pi/z)\xi^2}\widehat{\phi}(\xi)\,d\xi = \sqrt{z}\int_{-\infty}^{\infty} e^{-\pi z x^2}\phi(x)\,dx,$$

where both sides are well defined and analytic whenever z has positive real part. But we already know from the Fourier transform of the Gaussian that this identity holds whenever z is positive and real, and so the remaining identities follows by analytic continuation. We note that the higher dimensional identity is invariant under changes of coordinates in $SO(n)$. Thus it suffices to prove the remaining theorem when T is diagonal. But then everything tensorizes and reduces to the one dimensional case. More generally, if $T = T_0 - iT_1$ is a complex symmetric matrix, which is well defined if T_1 is positive semidefinite, then

$$\widehat{G_T} = \frac{1}{\sqrt{i\det(T)}}\cdot G_{-T^{-1}},$$

which follows from analytic continuation of the case for real T.

Thus we conclude that

$$|\langle\widehat{u},\phi\rangle| = |\langle u,\widehat{\phi}\rangle| \lesssim \|(1+|x|)^K\phi\|_{L^\infty(\mathbf{R}^d)}.$$

Thus $\widehat{u}$ is a distribution of order zero, and thus a measure. But $x^\alpha u$ is a compactly supported distribution for all α, which implies that $D^\alpha\widehat{u}$ is a distribution of order zero, and thus a measure.

Example. *We know $((-2\pi ix)^\alpha)^\wedge = ((-2\pi ix)^\alpha\cdot 1)^\wedge = D^\alpha\delta$, which essentially provides us a way to compute the Fourier transform of any polynomial, i.e. as a linear combination of dirac deltas and the distribution derivatives of dirac deltas, which are derivatives evaluated at points.*

Example. *Consider the Hilbert kernel $\Lambda = p.v(1/x)$. We have seen this distribution is tempered, so we can take it's Fourier transform. Now $x\Lambda = 1$, so the dericative of $\widehat{\Lambda}$ is $-2\pi i\delta$, where δ is the Dirac delta function at the origin. But this means there exists A such that $\widehat{\Lambda}(\xi) = A - 2\pi i\cdot\mathbf{I}(\xi > 0)$. But*

$\Lambda(-x) = -\Lambda(x)$, implying that $\widehat{\Lambda}(-\xi) = -\widehat{\Lambda}(\xi)$, and thus $A - 2\pi i = -A$, i.e. $A = i\pi$. Thus

$$\widehat{\Lambda}(\xi) = \pi i - 2\pi i \cdot \mathbf{I}(\xi > 0) = -i\pi \cdot sgn(\xi).$$

Theorem 9.41. *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*

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

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 must understand the integral

$$\int_{\mathbf{R}^d} \widehat{\phi}(x)\, d\mu(x).$$

Applying Fubini's theorem, which applies since μ has finite mass, we conclude that

$$\int_{\mathbf{R}^d} \widehat{\phi}(x)\, d\mu(x) = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \phi(\xi) e^{-2\pi i \xi \cdot x} d\mu(x)\, d\xi = \int_{\mathbf{R}^d} \phi(\xi) f(\xi)\, d\xi,$$

where

$$f(\xi) = \int_{\mathbf{R}^d} e^{-2\pi i \xi x} d\mu(x).$$

Thus $\widehat{\mu}$ is precisely f, and it suffices to show that $\|f\|_{L^\infty(\mathbf{R}^d)} \leqslant \|\mu\|$, and that f is uniformly continuous. The inequality follows from a simple calculation of the triangle inequality, and the second inequality follows because for some y,

$$\begin{aligned} |f(\xi + \eta) - f(\xi)| &= \left| \int_{\mathbf{R}^d} e^{-2\pi i \xi \cdot x} (e^{-2\pi i \eta \cdot x} - 1)\, d\mu(x) \right| \\ &\leqslant \int_{\mathbf{R}^d} |e^{-2\pi i \eta \cdot x} - 1|\, d|\mu|(x). \end{aligned}$$

As $\eta \to 0$, the dominated convergence theorem implies that this quantity tends to zero, which proves uniform continuity. On the other hand, if $x_i \mu$ is finite for some i, then

$$\frac{f(\xi + \varepsilon e_i) - f(\xi)}{\varepsilon} = \int_{\mathbf{R}^d} e^{-2\pi i \xi \cdot x} \frac{(e^{-2\pi \varepsilon i x_i} - 1)}{\varepsilon} d\mu(x).$$

We can apply the dominated convergence theorem to show that as $\varepsilon \to 0$, this quantity converges to the classical partial derivative f_i, which has the integral formula

$$f_i(\xi) = (-2\pi i) \int_{\mathbf{R}^d} e^{-2\pi i \xi \cdot x} x_i d\mu(x),$$

which is the Fourier transform of $x_i \mu$. Higher derivatives are similar. $\square$

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 $D^\alpha(\Lambda * \phi) = (D^\alpha \Lambda) * \phi = \Lambda * (D^\alpha \phi)$ just as for $\phi \in \mathcal{D}(\mathbf{R}^d)$. Moreover, $\Lambda * \phi$ is a slowly increasing function; to see this, we know there is n such that

$$|\Lambda \phi| \lesssim \|\phi\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)}.$$

Now for $|y| \geqslant 1$,

$$\|T_y \phi\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)} \leqslant |x-y|^n \leqslant 2^n(1+|y|^n)\|\phi\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)},$$

and so

$$(\Lambda * \phi)(x) = \Lambda(T_{-x}\phi^*) \lesssim_n (1+|x|^n)\|\phi\|_{\mathcal{S}^{n,m}(\mathbf{R}^d)},$$

which gives that $\Lambda * \phi$ is slowly increasing. In particular, we can take the Fourier transform of $\Lambda * \phi$. Now for any $\psi \in \mathcal{S}(\mathbf{R}^d)$ with $\widehat{\psi} \in \mathcal{D}(\mathbf{R}^d)$,

$$\begin{aligned}
\int_{\mathbf{R}^d} \widehat{\Lambda * \phi}(\xi)\psi(\xi)\, d\xi &= \int_{\mathbf{R}^d} (\Lambda * \phi)(x)\widehat{\psi}(x)\, dx \\
&= \int_{\mathbf{R}^d} \Lambda(\widehat{\psi}(x) \cdot T_{-x}\phi^*)\, dx \\
&= \Lambda\left(\int_{\mathbf{R}^d} \widehat{\psi}(x) T_{-x}\phi^*\, dx \right) \\
&= \Lambda\left(\widehat{\psi} * \phi^*\right) = \Lambda\left(\widehat{\psi} * \widehat{\widehat{\phi}}\right) \\
&= \Lambda\left(\widehat{\psi\widehat{\phi}}\right) = \widehat{\Lambda}\left(\psi\widehat{\phi}\right) = \widehat{\phi}\widehat{\Lambda}(\psi).
\end{aligned}$$

We therefore conclude that $\widehat{\Lambda * \phi} = \widehat{\phi}\widehat{\Lambda}$.

Because of the dilation symmetry of the Fourier transform, the family of homogeneous distributions (which are all tempered) is invariant under the Fourier transform. More precisely, the Fourier transform of a distribution on $\mathbf{R}^d$ which is homogeneous of degree σ is a homogeneous distribution of degree $-d-\sigma$.

Lemma 9.42. *If u is homogeneous and in $C^\infty(\mathbf{R}^d - \{0\})$, then so is $\widehat{u}$.*

Proof. Suppose first that u is homogeneous of order a, with $\text{Re}(a) < -n$. Then we can write $u = u_0 + u_1$, where u_0 is supported near the origin, and u_1 is an integrable function. But the Fourier transform of both of these terms is continuous. Thus $\widehat{u}$ is continuous.

To upgrade this fact, given any homogeneous function u in $C^\infty(\mathbf{R}^d - \{0\})$, if α is suitably large, then $D^\alpha u$ is homogeneous of degree less than $-n$ and in $C^\infty(\mathbf{R}^d - \{0\})$, and so $\widehat{D^\alpha u} = \xi^\alpha \widehat{u}$ is continuous. But this means tht $|\xi|^\alpha \widehat{u}$ is continuous, which means $\widehat{u}$ is continuous. But since $x^\beta u$ is homogeneous if u is homogeneous, we conclude that $\widehat{x^\beta u} = D^\beta \widehat{u}$ is continuous for all β, so that $\widehat{u}$ lies in $C^\infty(\mathbf{R}^d - \{0\})$. □

9.9 Test Functions on a Manifold

Most of the theory of distributions generalizes easily to a smooth manifold M. One can define the topology on $C^\infty(M)$ such that for each compact set $K \subset M$, and each family of smooth vector fields $X_1, \dots, X_k$ defined on a neighborhood of K, we have a seminorm

$$\rho_{K,X}(f) = \|(X_1 \circ \dots \circ X_k) f\|_{L^\infty(K)}.$$

This induces a relative topology on $C_c^\infty(U)$ for any open set U with $\overline{U}$ closed. The inductive limit gives the topology on the space $C_c^\infty(M)$ of compactly supported test functions. We define $\mathcal{D}^*(M)$ to be the dual of $C_c^\infty(M)$, the space of distributions on M, and let $\mathcal{E}^*(M)$ to be the dual of $C^\infty(M)$. One can define the support of a distribution on a manifold, and $\mathcal{E}^*(M)$ can then be seen as the space of compactly supported distributions on M. It is more difficult to build a canonical definition of the space $\mathcal{S}^*(M)$ of tempered distributions; there is no natural direction to measure the rate of decay of a function, so it is difficult to even define the Schwartz space $\mathcal{S}(M)$. Thus to do the things one does in Schwartz space on a general manifold (Fourier-type arguments), one must either work locally, or on a compact manifold where $C^\infty(M)$ should coincide with $\mathcal{S}(M)$.

In the general setting of a manifold, there is no natural way to embed $C^\infty(M)$ in $\mathcal{D}^*(M)$, nor embed $C_c^\infty(M)$ in $\mathcal{E}^*(M)$, since there is no natural way to integrate the product of two functions on a manifold. Nonetheless, given any smooth, non-vanishing positive density ω on M, we can define, for $f \in C_c^\infty(M)$ and $g \in C^\infty(M)$,

$$\langle f, g \rangle = \int_M f(x) g(x) d\omega(x).$$

This is a continuous bilinear map, and thus induces a canonical way to embed the functions as distributions.

9.10 Paley-Wiener Theorem

TODO: See Rudin, Functional Analysis.

Theorem 9.43. *Let f be an entire function on $\mathbf{C}^n$. Then there exists a compactly supported distribution u on $\mathbf{R}^n$ such that for all $z \in \mathbf{C}^n$,*

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

if and only if there exists $n > 0$ and $B > 0$ such that

$$|f(x+iy)| \lesssim (1+|z|)^n e^{B|y|}.$$

In this case, the distribution u will be supported on the closed ball of radius B centered at the origin.

Chapter 10

Spectral Analysis of Singularities

Suppose u is a distribution on $\mathbf{R}^d$. The *singular support* of u is the set of points $x_0 \in \mathbf{R}^d$ which *do not* have an open neighbourhood upon which u acts as integration against a C^∞ function. Understanding the singular support of a distribution, and how to control it, is often a useful perspective in harmonic analysis; to reduce the study of u to the study of a C^∞ function one need only smoothen around the singular support of u.

The smoothess of a distribution is linked to the decay of it's Fourier transform. In particular, suppose there is a compactly supported bump function $\phi \in C^\infty(\mathbf{R}^d)$ with $\phi(x) = 1$ in a neighbourhood of some point $x_0 \in \mathbf{R}^d$. Since ϕu is compactly supported, the Paley-Wiener theorem implies $\widehat{\phi u}$ is an entire function with polynomial growth at infinity. The Fourier inversion formula implies that $\phi u \in \mathcal{D}(\mathbf{R}^d)$ if and only if for all $N \geqslant 0$, $|\widehat{\phi u}(\xi)| \lesssim_N |\xi|^{-N}$. Thus we can infer the singular support of u via purely spectral means, provided we are first able to localize about a point.

We can therefore gain more detailed information about singularities of a distribution u through the Fourier transform. If x_0 is a singularity of u, then for any bump function $\phi \in C^\infty(\mathbf{R}^d)$ with $\phi(x) = 1$ in a neighbourhood of x_0, there must exist some direction in frequency space on which $\widehat{\phi u}$ does not decay. However, this does not mean that ϕ is unable to decay in certain directions; there might exist a conical neighbourhood U about the origin containing some frequency ξ_0 such that for all $\xi \in U$ and all $N > 0$,

$$|\widehat{u\phi}(\xi)| \lesssim_N |\xi|^{-N}. \tag{10.1}$$

the set of values ξ_0 which do *not* satisfy (10.1) for any choice of bump function ϕ about x_0 forms a closed conical subset of $\mathbf{R}^d$, and we call this the *wavefront* of u about the singularity x_0. The set

$$\text{WF}(u) = \{(x_0, \xi_0) : \xi_0 \text{ is in the wavefront of } u \text{ at } x_0\}$$

is the *wavefront set* of the distribution, and provides a deeper characterization of the singularities of u. For instance, in order to smoothen out a distribution u one need only average along the directions in the wavefront set.

Let us now discuss the wavefront set a little more precisely. If u is a compactly supported distribution on $\mathbf{R}^d$, we define $\Gamma(u)$ to be the set of $\xi_0 \in \mathbf{R}^d$ which have no conical neighbourhood U such that for each $N > 0$ and $\xi \in U$,

$$|\widehat{u}(\xi)| \lesssim_N |\xi|^{-N}. \tag{10.2}$$

It is simple to verify via a compactness argument that if $\Gamma(u) = \varnothing$, then $u \in C^\infty(\mathbf{R}^d)$.

Lemma 10.1. *If u is a compactly supported distribution and $\phi \in \mathcal{D}(\mathbf{R}^d)$, then*

$$\Gamma(\phi u) \subset \Gamma(u).$$

Proof. Suppose $\xi_0 \notin \Gamma(u)$, so ξ_0 has a conical neighbourhood U such that (10.2) holds. Then there exists $\varepsilon > 0$ such that U contains

$$\left\{ \eta \in \mathbf{R}^d : \frac{\xi_0 \cdot \eta}{|\xi_0||\eta|} \geqslant 1 - 2\varepsilon \right\}$$

Let V be the conical neighbourhood of ξ_0 defined by setting

$$V = \left\{ \eta \in \mathbf{R}^d : \frac{\xi_0 \cdot \eta}{|\xi_0||\eta|} \geqslant 1 - \varepsilon \right\}.$$

We claim V satisfies (10.2). Fix $\xi \in V$. Then

$$|\widehat{\phi u}(\xi)| = (\widehat{\phi} * \widehat{u})(\xi) = \int_{\mathbf{R}^d} \widehat{\phi}(\eta) \widehat{u}(\xi - \eta)\, d\xi.$$

If $|\xi - \eta| \leqslant 0.25\varepsilon|\xi|$, then it is simple to verify that

$$(\xi_0 \cdot \eta) \geqslant (1 - 2\varepsilon)|\xi_0||\eta|$$

so $\eta \in U$. Thus for any $N > 0$, $\widehat{u}(\eta) \lesssim_N 1/(1+|\eta|)^N$. Since $\phi \in L^\infty(\mathbf{R}^d$, we conclude

$$\begin{aligned}\int_{|\eta| \leqslant 0.25\varepsilon|\xi|} \widehat{\phi}(\eta)\widehat{u}(\xi - \eta)\, d\xi &\lesssim_\phi \int_{|\eta| \leqslant 0.25\varepsilon|\xi|} \frac{1}{1 + |\xi - \eta|^N} \\ &\lesssim_{\varepsilon,d} \frac{|\xi|^d}{(1 + 2|\xi|^N)} \lesssim \frac{1}{1 + |\xi|^{N-d}}.\end{aligned}$$

On the other hand, since u is compactly supported, $\widehat{u}$ is slowly increasing, i.e. there exists $m > 0$ such that

$$|\widehat{u}(\xi)| \leqslant 1 + |\xi|^m.$$

Since $\phi \in \mathcal{D}(\mathbf{R}^d)$, we have $|\widehat{\phi}(\eta)| \lesssim_M 1/(1+|\eta|^M)$ for all $M > 0$ and thus we conclude that if $M > m + d$

$$\begin{aligned}\int_{|\eta| \geqslant 0.25\varepsilon|\xi|} \widehat{\phi}(\eta)\widehat{u}(\xi - \eta) &\lesssim_M \int_{|\eta| \geqslant 0.25\varepsilon|\xi|} \frac{1 + |\xi - \eta|^m}{1 + |\eta|^M} \\ &\lesssim_{\varepsilon,m} \int_{|\eta| \geqslant 0.25\varepsilon|\xi|} \frac{1 + |\eta|^m}{1 + |\eta|^M} \\ &\lesssim_{\varepsilon,d} \frac{1}{1 + |\xi|^{M-m-d}}.\end{aligned}$$

Choosing the parameter M and N appropriately, we obtain the required bound which shows that $\xi_0 \notin \Gamma(\phi u)$. $\square$

This fact means we can obtain a consistant localization about a point. If u is a distribution, $\phi_1, \phi_2 \in \mathcal{D}(\mathbf{R}^d)$ are given, and the support of ϕ_2 is compactly supported on the support of ϕ_1, then $\phi_2/\phi_1 \in \mathcal{D}(\mathbf{R}^d)$, and so we conclude that

$$\Gamma(\phi_2 u) = \Gamma((\phi_2/\phi_1)\phi_1 u) \subset \Gamma(\phi_1 u).$$

Thus if u is a distribution, and $x \in \mathbf{R}^d$, then we define $\Gamma_x(U)$ to be equal to

$$\bigcap \Big\{ \Gamma(\phi u) : \phi \in \mathcal{D}(\mathbf{R}^d), x \in \text{supp}(\phi) \Big\}.$$

It is simple to see that if $\{\phi_n\}$ is a sequence in $\mathcal{D}(\mathbf{R}^d)$ such that $\text{supp}(\phi_{n+1})$ is compactly supported in $\text{supp}(\phi_n)$ for each n, and if $\bigcap \text{supp}(\phi_n) = \{x\}$, then $\Gamma_x(u) = \lim_{n\to\infty} \Gamma(\phi_n u)$. Finally, we define

$$\text{WF}(u) = \{(x,\xi) : \xi \in \Gamma_x(u)\}.$$

This is the *wavefront set* of u.

Lemma 10.2. *If u is a distribution, then $\pi_x(WF(u))$ is the singular support of u. If u is compactly supported, then $\pi_\xi(WF(u)) = \Gamma(u)$.*

Proof. Fix $x_0 \in \mathbf{R}^d$. If $(x_0, \xi_0) \notin \text{WF}(u)$ for all $\xi_0 \in \mathbf{R}^d$, then there exists $\phi \in \mathcal{D}(\mathbf{R}^d)$ such that $\phi(x_0) \neq 0$ and $\Gamma(\phi u) = \varnothing$. But this means $\phi u \in \mathcal{D}(\mathbf{R}^d)$, so x_0 is not in the singular support of u. This shows $\pi_x(\text{WF}(u))$ is a subset of the singular support. The converse is obvious.

On the other hand, let us assume u is compactly supported, and that $\xi_0 \notin \Gamma(u)$. Then $(x_0, \xi_0) \notin \text{WF}(u)$ for any $x_0 \in \mathbf{R}^d$ since $\Gamma(\phi u) \subset \Gamma(u)$ for any $\phi \in \mathcal{D}(\mathbf{R}^d)$. But if $(x_0.\xi_0) \notin \text{WF}(u)$ for any $x_0 \in \mathbf{R}^d$ we can cover the support of u by a partition of unity $\phi_1, \dots, \phi_N \in \mathcal{D}(\mathbf{R}^d)$ such that $\xi_0 \notin \Gamma(\phi_i u)$ for each i, and summing up shows $\xi_0 \notin \Gamma(u)$. □

Example. *Suppose u is a homogenous distribution which is C^∞ away from the origin. Then $\widehat{u}$ is homogenous and C^∞ away from the origin, and we claim that*

$$WF(u) = \{(0, \xi) : \xi \in supp(\widehat{u})\}.$$

Since the singular support of u is $\{0\}$, we know $WF(u) \subset \{0\} \times \mathbf{R}^d$, and so it suffices to calculate $\Gamma_0(u)$. Fix a radial function $\phi \in \mathcal{D}(\mathbf{R}^d)$. Let $\phi_t(x) = t^{-d}\phi(tx)$, let $\psi(\xi) = \widehat{\phi}(\xi)$, and let $\psi_t(\xi) = \widehat{\phi_t}(\xi) = \psi(\xi/t)$. If v is the homogeneous distribution given by the Fourier transform of u, then

$$v_t = \widehat{\phi_t u} = \psi_t * v$$

If $\xi \notin supp(v)$, then it is obvious that $(0, \xi) \in WF(u)$. Conversely, if $(0, \xi) \notin WF(u)$, then for suitably large t, there is an open cone $U_t \subset \mathbf{R}^d$ containing ξ such that $v_t(\eta) \lesssim_{t,N} |\eta|^{-N}$ for all $\eta \in U_t$. Now if v is homogeneous of degree s, then

$$\begin{aligned}(\psi_t * v)(r\eta) &= \int v(\eta')\psi_t\left(r\eta - \eta'\right) d\eta' \\ &= r^{d+s}(v * Dil_r \psi_t)(\eta),\end{aligned}$$

where $Dil_r \psi_t(\eta) = \psi_t(r\eta)$. But this means that

$$|r^d(v * Dil_r \psi_t)(\eta)| \lesssim_{t,N} r^{-N-s}.$$

As $r \to \infty$, $r^d(v * Dil_r \psi_t)(\eta) \to v(\eta)$. Taking $r \to \infty$, choosing $N > -s$, we conclude that $v(\eta) = 0$. Thus we find that v vanishes on U_t. Thus $\xi \notin supp(v)$, which completes the argument.

Example. *Thus if δ is the Dirac delta function at the origin in $\mathbf{R}^d$, which is homogeneous of degree $-d$, then $\widehat{\delta}(\xi) = 1$, hence $WF(\delta) = \{0\} \times \mathbf{R}^d$.*

Example. *If $u(x) = p.v(1/x)$ is the distribution on $\mathbf{R}$, then*

$$\widehat{u}(\xi) = -i\pi \cdot sgn(\xi).$$

Thus $WF(u) = \{0\} \times \mathbf{R}$.

The fact that $(x_0, \xi_0) \notin \mathrm{WF}(u)$ implies precisely that there exists a neighbourhood U_0 of x_0 such that for any $\phi \in C_c^\infty(U_0)$, and any $N > 0$,

$$\int_{\mathbf{R}^d} u(x)\phi(x)e^{-2\pi i \lambda \xi \cdot x}\, dx \lesssim_N \langle \xi \rangle^{-N}.$$

It will be useful to consider a nonlinear analogue of this statement, which will be useful for showing the invariance of the wavefront set under changes of variables.

Theorem 10.3. *Let u be a distribution, and let $(x_0, \xi_0) \notin WF(u)$. Let U be an open subset of $\mathbf{R}^d$ containing x_0, let V be an open subset of $\mathbf{R}^p$ containing a_0, and let $\psi : U \times V \to \mathbf{R}$ be a C^∞ function with $\nabla_x \psi(x_0, a_0) = \xi_0$. Then there is an open set U_0 of x_0, an open set V_0 of a_0 such that for any $\phi \in C_c^\infty(U_0)$, and any $N > 0$,*

$$\left| \int u(x)\phi(x)e^{-2\pi i \lambda \psi(x,a)}\, dx \right| \lesssim_N \lambda^{-N}$$

where the bound is uniform on V_0.

Proof. Fix $\varepsilon > 0$, to be chosen later, and choose U_0 and V_0 such that $|\nabla_x \psi(x,a) - \xi_0| \leqslant \varepsilon/2$ for $x \in U_0$ and $a \in V_0$. For any given $\phi \in C_c^\infty(U_0)$, consider $\tilde{\phi} \in C_c^\infty(U_0)$ with $\tilde{\phi}\phi = \phi$. Then

$$\begin{aligned}
\int u(x)\phi(x)e^{-2\pi i \lambda \psi(x,a)}\, dx &= \int u(x)\phi(x)\phi_1(x)e^{-2\pi i \lambda \psi(x,a)}\, dx \\
&= \int \widehat{u\phi}(\xi) \left(\int \phi_1(x) e^{-2\pi i (\lambda \psi(x,a) - \xi)}\, dx \right) d\xi \\
&= \lambda^d \int \widehat{u\phi}(\xi) \left(\int \phi_1(x) e^{-2\pi \lambda i (\psi(x,a) - \xi)}\, dx \right) d\xi \\
&= \lambda^d \int \widehat{u\phi}(\lambda \xi) J(\lambda, \xi, a)\, d\xi.
\end{aligned}$$

Let $\eta \in \mathcal{D}(\mathbf{R}^d)$ be a smooth bump function supported on $|\xi| \leqslant 1$ and with $\eta(\xi) = 1$ for $|\xi| \leqslant 1/2$. Fix $\varepsilon > 0$, and write $J(\lambda, \xi, a) = J_1(\lambda, \xi, a) + J_2(\lambda, \xi, a)$, where

$$J_1(\lambda, \xi, a) = \eta\left(\frac{\xi - \xi_0}{\varepsilon}\right) \int_{\mathbf{R}^d} \phi_1(x) e^{-2\pi\lambda i(\psi(x,a) - \xi)}\, dx$$

and

$$J_2(\lambda, \xi, a) = \left(1 - \eta\left(\frac{\xi - \xi_0}{\varepsilon}\right)\right) \int_{\mathbf{R}^d} \phi_1(x) e^{-2\pi\lambda i(\psi(x,a) - \xi)}\, dx.$$

If ε is chosen appropriately small, then $|\widehat{u\phi}(\lambda\xi)| \lesssim_N \lambda^{-N}$ uniformly for $|\xi - \xi_0| \leqslant \varepsilon$. Since $|J_1(\lambda, \xi, a)| \lesssim 1$, this implies

$$\left| \int \widehat{u\phi}(\lambda\xi) J_1(\lambda, \xi, a) \right| d\xi \lesssim_N \lambda^{-N}.$$

On the other hand, if $|\xi - \xi_0| \geqslant \varepsilon$, then $|\nabla_x \phi(x,a) - \xi| = |\xi_0 - \xi| - \varepsilon/2 \geqslant \varepsilon/2$. Thus the method of stationary phase implies that

$$|J_2(\lambda, \xi, a)| \lesssim_N \lambda^{-N},$$

uniformly in a. Combined with the fact that $\widehat{u\phi}$ is of polynomial growth, this implies that

$$\left| \int \widehat{u\phi}(\lambda\xi) J_1(\lambda, \xi, a)\, d\xi \right| \lesssim_N \lambda^{-N}.$$

Combining these two estimates completes the proof. $\square$

For a smooth function $\phi \in C_c^\infty(V)$ and a smooth diffeomorphism $f : U \to V$, we can define $f^*\phi \in C_c^\infty(U)$ by setting $f^*\phi(x) = \phi(f(x))$. Then for $\psi \in C_c^\infty(V)$,

$$\int f^*\phi(x)\psi(x) = \int \phi(f(x))\psi(x) = \int \phi(y)\psi(f^{-1}(y)) \cdot \frac{1}{|f'(f^{-1}(y))|}\, dy.$$

Thus for a distribution u on V, to define a distribution f^*u on U such that for $\psi \in \mathcal{D}(\mathbf{R}^d)$,

$$\int (f^*u)(x)\psi(x) = \int u(y)\phi(f^{-1}(y)) \cdot \frac{1}{|f'(f^{-1}(y))|}\, dy.$$

There is a simple relation between the wavefront set of u and f^*u. We consider $f^* : V \times \mathbf{R}^d \to U \times \mathbf{R}^d$ by defining $f^*((y,v)) = (f^{-1}(y), f'(y)^T v)$. This agrees with the definition of f^* encountered in differential geometry if we identify $V \times \mathbf{R}^d$ and $U \times \mathbf{R}^d$ with the cotangent bundle T^*V and T^*U.

Theorem 10.4. *For any distribution u on V, $WF(f^*u) = f^*(WF(U))$.*

Proof. Assume $(y_0, \eta_0) \notin \text{WF}(u)$, let $(x_0, \xi_0) = f^*((y_0, \eta_0))$, and then define $\psi(y, \xi) = f^{-1}(y) \cdot \xi$. Then

$$\nabla_y \psi(y_0, \xi_0) = (f^{-1}(y_0)')^T(\xi_0) = \eta_0.$$

Thus, applying the previous theorem, since

$$\widehat{f^*(u\phi)}(\lambda\xi) = \int u(y) \frac{\phi(f^{-1}(y))}{|f'(f^{-1}(y))|} e^{-2\pi i \lambda \xi \cdot f^{-1}(y)} = \int u(y) \tilde{\phi}(y) e^{-2\pi i \lambda \psi(y,\xi)}\, dy,$$

we conclude that $|\widehat{f^*(u\phi)}(\lambda\xi) \lesssim_N \lambda^{-N}$, which implies $(x_0, \xi_0) \in \text{WF}(f^*(u))$. Thus $\text{WF}(f^*u) \subset f^*(\text{WF}(u))$. The converse statement that $f^*(\text{WF}(u)) \subset \text{WF}(f^*u)$ is obtained by symmetry. $\square$

Using this change of variables formula, we see that the wavefront set transforms under a change of coordinates like a covector. Since this gives an invariant definition, we can define the wavefront set of distributions on any smooth manifold M, and the wavefront set will then be a closed, conical subset of T^*M. In the next, optional section, we develop this theory, defining the faimly of distributions on a manifold.

10.1 Wavefront Sets on Manifolds

The lack of a natural volume form on a manifold M prevents us from obtaining a canonical, invariant definition of integration on a manifold, which enables us to define the family of distributions from compactly supported, smooth test functions on the manifold.

Recall that for any manifold M, we can define a one-dimension bundle $\text{Vol}(TM)$, sections of which we call *scalar densities*. In coordinates, a scalar density corresponds to a family of functions $\omega_x \in C^\infty(U)$ for each coordinate chart (x, U) such that for any other coordinate chart (y, V), on $U \cap V$,

$$\omega_y = \omega_x \cdot \left|\det\left(\frac{\partial y^i}{\partial x^j}\right)\right|.$$

Given any $f \in C^\infty(M)$, and any *compactly supported* scalar density $\omega \in \Gamma(\text{Vol}(TM))$, the quantity

$$\langle f, \omega \rangle = \int_M f \cdot \omega$$

is well defined. We can equip the family of *compactly supported* scalar densities with a topology analogous to the topology on the family of compactly supported test functions on some open subset of Euclidean space. The operator $\omega \mapsto \langle f, \omega \rangle$ is then continuous, and so we can view $C^\infty(M)$ as a subfamily of the continuous dual of the space of compactly supported scalar densities. We therefore obtain a satisfactory, coordinate independent definition of $\mathcal{D}^*(M)$ on any manifold M. With this definition, the theory of wavefront sets follows pretty much automatically by working in coordinates. For instance, an analogous version of the Schwartz kernel theorem holds, and for each $u \in \mathcal{D}^*(M)$, we have a natural closed, conic subset $\text{WF}(u)$ of T^*M. The local theory is essentially the same as the Euclidean space, so for simplicity, we stick to this case in the sequel.

Remark. We can define a bundle $\text{Vol}^\alpha(TM)$ for each $0 \leqslant \alpha \leqslant 1$ of *scalar densities of order* α, such that $\text{Vol}(TM) = \text{Vol}^1(TM)$, and $C^\infty(M) = \text{Vol}^0(TM)$. We have a natural bilinear pairing $\text{Vol}^\alpha(TM) \times \text{Vol}_c^{1-\alpha}(TM) \to \mathbf{R}$, where $\text{Vol}_c^{1-\alpha}(TM)$ is the family of compactly supported scalar densities of order $1-\alpha$. Equipping $\text{Vol}_c^{1-\alpha}(TM)$ enables us to define the family $(\mathcal{D}^\alpha)'(M)$ of distributional scalar densities as the dual of $\text{Vol}_c^{1-\alpha}(TM)$. There is a similar theory of wavefront sets for this family.

The theory of wavefront sets can also be considered in a sheaf theoretic framework. Given a manifold M, consider the 'unit' tangent bundle $U(M)$, which, as a space, is the quotient of TM identifying vectors which are constant multiplers of one another. We can view the family of distributions on a manifold as sections of a sheaf $\mathcal{D}^*$, since one can restrict and glue distributions defined on open subsets of a manifold. Similarily, the smooth functions on a manifold form a sheaf C^∞, which is a subsheaf of $\mathcal{D}^*$. The sheaf $\mathcal{D}^*$ is flasque, so we can restrict ourselves to studying global sections. Given an open subset V of $U(M)$, we consider the family $\mathcal{F}(U)$ of all *equivalence classes* of distributions in $\mathcal{D}^*(M)$, where u and v are identified if $\mathrm{WF}((u-v)\cap V) = \emptyset$. It is simple to see that $\mathcal{F}$ gives the structure of a *presheaf*. Global sections of this presheaf corresponds to elements of $\mathcal{D}^*(M)/C^\infty(M)$, and the support of any $u \in \mathcal{D}^*(M)$ in $\mathcal{F}$ is then $\mathrm{WF}(u)$.

10.2 Oscillatory Integral Distributions

In this section, we consider distributions on an open subset U of $\mathbf{R}^d$, formally defined by the formula

$$I_{a,\phi}(x) = \int a(x,\theta) e^{2\pi i \phi(x,\theta)}\, d\theta.$$

Here a is a *symbol* lying in some class $\mathcal{S}^t(U \times \mathbf{R}^p)$, i.e. satisfying bounds of the form

$$|\nabla_x^n \nabla_\theta^m a(x,\theta)| \lesssim_{n,m} \langle \theta \rangle^{t-m},$$

and $\phi \in C^\infty(U \times (\mathbf{R}^d - \{0\}))$ is homogeneous of degree one in θ, such that $(\nabla_x \phi, \nabla_\theta \phi)$ is nonvanishing on the support of a.

If $t < -d$, then $I_{a,\phi}$ can be interpreted as an absolutely convergent integral, and in this case $I_{a,\phi}$ is actually a locally integrable function. On the other hand, if $t \geqslant -d$, then $I_{a,\phi}$ will no longer act as a locally integrable function; for instance, our definition will show that the distribution

$$\int_{\mathbf{R}^d} \xi^t e^{2\pi i x \cdot \xi}\, d\xi$$

acts on functions as a constant multiple of the differential operator D^t.

To define the oscillatory integral distribution rigorously, we fix $\psi \in \mathcal{D}(\mathbf{R}^d)$, and $\rho \in \mathcal{D}(\mathbf{R}^p)$, equal to one in a neighborhood of the origin. We claim that the quantity

$$\lim_{R \to \infty} \int a(x,\theta)\psi(x)\rho(\theta/R) e^{2\pi i \phi(x,\theta)}\, d\theta$$

exists, and is independent of the choice of bump function ρ. This will be our definition of

$$\int I_{a,\phi}(x)\psi(x)\, dx.$$

To prove the limit exists, we fix $R_1 \leqslant R_2$, and let $\tilde{\rho}(\theta) = \rho(\theta/R_2) - \rho(\theta/R_1)$. Then $\tilde{\rho}$ is supported on $R_1 \lesssim |x| \lesssim R_2$. Assume first that $R_2 \leqslant 2R_1$. Rescaling, we find that if $\eta(x,\theta) = a(x, R_2\theta)\psi(x)\rho(\theta)$, then

$$\begin{aligned}\int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i\phi(x,\theta)} a(x,\theta)\psi(x)\tilde{\rho}(\theta) &= R_2^m \int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i R_2\phi(x,\theta)} a(x, R_2\theta)\psi(x)\tilde{\rho}(\theta)\\ &= R_2^p \int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i R_2\phi(x,\theta)} \eta(x,\theta).\end{aligned}$$

Then η is supported on $1/2 \lesssim |\theta| \lesssim 1$ and $|x| \lesssim 1$. Thus the support of η is independant of R_1 and R_2. It is simple to verify that

$$|\nabla_x^n \nabla_\theta^m \eta(x,\theta)| \lesssim_{n,m} R_2^t \cdot |\nabla_x^n \psi(x)|,$$

where the bound is independant of R_1 and R_2. Since $\nabla_x\phi$ and $\nabla_\theta\phi$ have no common zeroes on the support of a, and thus ψ, we can apply the principle of stationary phase to conclude that

$$\left| R_2^p \int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i R_2\phi(x,\theta)} \eta(x,\theta) \right| \lesssim_N R_2^{p+m-N} \cdot \|\nabla^{\leqslant N}\psi\|_{L^\infty(\mathbf{R}^d)}.$$

In general, if $R_2 \geqslant 2R_1$, we consider the largest l such that $2^l R_1 \leqslant R_2$. If we set $a_k = 2^k R_1$ for $0 \leqslant k \leqslant l$, and $a_{l+1} = R_2$, then we write

$$\begin{aligned}&\left| \int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i\phi(x,\theta)} a(x,\theta)\phi(x)\tilde{\rho}(\theta) \right|\\ &\quad = \left| \sum_{k=0}^{l} \int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i\phi(x,\theta)} a(x,\theta)\phi(x)(\rho(\theta/a_{k+1}) - \rho(\theta/a_k)) \right|\\ &\quad \lesssim \sum_{k=0}^{l} a_{k+1}^{p+m-N} \|\nabla^{\leqslant N}\psi\|_{L^\infty(\mathbf{R}^d)}.\end{aligned}$$

If we choose $N > p + m$, then we conclude that

$$\begin{aligned}\left| \int_{\mathbf{R}^n}\int_{\mathbf{R}^p} e^{2\pi i\phi(x,\theta)} a(x,\theta)\phi(x)\tilde{\rho}(\theta) \right| &\lesssim (R_1^{p+m-N} + R_2^{p+m-N}) \|\nabla_x^{\leqslant N}\psi\|_{L^\infty(\mathbf{R}^d)}\\ &\lesssim R_1^{p+t-N} \|\nabla^{\leqslant N}\psi\|_{L^\infty(\mathbf{R}^d)}.\end{aligned}$$

In particular, this quantity tends to zero as $R_1 \to \infty$, which gives convergence of the limit, and also gives boundedness, showing $I_{a,\phi}$ is a distribution of order N, where N is the smallest integer bigger than $p+m$. A very similar argument shows that if ρ is equal to zero in a neighborhood of the origin, then

$$\lim_{R\to\infty} \int_{\mathbf{R}^n} \int_{\mathbf{R}^p} e^{2\pi i \phi(x,\theta)} a(x,\theta)\phi(x)\rho(\theta)\phi(x) = 0.$$

It follows from the above observation that the definition is independent of the original choice of ρ. It is left as an exercise to show that the map $a \mapsto I_{a,\phi}$ is continuous from $S^t(U \times \mathbf{R}^p)$ to $\mathcal{D}^*(U)$.

Let us now consider the wavefront set of $I_{a,\phi}$. If ψ is a bump function supported in a neighbourhood of some point x_0, then

$$\widehat{I_{a,\phi}\psi}(\lambda\xi_0) = \lambda^d \int\int e^{2\pi i\lambda(\phi(x,\theta) - x\cdot\xi_0)} a(x,\lambda\theta)\psi(x)\, dx\, d\theta.$$

Since ψ is compactly supported, the method of stationary phase would imply this integral would decay rapidly as $\lambda \to \infty$ if $\nabla_\theta \phi(x_0,\theta_0) \neq 0$, or if $\nabla_\theta\phi(x_0,\theta_0) = 0$, but $\nabla_x \phi(x,\theta) \neq \xi_0$, provided that ψ was supported on a small enough neighbourhood of x_0. Thus

$$\mathrm{WF}(I_{a,\phi}) \subset \{(x_0, \nabla_x\phi(x_0,\theta_0)) : (x_0,\theta_0) \in \text{ess supp}(a) \text{ and } \nabla_\theta\phi(x_0,\theta_0) = 0\}.$$

Here ess supp(a) is the complement of the set of pairs (x_0,θ_0) which have a conical neighborhood upon which a lies in $S^{-\infty}$.

Near the wavefront set, we can compute an asymptotic formula which characterizes the behaviour of the distribution near the wavefront set up to integration against a function in $C^\infty(U)$.

Theorem 10.5. *Consider a phase function ϕ_1. Fix $(x_0,\theta_0) \in U \times \mathbf{R}^p$ such that $\nabla_\theta\phi_1(x_0,\theta_0) = 0$. Let $\xi_0 = \nabla_x\phi_1(x_0,\theta_0)$. Consider any phase function $\phi_2 \in C^\infty(U \times \mathbf{R}^q)$ and $\sigma_0 \in \mathbf{R}^q$ with*

$$\nabla_x\phi_1(x_0,\theta_0) = \nabla_x\phi_2(x_0,\sigma_0).$$

Furthermore, assume that the Hessian $H_{x,\theta}(\phi-\psi)$ is nondegenerate at (x_0,θ_0,σ_0). Then there exists a conical neighborhood Γ of (x_0,θ_0), an open neighborhood V of x_0, and an open neighborhood Σ of σ_0, such that if $\psi \in C_c^\infty(V)$, and a is *a symbol on $U \times \mathbf{R}^p$ with ess sup$(a) \subset \Gamma$, then there exists a family of smooth functions a_k such that as $\lambda \to \infty$,*

$$\int I_{a,\phi_1}(x)\psi(x)e^{-2\pi i\lambda\phi_2(x,\sigma)}\,dx$$
$$\sim e^{-2\pi i\lambda\phi_2(x(\sigma),\sigma)}|\det Q(\sigma)|^{-1/2}e^{(i\pi/4)sgn(Q(\sigma))}\lambda^{(p-d)/2}\sum_{k=0}^{\infty}a_k(\sigma,\lambda)\cdot\lambda^{-k},$$

Here $a_k(\sigma,\lambda)$ is a linear differential operator in a and u at $(x(\sigma),\theta(\sigma),\sigma)$

Example. *If we set $\phi(x,\xi) = x\cdot\xi$, then we find that for any symbol a,*

$$WF(I_{a,\phi}) \subset \{0\}\times\mathbf{R}^n$$

Thus we conclude that $I_{a,\phi}$ is smooth away from the origin. If $a(x,\theta) = a(x)$, then this result says that the Fourier transform of any symbol is smooth away from the origin. This should be compared to the result that the Fourier transform of a homogeneous distribution which is C^∞ away from the origin is C^∞.

The phase ϕ of an oscillatory integral distribution is called *nondegenerate* if whenever $\nabla_\theta\phi(x,\theta) = 0$, the matrix $D(\nabla_\theta\phi)(x,\theta)$ has full rank p. It follows that

$$\Sigma_\phi = \{(x,\theta) : \nabla_\theta\phi(x,\theta) = 0\}$$

is a d dimensional submanifold of $U\times\mathbf{R}^p$. Moreover, the map f from Σ_ϕ to $U\times\mathbf{R}^d$ given by $(x,\theta)\mapsto(x,\nabla_x\phi(x,\theta))$ is an immersion, the image $f(\Sigma_\phi)$ being denoted Λ_ϕ. To verify the map is an immersion, we note that at a point (x,θ) the tangent space of Σ_ϕ consists of vectors $(v,w)\in\mathbf{R}^d\times\mathbf{R}^p$ such that

$$D_x\nabla_\theta\phi(x,\theta)\cdot v + D_\theta\nabla_\theta\phi(x,\theta)\cdot w = 0.$$

Now

$$Df(x,\theta)(v,w) = (v, D_x\nabla_x\phi(x,\theta)\cdot v + D_\theta\nabla_x\phi(x,\theta)\cdot w).$$

Thus if (v,w) lies in the tangent space and $Df(x,\theta)(v,w) = 0$, then $v = 0$, which implies

$$D_\theta\nabla_\theta\phi(x,\theta)\cdot w = D_\theta\nabla_x\phi(x,\theta)\cdot w = 0.$$

Since mixed partials commute, this says exactly that $D(\nabla_\theta\phi)^T\cdot w = 0$. The full rank condition thus implies that $w = 0$. Thus $(v,w) = 0$, completing the argument that f is an immersion.

Many properties about the phase function can be summarized via the immersed manifold Λ_ϕ. For instance, given a function $\psi(x,\sigma)$, the function $\eta(x,\theta,\sigma) = \phi(x,\theta) - \psi(x,\sigma)$ has a nondegenerate stationary point as a function of x and θ at a point (x_0,θ_0,σ_0) if and only if ϕ is nondegenerate phase function in a neighborhood of (x_0,θ_0), and the covector field $d_x\psi$ intersects Λ_ϕ transversally at (x_0,ξ_0), where $\xi_0 = \nabla_x\phi(x_0,\theta_0) = \nabla_x\psi(x_0,\sigma_0)$. In particular, we see that nondegenerate phase functions are 'generic'.

TODO: If $I_{a_1,\phi_1} = I_{a_2,\phi_2}$, prove that $\Lambda_{\phi_1} = \Lambda_{\phi_2}$.

The converse is also true.

Theorem 10.6. *Suppose ϕ_1 and ϕ_2 are nondegenerate phase functions on $U \times \mathbf{R}^{p_1}$ and $U \times \mathbf{R}^{p_2}$. Let*

The immersed manifold Λ_ϕ of $U \times \mathbf{R}^d$ actually has a particular geometric character. Consider the two form

$$\sigma = dx^1 \wedge d\xi^1 + \dots + dx^d \wedge d\xi^d.$$

The $\sigma = d\omega$, where $\omega = \xi^1 dx^1 + \dots + \xi^d dx^d$. We claim that for any $p = (x,\theta) \in \Sigma_\phi$, and any $v,w \in T_p\Sigma_\phi$, $\sigma(f_*v, f_*w) = 0$. To see this, we calculate that

$$f^*(\sigma) = f^*(d\omega) = d(f^*\omega),$$

and

$$f^*\omega = \nabla_x\phi \cdot dx = d\phi - \nabla_\theta\phi \cdot d\theta.$$

On Σ_ϕ, $\nabla_\theta\phi = 0$, so $f^*\omega = d\phi$, and so $f^*(\sigma) = d(f^*\omega) = d^2\phi = 0$. Thus Λ_ϕ is a Lagrangian submanifold of $T^*\mathbf{R}^d$.

For any phase function ϕ (possibly degenerate), we can define $\Sigma_\theta = \{(x,\theta) : \nabla_\theta\phi(x,\theta) = 0\}$, and thus the set Λ_ϕ. If Λ_ϕ is an immersed Lagrangian submanifold, we say $I_{a,\phi}$ is a *Lagrangian distribution*. A degenerate example is given when $p = d+1$, and for $\theta = (\tilde{\theta}, \theta_{d+1})$, $\phi(x,\theta) = x \cdot \tilde{\theta}$. Then $\Lambda_\phi = \{0\} \times \mathbf{R}^d$, which is Lagrangian.

If $f : U \to V$ is a diffeomorphism between open subsets of $\mathbf{R}^d$, and we equip T^*U with coordinates (x,ξ), and T^*V with coordinates (y,η), then we obtain an isomorphism $g : T^*U \to T^*V$ mapping (x,ξ) in T^*V to $(x, (Df(x)^T)^{-1}\eta)$ in T^*U. Under this correspondence, if we consider the two-form $\omega_V = \sum \eta_i \wedge dy^i$ on U, then

$$g^*\omega_V = \sum (\eta^i \circ g) \cdot d(y^i \circ g) = \sum ((Df(x)^T)^{-1}\xi)_i df^i(x) = \sum \xi_i dx^i.$$

Thus the Lagrangian form is invariant under coordinate changes, and can thus be well defined on the cotangent bundle of any manifold M. Thus we can discuss the Lagrangian submanifolds of T^*M for any manifold M.

Example. *Consider a one-form ψ on M, i.e. a smooth function $\psi : M \to T^*M$. Working in coordinates (x,ξ) on T^*M, we have*

$$\psi^*\omega = \sum \psi^i dx^i = d\psi.$$

Thus we see that $\psi^*\sigma = 0$ if and only if $d\psi = 0$, so ψ defines a Lagrangian submanifold of T^*M if and only if it is closed.

10.3 Singular Operations on Distributions

A subset Γ of $\Omega \times \mathbf{R}^d$ is *conic* if $(x,\xi) \in \Gamma$ implies that $(x,\lambda\xi) \in \Gamma$. Given a closed conic set Γ, let $\mathcal{D}^*_\Gamma(\Omega)$ denote the family of all distributions u with $\mathrm{WF}(u) \subset \Gamma$, with the seminorms

$$u \mapsto \sup_{\substack{\xi \in V \\ \lambda > 0}} \lambda^N |\widehat{\phi u}(\lambda\xi)|$$

where V is a closed conic set disjoint from Γ. Then $\mathcal{D}^*_\Gamma(\Omega)$ is a Fréchet space. To analyze this space, we require a lemma about general distributions.

Lemma 10.7. *Let $\mathcal{U}$ be a family in $\mathcal{D}^*(\Omega)$ such that $\sup_{u \in \mathcal{U}} |u(\phi)| < \infty$. Then for any $\phi \in \mathcal{D}(\Omega)$, there exists $m > 0$ such that for any $\xi \in \mathbf{R}^d$,*

$$|\widehat{\phi u}(\xi)| \lesssim (1 + |\xi|)^m,$$

uniformly in u and ξ. If we have a sequence $\{u_n\}$ converging to some distribution u, then as $n \to \infty$,

$$|\widehat{\phi u_n}(\xi) - \widehat{\phi u}(\xi)| = o((1 + |\xi|^m)).$$

In particular, $\widehat{\phi u_n}$ converges to $\widehat{\phi u}$ on compact subsets.

Proof. This result follows from a previous result we proved about bounded and convergent families of distributions and bounds in terms of the norm spaces $C^m_c(K)$ for a compact subset K of Ω, when we plug in the test function $\phi e^{-2\pi i \xi \cdot x}$. ☐

Theorem 10.8. *A sequence of distributions u_n converges to u in $\mathcal{D}^*_\Gamma(U)$ if and only if u_n converges to u distributionally, and for any conic set V disjoint from Γ, and $N > 0$,*

$$\sup_{\xi \in V} \lambda^N |\widehat{\phi u_n}(\lambda \xi)|$$

is bounded independantly of n.

Proof. This conditions are certainly necessary for convergence. Conversely, if these conditions are satisfied, the previous lemma implies that as $n \to \infty$,

$$\sup_{\lambda \geqslant 1} \sup_{\xi \in V} \lambda^{-m} |\widehat{\phi u_n}(\lambda \xi) - \widehat{\phi u}(\lambda \xi)| = o(1).$$

We also know that

$$\sup_{\lambda \geqslant 1} \sup_{\xi \in V} \lambda^{N+1} |\widehat{\phi u_n}(\lambda \xi) - \widehat{\phi u}(\lambda \xi)| < \infty.$$

Let us call this supremum $C > 0$. Given any $\varepsilon > 0$, we find that

$$\sup_{\lambda \geqslant C/\varepsilon} \lambda^N |\widehat{\phi u_n}(\lambda \xi) - \widehat{\phi u}(\lambda \xi)| \leqslant \varepsilon.$$

But

$$\sup_{1 \leqslant \lambda \leqslant C/\varepsilon} \sup_{\xi \in V} \lambda^N |\widehat{\phi u_n}(\lambda \xi) - \widehat{\phi u}(\lambda \xi)| = o((C/\varepsilon)^{N+m}).$$

Taking n suitably large, depending on C, ε, N, and m, we conclude that

$$\sup_{1 \leqslant \lambda \leqslant C/\varepsilon} \sup_{\xi \in V} \lambda^N |\widehat{\phi u_n}(\lambda \xi) - \widehat{\phi u}(\lambda \xi)| \leqslant \varepsilon.$$

Combining this with the supremum above shows that we have convergence in $\mathcal{D}^*_\Gamma(U)$. □

Theorem 10.9. *$\mathcal{D}(\Omega)$ is sequentially dense in $\mathcal{D}^*_\Gamma(\Omega)$.*

Proof. Consider a distribution $u \in \mathcal{D}^*_\Gamma(\Omega)$. Without loss of generality we may assume u is compactly supported. Consider an approximation to the identity $\{\phi_\delta\}$. Then $u * \phi_{1/n} \in C^\infty(\Omega)$, and thus an element of $\mathcal{D}^*_\Gamma(\Omega)$. $u * \phi_{1/n}$ converges to u in $\mathcal{D}^*(\Omega)$, so by the last result, it suffices to show that for any admissable choice of ψ, V, and $N > 0$,

$$\sup_{\xi \in V} \sup_n \lambda^N |\psi \widehat{(u * \phi_n)}(\lambda \xi)| < \infty.$$

But this is simple, for we get arbitrarily fast decay if n is suitably large, depending on the distance from V to Γ, and the finitely many smaller choices of n are negligible. $\square$

We have a continuous map $(\phi,\psi) \to \phi\psi$ from $\mathcal{D}(\Omega) \times \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$, which extends to a continuous map $(\phi,u) \to \phi u$ from $\mathcal{D}(\Omega) \times \mathcal{D}^*(\Omega) \to \mathcal{D}^*(\Omega)$. However, it is *not* possible to extend this to a continuous map $(u,v) \mapsto uv$ from $\mathcal{D}^*(\Omega) \times \mathcal{D}^*(\Omega) \to \mathcal{D}^*(\Omega)$. For instance, if $\{\phi_\varepsilon\}$ is an approximation to the identity, then ϕ_ε converges to the Dirac delta distribution δ at the origin, so we would expect ϕ_ε^2 to converge to a distribution representing the product $\delta \cdot \delta$, but this does not happen because if $\psi \in \mathcal{D}(\mathbf{R}^d)$ and $\psi(x) = 1$ for $|x| \leqslant 1$, then

$$\left| \int \phi_\varepsilon^2(x)\psi(x) \right| \gtrsim 1/\varepsilon$$

and thus does not converge. It is a surprising fact that we can use the wavefront set of a distribution to define the product of two distributions, *provided that the wavefront sets satisfy a disjointness relation*.

To see how this is possible, we note that for $\phi,\psi \in \mathcal{D}(\mathbf{R}^d)$, we might expect us to be able to take Fourier transforms, so that

$$\begin{aligned}\int u(x)v(x)\phi(x)\psi(x)\, dx &= \int (\phi u)(x)(\psi v)(x)\, dx \\ &= \int\int (\widehat{\phi u} * \widehat{\psi v})(\xi) e^{2\pi i \xi \cdot x}\, d\xi\, dx \\ &= \int\int \widehat{\phi u}(\eta)\widehat{\psi v}(\xi - \eta) e^{2\pi i \xi \cdot x}\, d\xi\, dx.\end{aligned}$$

The only problem with taking this as the *definition* of the product is that the integral we have obtained might not converge in general. However, if at least one of the Fourier transforms decreases rapidly in the right directions.

Theorem 10.10. *Fix conic sets $\Gamma_1, \Gamma_2 \subset \Omega \times \mathbf{R}^d - \{0\}$. If $\Gamma_3 = \Gamma_1 + \Gamma_2$ does not contain any points in $0_\Omega = \Omega \times \{0\}$, then we have a unique continuous map from $\mathcal{D}^*_{\Gamma_1}(\Omega) \times \mathcal{D}^*_{\Gamma_2}(\Omega) \to \mathcal{D}^*_{\Gamma_3}(\Omega)$ which agrees with multiplication for elements of $C^\infty(\Omega)$.*

Example. *Consider the distributions Λ_1 and Λ_2 on $\mathbf{R}^2$, given by integration along the x and y axis respectively, i.e.*

$$\int \Lambda_1(x,y)\phi(x,y)\,dx\,dy = \int \phi(x,0)\,dx$$

and

$$\int \Lambda_2(x,y)\phi(x,y)\,dx\,dy = \int \phi(0,y)\,dy.$$

We have seen that $WF(\Lambda_1) = \{(x,0;0,\eta) : \eta \neq 0\}$ and $WF(\Lambda_2) = \{(0,\xi;y,0) : \xi \neq 0\}$. Now

$$WF(\Lambda_1) + WF(\Lambda_2) = \{(x,\xi,y,\eta) : \xi,\eta \neq 0\},$$

which is disjoint from $0_{\mathbf{R}^2}$, and so a product $\Lambda_1 \cdot \Lambda_2$ is well defined. To determine what the product is, we consider a non-negative bump function $\phi \in \mathcal{D}(RR^d)$ equal to one in a neighborhood of the origin, and define

$$\phi_{x,\delta}(x,y) = (1/2\delta)\mathbf{I}(|x| \leqslant 1/\delta, |y| \leqslant \delta)$$

$$\phi_{y,\delta}(x,y) = (1/2\delta)\mathbf{I}(|x| \leqslant \delta, |y| \leqslant 1/\delta).$$

Then as $\delta \to 0$, $\phi_{x,\delta} \to \Lambda_1$ and $\phi_{y,\delta} \to \Lambda_2$. We find that

$$\phi_{x,\delta}\phi_{y,\delta} = (1/4\delta^2)\mathbf{I}(|x| \leqslant \delta, |y| \leqslant \delta).$$

As $\delta \to 0$, $\phi_{x,\delta}\phi_{y,\delta}$ thus converges to the Dirac delta distribution δ at the origin. Thus $\Lambda_1 \cdot \Lambda_2 = \delta$.

Example. *Let $\Lambda = (x + i0^+)^{-1}$, i.e. the distribution*

$$\int \Lambda(x)\phi(x)\,dx = \lim_{y \to 0^+} \int \frac{\phi(x)}{x + iy}\,dx = \lim_{y \to 0^+} \Lambda_y(\phi).$$

The Λ is homogeneous. Moreover, some formal manipulations, plus some contour integrals, show that

$$\widehat{\Lambda}(\xi) = -2\pi i \cdot \mathbf{I}(\xi < 0).$$

In particular, $WF(\Lambda) = \{(0,\xi) : \xi < 0\}$. this means $WF(\Lambda) + WF(\Lambda)$ does not contain any zero vectors, so the product $\Lambda \cdot \Lambda$ is well defined. Now Λ is *the limit of the C^∞ functions $\phi_y(x) = 1/(x+iy)$ in $\mathcal{D}^*(\mathbf{R})$, and it requires only a simple calculation to show that Λ is also the limit in $\mathcal{D}^*_\Gamma(\mathbf{R})$, where $\Gamma = \{(0,\xi) : \xi < 0\}$. Since*

$$\phi_y(x)^2 = 1/(x+iy)^2,$$

we find by continuity that

$$\Lambda \cdot \Lambda = (x + i0^+)^{-2},$$

i.e.

$$\int \Lambda(x)\Lambda(x)\phi(x)\, dx = \lim_{y \to 0} \int \frac{\phi(x)}{(x+iy)^2}\, dx.$$

To define more sophisticated operations on distributions, we define the generic operations of *pullback*, *pushforward*, and *tensoring*. Intuitively, the pullback of a distribution gives a way to 'compose' a distribution with a smooth function in the domain, the push forward enables one to 'integrate a distribution along fibres', and tensoring enables us to take the product of distributions.

Let us begin with the pullback. For a smooth map $f : \Omega \to \Psi$, not necessarily a diffeomorphism, and $\phi \in \mathcal{D}(\Psi)$, we can define $f^*\phi = \phi \circ f \in C^\infty(\Omega)$. This map is continuous in the appropriate topology, and if f is a proper map, f^* is continuous from $\mathcal{D}(\Psi) \to \mathcal{D}(\Omega)$. To obtain a distributional definition, we apply the Fourier inversion formula; if $\psi \in \mathcal{D}(\mathbf{R}^d)$, then

$$\int (f^*\phi)(x)\psi(x)\, dx = \int \phi(f(x))\psi(x)\, dx = \int\int \widehat{\phi}(\eta)\psi(x)e^{2\pi i \eta \cdot f(x)}\, d\xi\, dx.$$

For a compactly supported distribution u on Ψ, it is therefore natural to define f^*u on Ω such that

$$\int (f^*u)(x)\psi(x)\, dx = \int \widehat{u}(\eta) \left(\int \psi(x) e^{2\pi i \eta \cdot f(x)}\, dx \right) d\eta.$$

We can decompose this integral so that ψ is supported on various small sets. If ψ is supported in a neighbourhood of x_0, then the oscillatory integral on the inside decays fast as $\eta \to \infty$ provided that $Df(x_0)^T \eta \neq 0$. Thus, provided that $\widehat{u}(\eta)$ decays fast whenever $Df(x_0)^T\eta = 0$, the integral above is well defined. Proceeding through this argument more rigorously gives the following result, left as an exercise.

Theorem 10.11. *Given a smooth map $f : \Omega \to \Psi$, let*

$$N = \{(f(x), \eta) : Df(x)^T \eta = 0\}.$$

Fix a closed cone Γ with $\Gamma \cap N = \varnothing$. Then $f^* : \mathcal{D}(\Psi) \to C^\infty(\Omega)$ extends to a continuous map from $\mathcal{D}^*_\Gamma(\Psi) \to \mathcal{D}^*_{f^*\Gamma}(\Omega)$, where

$$f^*\Gamma = \{(x, Df(x)^T \xi) : (f(x), \xi) \in \Gamma\}.$$

Example. *if $\pi_x : \mathbf{R}^2 \to \mathbf{R}$ and $\pi_y : \mathbf{R}^2 \to \mathbf{R}$ are the obvious projection maps, then we have*

$$\begin{aligned}\int (\pi_x^* \delta)(x,y)\phi(x,y)\, dx\, dy &= \int \widehat{\delta}(\xi)\phi(x,y)e^{2\pi i \xi \cdot x}\, dx\, dy\, d\xi \\ &= \int \phi(x,y) e^{2\pi i \xi \cdot x}\, dx\, dy\, d\xi \\ &= \int \phi(0,y)\, dy.\end{aligned}$$

Thus $\pi_x^*\delta$ is the distribution given by integration on the y-axis. Similarily, one can calculate that $\pi_y^*\delta$ is the distribution given by integration on the x-axis. It is simple to calculate explicitly, or using the properties of pullback, that

$$WF(\pi_x^*\delta) \subset \{(0, \xi, y, 0) : \xi \neq 0\}$$

and

$$WF(\pi_y^*\delta) \subset \{(x, 0, 0, \eta) : \eta \neq 0\}.$$

In fact, in these two cases these equations are equalities.

Next, let us define the tensor product. Given a distribution u_1 on Ω_1 and a distribution u_2 on Ω_2, we define a distribution $u_1 \otimes u_2$ on $\Omega_1 \times \Omega_2$ such that for $\phi \in \mathcal{D}(\Omega_1 \times \Omega_2)$,

$$\int (u_1 \otimes u_2)(x_1, x_2)\phi(x_1, x_2)\, dx_1\, dx_2 = \int u_1(x_1) \left(\int u_2(x_2)\phi(x_1, x_2)\, dx_2 \right) dx_1,$$

where the function

$$\tilde{\phi}(x_1) = \int u_2(x_2)\phi(x_1, x_2)\, dx_2$$

is smooth, where one can easily verify that

$$D^\alpha \tilde{\phi}(x_1) = \int u_2(x_2) D^\alpha \phi(x_1, x_2)\, dx_2.$$

Thus the tensor product of any two distributions is well defined. It is simple to check that

$$\mathrm{WF}(u_1 \otimes u_2) \subset \mathrm{WF}(u_1) \times \mathrm{WF}(u_2) \cup \mathrm{WF}(u_1) \times \{0\} \cup \{0\} \times \mathrm{WF}(u_1).$$

obtained by isolating each variable separately with a bump function and then tensoring the Fourier transform.

Finally, we define the pushforward of a distribution. This is most naturally defined distributionally. Given a smooth map $f : \Omega \to \Psi$, $\phi \in \mathcal{D}(\Omega)$, and $\psi \in \mathcal{D}(\Psi)$, we define

$$\int f_* \phi(y) \psi(y) dy = \int \phi(x) \psi(f(x))\, dx.$$

Thus f_* is just the adjoint of f^*. One problem which prevents us from directly using this definition to extend the definition to distributions is that $\psi \circ f$ need not be compactly supported if ψ is compactly supported. But this technical issue is removed provided that f is a *proper map*, i.e. inverse images of compact sets are compact. It is then simple to define

$$\int f_* u(y) \psi(y)\, dy = \int u(x) \psi(f(x))\, dx.$$

for a distribution u on Ω and $\psi \in \mathcal{D}(\Psi)$. To understand the wavefront set of u, we consider a bump function ϕ supported in a neighbourhood on Ω and consider

$$\int f_*(u\phi)(y) e^{-2\pi i \eta \cdot y}\, dy = \int u(x) \phi(x) e^{-2\pi i \eta \cdot f(x)}\, dx.$$

We have already show that for such an oscillatory integral, provided that $(f(x_0), Df(x_0)^T \eta) \notin \mathrm{WF}(u)$, this integral converges. Thus

$$\mathrm{WF}(f_* u) \subset \{(y, \eta) : \text{There is } (x, \xi) \in \mathrm{WF}(u) \text{ and } Df(x)^T \eta = \xi\}.$$

Now we have defined pushforward, pullback, and tensoring, let us see how they can be used to define useful operations on distributions.

Example. *Given $\phi, \psi \in \mathcal{D}(\Omega)$, we have*

$$\phi \cdot \psi = i^*(\phi \otimes \psi),$$

where $i(x) = (x,x)$, which gives us another way to define the product of distributions by a tensoring, combined with a pullback.

Let us consider an important example which occurs in the theory of kernel operators. Recall that the Schwartz kernel theorem says that if $T : \mathcal{D}(\Omega) \to \mathcal{D}^*(\Psi)$ is any continuous linear map, then there exists a distribution $K \in \mathcal{D}^*(\Omega \times \Psi)$ such that for any $\phi \in \mathcal{D}(\Omega)$ and $\psi \in \mathcal{D}(\Psi)$,

$$\int T\phi(y)\psi(y)\, dy = \int K(x,y)\phi(x)\psi(y)\, dx\, dy.$$

In other words, if $\pi(x,y) = y$ and $\Delta(x,y) = (x,x,y)$, then

$$T\phi = \pi_*(\Delta^*(\phi \otimes K)). \tag{10.3}$$

Going through the definitions shows that $\text{WF}(T\phi) \subset \text{WF}(K)_Y$, where

$$\text{WF}(K)_Y = \{(y,\eta) : (x,0;y,\eta) \in \text{WF}(K) \text{ for some } x \in \Omega\}.$$

We can also use equation (10.3) to extend the domain of T to certain compactly supported distributions. Going through the definition shows that for a compactly supported distribution u, the expression $\pi_*(\Delta^*(\phi \otimes K))$ is well defined provided that

$$\text{WF}'_X(K) \cap \text{WF}(u) = \varnothing,$$

where

$$\text{WF}'(K)_X = \{(x,-\xi) : (x,\xi;y,0) \in \text{WF}(K) \text{ for some } y\}.$$

In this case, we define $Tu = \pi_*(\Delta^*(\phi \otimes K))$. This gives a sequentially continuous map from the subspace of compactly supported distributions in $\mathcal{D}^*_\Gamma(\Omega)$ to $\mathcal{D}^*(\Omega)$ for any Γ with $\text{WF}'_X(K) \cap \Gamma = \varnothing$. If, in addition, the projection map $\pi(x,y) = y$ is proper on $\text{supp}(K)$, then this can be extended to a sequentially continuous map from $\mathcal{D}^*_\Gamma(\Omega)$ to $\mathcal{D}^*(\Omega)$. It is a simple exercise to show that

$$\text{WF}(Tu) \subset \text{WF}_Y(K) \cup \text{WF}'(K) \circ \text{WF}(u)$$

where

$$\mathrm{WF}'(K) = \{(x, -\xi; y, \eta) : (x, \xi; y, \eta) \in \mathrm{WF}(K)\},$$

is the *wavefront relation* of K, and the composition $R \circ E$ of a subset E of $\Omega \times \mathbf{R}^n$ and a subset R of $(\Omega \times \Psi) \times (\mathbf{R}^n \times \mathbf{R}^m)$ is defined to be the set of $(y, \eta) \in \Psi \times \mathbf{R}^m$ such that $(x, \xi) \in E$ for some x and ξ, and $(x, \xi; y, \eta) \in R$. A simple way to remember the results of this construction is that K can be applied to any distribution u such that $\mathrm{WF}'(K) \circ (\mathrm{WF}(u) \cup 0_\Omega)$ contains no zero vector, and then $\mathrm{WF}(Ku)$ is equal to this composition.

Example. *Consider a Psuedodifferential operator T given by a symbol a, i.e.*

$$T\phi(y) = \int a(y, \xi)\widehat{\phi}(\xi) e^{2\pi i \xi \cdot y}\, d\xi = \int a(y, \xi)\phi(x) e^{2\pi i \xi \cdot (y - x)}\, d\xi\, dx.$$

We can also think of T as a kernel operator with kernel

$$K(x, y) = \int a(y, \xi) e^{2\pi i \xi \cdot (y - x)}\, d\xi.$$

The kernel is a distribution defined by an oscillatory integral distribution, and our calculations for such distributions show that

$$WF(K) \subset \{(x, -\xi; x, \xi) : x \in \Omega, \xi \in \mathbf{R}^n - \{0\}\}.$$

Thus

$$WF'(K) \subset \{(x, \xi; x, \xi) : x \in \Omega, \xi \in \mathbf{R}^n - \{0\}\}.$$

In particular, $WF'(K)$, viewed as a relation, contains no zero vectors, and so Tu is well defined for any compactly supported distribution u. Moreover, we find $WF(Tu) \subset WF(u)$. This is part of the pseudolocal *nature of pseudodifferential operators; when T is applied to some distribution u supported near (x_0, ξ_0) in phase space, we should expect the same will be true of Tu.*

Example. *Given a distribution u, convolution with u is given by the Schwartz kernel $K(x, y) = u(y - x)$, i.e. $K = f^* u$, where $f : \mathbf{R}^{2d} \to \mathbf{R}^d$ is given by $f(x, y) = y - x$. Since f is surjective, the resulting set N is empty, so the pullback K is always a well defined distribution. Moreover,*

$$WF(K) \subset f^* WF(u) = \{(x_1, -\xi, x_2, \xi) : (x_2 - x_1, \xi) \in WF(u)\},$$

and therefore

$$WF'(K) \subset \{(x_1, \xi, x_2, \xi) : (x_2 - x_1, \xi) \in WF(u)\}.$$

We actually have equality here. To see this, for $a \in \mathbf{R}^d$, and let $g : \mathbf{R}^d \to \mathbf{R}^{2d}$ such that $g(x) = (x+a, a)$. Then $u = g^* K$, and so it follows htat

$$WF(u) \subset \{(x,\xi) : (x+a, a, \xi, -\xi) \in WF(K)\}.$$

It follows from this that we have equality.

If u is a distribution supported at the origin, then $WF'(K)$ is a subset of the diagonal in $T^*X \times T^*X$. Thus if P is a linear differential operator, then $WF(Pu) \subset WF(u)$.

How about the *composition* of kernel operators? Intuitively, if $B : \mathcal{D}(\Psi) \to \mathcal{D}^*(\Phi)$, $A : \mathcal{D}(\Omega) \to \mathcal{D}^*(\Psi)$, with kernels $K_A(x,y)$ and $K_B(y,z)$, then, if we could define a kernel operator $C = B \circ A : \mathcal{D}(\Omega) \to \mathcal{D}^*(\Phi)$, then it should have kernel

$$K_C(x,z) = \int K_A(x,y) K_B(y,z)\, dy.$$

Slightly more precisely, we might want to define $K_C = \pi_* \Delta^*(K_A \otimes K_B)$, where $\Delta : X \times Y \times Z \to X \times Y \times Y \times Z$, and $\pi : X \times Y \times Z \to X \times Z$. In order for this to make sense, we require the projection map $(x,y) \to x$ to be proper on $\text{supp}(K_A)$. Provided that $\text{WF}'(K_B) \circ \text{WF'}(K_A) \circ 0_\Omega$ does not contain a zero section, the composition $B \circ A$ is well defined as an operator from $\mathcal{D}(\Omega) \to \mathcal{D}^*(\Phi)$, where the kernel is defined by the formula above. Moreover,

$$\text{WF}'(K_C) \subset \text{WF}'(K_B) \circ ((\text{WF}'(K_A) \cup 0_\Omega) \cup 0_\Psi).$$

For instance, this theorem says that if P and Q are pseudodifferential operators, then $P \circ Q$ is well defined, which is also pseudolocal in the sense that $\text{WF}((P \circ Q)(u)) \subset \text{WF}(u)$. Of course, one can show $P \circ Q$ is *also* a pseudodifferential operator, from which this result automatically follows from the previous results we talked about in this chapter.

Example. *Given a kernel $K \in \mathcal{D}^*(\Omega \times \Psi)$ generating an operator $T : \mathcal{D}(\mathbf{R}^n) \to \mathcal{D}^*(\mathbf{R}^m)$, the adjoint map $T^* : \mathcal{D}(\mathbf{R}^m) \to D'(\mathbf{R}^n)$ is induced by the kernel $K^*(y,x) = \overline{K(x,y)}$. It is simple to verify that*

$$WF'(K^*) = \{(y,\eta; x,\xi) : (x,\xi; y,\eta) \in WF(K)\}.$$

Thus provided the projection $(x,y) \mapsto x$ is proper on $\text{supp}(K)$, and

$$WF'(K^*) \circ WF'(K) \circ 0_\Omega$$

does not contain a zero section, i.e. there does not exist any $(x,0;y,\eta) \in WF'(K)$. If $(x,y) \mapsto x$ is proper, then we can define T^*T. Similarily, if there does not exist any $(x,\xi;y,0) \in WF'(K)$, and $(x,y) \mapsto y$ is proper, then we can define TT^*.

$WF'_Y(K^*) \cap WF_Y(K) = \varnothing$, we can define the kernel $K^* \circ K$. This means precisely that there doesn't exist any $(x,y,0,\eta) \in WF(K)$. We then calculate that

$$WF(K^* \circ K) \subset \{(x_1,x_2,\xi_1,\xi_2) : (x_1,y,\xi_1,\eta),(x_2,y,\xi_2,\eta) \in WF(K)\}$$

In particular, if $WF(K)$ is the graph of a function, then $WF(K \circ K^*)$ is a subset of the diagonal of T^*X.

10.4 Propogation of Singularities Theorem

One important relation between u and $\mathrm{WF}(u)$ is the *propogation of singularities theorem*. If u is a solution to a linear partial differential equation

$$\sum_{|\alpha| \leqslant K} a_\alpha(x)(\partial_\alpha u)(x) = v$$

where v is a distribution, then for any $(x,\xi) \in \mathrm{WF}(u) - \mathrm{WF}(v)$,

$$q(x,\xi) = \sum_{|\alpha| \leqslant K} a_\alpha(x)\xi^\alpha = 0,$$

and $\mathrm{WF}(u) - \mathrm{WF}(v)$ is invariant under the flow generated by the Hamiltonian vector field

$$H_{x,\xi} = \sum_{i=1}^d \frac{\partial q}{\partial x^j}\frac{\partial}{\partial \xi^j} - \frac{\partial q}{\partial \xi_j}\frac{\partial}{\partial x^j}.$$

As a particular example, if $u(t,x,y)$ is a distributional solution to the wave equation $u_{tt} = \Delta u$ and we let $v_t(x,y) = u(t,x,y)$, then $\Delta v_t = u_{tt}$, and so by the propogation of singularities theorem $\mathrm{WF}(v_t) \subset \mathrm{WF}(u_{tt})$.

Then the Paley-Wiener theorem implies that $\widehat{u}$ is an analytic function on $\mathbf{R}^d$. If $\widehat{u}$ decays rapidly, then u is also a smooth function. However, even if u is not smooth, $\widehat{u}$ may still decrease rapidly in certain directions, which implies that the singularities of u 'propogate' in certain directions and understanding these directions is often useful to understanding the distribution u. We can also get even more information about the distribution u by looking at the singular frequencies.

To begin with, let

To begin with, a distribution u is *nonsingular* at a point $x \in \mathbf{R}^d$ if u is locally a C^∞ function in a neighbourhood of x, i.e. there exists a bump function $\phi \in C^\infty(\mathbf{R}^d)$ with $\phi(x) \neq 0$ such that $\phi u \in C^\infty(\mathbf{R}^d)$. The *singular support* of a compactly supported distribution u to be the set of all points $x \in \mathbf{R}^d$ upon which u is not nonsingular.

Chapter 11

Symbol Classes

In various settings in harmonic analysis, especially generalizations of settings where *homogeneous functions* are prime examples under study, it is useful to study various *symbol classes*. For instance, psuedodifferential operators historically dealt with operators $a(x,D)$, where a is a function defined by a sum of homogeneous functions of various orders in the frequency variables. If the highest degree of the terms in the sum was α, then for any N and M, $\nabla_x^N \nabla_\theta^M a(x,\theta)$ is a sum of homogeneous functions, with highest degree $\alpha - M$. Thus we have bounds of the form

$$|\nabla_x^N \nabla_\theta^M a(x,\theta)| \lesssim \langle \theta \rangle^{\alpha - M}.$$

Given d, p, and an open subset U of $\mathbf{R}^d$, we define the *classical symbol class* $\mathcal{S}^\alpha(U \times \mathbf{R}^p)$ *of order* α as consisting of all functions $a \in C^\infty(U \times \mathbf{R}^p)$ such that

$$|\nabla_x^N \nabla_\theta^M a(x,\theta)| \lesssim_{N,M} \langle \theta \rangle^{\alpha - M}$$

holds *uniformly* in x. We take the optimal constants in these inequalities as a family of seminorms which gives $\mathcal{S}^\alpha(U \times \mathbf{R}^p)$ the structure of a Frechét space. The family of functions for which these bounds hold *locally uniformly* in x form the family of symbols $\mathcal{S}^\alpha_{\text{loc}}(U \times \mathbf{R}^p)$.

The classes $\mathcal{S}^\alpha(U \times \mathbf{R}^p)$ are decreasing as $\alpha \to -\infty$, and we define $\mathcal{S}^{-\infty}(U \times \mathbf{R}^p)$ to be the intersection of all these classes of symbols. Operators defined by such functions are often highly regular. For instance, a pseudodifferential operator defined by such a symbol is called a *smoothing operator*, and maps any compactly supported distribution to a smooth function. The class $\mathcal{S}^{-\infty}(U \times \mathbf{R}^p)$ is dense in any of the classes $\mathcal{S}^\alpha(U \times \mathbf{R}^p)$, since it contains any symbol compactly supported in θ, and we can take cutoffs as $\theta \to \infty$.

A useful strategy to understand a symbol is to break it down into an asymptotic series of simpler symbols. Suppose $\{a_n\}$ is a sequence of symbols, then we write

$$a \sim \sum_{n=0}^{\infty} a_n$$

for some symbol a, if for any $\alpha \in \mathbf{R}$, there exists N_0 such that for $N \geqslant N_0$, $a - \sum_{n=0}^{N} a_n$ is a symbol of order α. If a_n is a symbol of order α_n, and $\lim_{n\to\infty} \alpha_n = -\infty$, then a symbol a always exists satisfying these asymptotics.

Theorem 11.1. *Consider a sequence of symbols $\{a_n\}$, with $a_n \in \mathcal{S}^{\alpha_n}(U \times \mathbf{R}^p)$, where $\lim_{n\to\infty} a_n = -\infty$, and let $\alpha = \max \alpha_n$. Then there exists a symbol $a \in \mathcal{S}^{\alpha}(U \times \mathbf{R}^p)$ such that $a \sim \sum a_n$.*

Proof. Fix a bump function $\phi \in \mathcal{D}(\mathbf{R}^p)$ equal to 0 when $|x| \leqslant 1/2$, and equal to one when $|x| \geqslant 1$. Find a rapidly increasing sequence $\{r_n\}$ such that

$$|\nabla_x^j \nabla_\lambda^k \{\phi(\theta/r_n) a_n(x,\theta)\}| \leqslant 2^{-n} \langle\theta\rangle^{\alpha_n + 1 - k}$$

for $x \in U$, where $i, j \leqslant n$. We define

$$a(x,\theta) = \sum_{n=0}^{\infty} \phi(\theta/r_n) \cdot a_n(x,\theta),$$

which is smooth, since it is a locally finite sum. For any N, if we set

$$R_N(x,\theta) = \sum_{n=N}^{\infty} \phi(\theta/r_n) \cdot a_n(x,\theta),$$

then

$$a - \sum_{n=0}^{N-1} a_n = \sum_{n=0}^{N-1} (\phi(\theta/r_n) - 1) a_n(x,\theta) + R_N(x,\theta)$$

If $x \in U$, we find that

$$|\nabla_x^j \nabla_\lambda^k R_N(x,\theta)| \lesssim_{N,i,j} \langle\theta\rangle^{\max_{n \geqslant N} \alpha_n + 1 - k}.$$

Thus $R_N \in \mathcal{S}^{\beta_N}(U \times \mathbf{R}^p)$, where $\beta_N = \max_{n \geqslant N} \alpha_n + 1$. On the other hand,

$$E_N(x,\theta) = \sum_{n=0}^{N-1} (\phi(\theta/r_n) - 1) a_n(x,\theta)$$

vanishes for $|\theta| \geqslant r_n$, and is thus compactly supported in θ, which implies that $E_N \in \mathcal{S}^{-\infty}(U \times \mathbf{R}^p)$. □

Remark. A similar formula holds for local families of symbols.

To verify asymptotic formulae, the following Lemma is often helpful.

Lemma 11.2. *Suppose $a \in C^\infty(U \times \mathbf{R}^p)$, and for any $n, m > 0$, there exists α_{nm} such that*

$$|\nabla_x^n \nabla_\theta^m a(x,\theta)| \lesssim_{n,m} \langle\theta\rangle^{\alpha_{nm}}.$$

If, for any $\alpha \in \mathbf{R}$,

$$|a(x,\theta)| \lesssim_\alpha \langle\theta\rangle^\alpha,$$

then $a \in \mathcal{S}^{-\infty}(U \times \mathbf{R}^p)$.

Proof. We begin by showing that if $f \in C^2(\mathbf{R})$, $\|f\|_{L^\infty(\mathbf{R})} \leqslant A$, and $\|f''\|_{L^\infty(\mathbf{R})} \leqslant B$, then $\|f'\|_{L^\infty(\mathbf{R})} \leqslant \sqrt{2AB}$. this follows because for any x, and $\varepsilon > 0$, there exists θ_1 lying between x and $x - \varepsilon$ such that

$$f(x) - f(x-\varepsilon) = \varepsilon f'(x) + \varepsilon^2 f''(\theta_1)/2$$

and θ_2 lying between x and $x + \varepsilon$ such that

$$f(x+\varepsilon) - f(x) = \varepsilon f'(x) + \varepsilon^2 f''(\theta_2)/2.$$

Thus

$$f(x+\varepsilon) - f(x-\varepsilon) = 2\varepsilon f'(x) + \varepsilon^2/2(f''(\theta_1) + f''(\theta_2)).$$

Rearranging gives

$$f'(x) = (f(x+\varepsilon) - f(x-\varepsilon))/2\varepsilon - (\varepsilon/4)(f''(\theta_1) + f''(\theta_2)),$$

and thus

$$|f'(x)| \leqslant A/\varepsilon + B\varepsilon/2.$$

Taking $\varepsilon = \sqrt{2A/B}$ completes the proof.

It follows from this that if K and K' are compact sets, with K contained in the interior of K', then

$$\|\nabla_\theta \phi\|_{L^\infty(K)} \lesssim_K \sqrt{\|\phi\|_{L^\infty(K')} \|\nabla_\theta^2 \phi\|_{L^\infty(K'')}}.$$

The theorem then follows by successively differentiating in θ. □

Corollary 11.3. *Suppose $\{a_n\}$ are a family of symbols, with $a_n \in \mathcal{S}^{\alpha_n}(U \times \mathbf{R}^p)$ for each n, and $\lim_{n\to\infty} \alpha_n = -\infty$. Then if $a \in C^\infty(U \times \mathbf{R}^p)$, and for each N and M, there exists α_{NM} such that*

$$|\nabla_x^N \nabla_\theta^M a(x,\theta)| \lesssim \langle \theta \rangle^{\alpha_{NM}}.$$

If for each n, there exists β_n such that

$$|a(x,\xi) - \sum_{k=0}^n a_n(x,\xi)| \lesssim_n \langle \theta \rangle^{\beta_n},$$

and $\lim_{n\to\infty} \beta_n = -\infty$, then $a \sim \sum a_n$.

Sometimes one has to use more powerful notions of homogeneity than the simple decay estimates above. In this case, it is useful to focus on *classical symbols*, i.e. symbols which satisfy an asymptotic formula of the form

$$a(x,\xi) \sim \sum_{k=-\infty}^n a_k(x,\xi),$$

where $a_k \in \mathcal{S}^k(U \times \mathbf{R}^p)$ is homogeneous of order k in ξ. We denote the class of such symbols of order α by $\mathcal{S}^\alpha_{\text{cl}}(U \times \mathbf{R}^p)$.

Chapter 12

Pseudodifferential Operators

The goal of this chapter is to consider a general family of operators that manipulate space and time *locally*. This is of course impossible to do simultaneously because of the uncertainty principle, but one can do things *pseudolocally*, i.e. the position of the support in time and space is approximately preserved up to a rapidly decaying error. Before we begin, let us consider some basic examples that allow us to control space or time exclusively, to get an idea of what we want out of such a theory.

The theory of Fourier multipliers can be used to understand constant coefficient differential operators. The most basic spatial multiplier in Fourier analysis are the *position operators* $X^\alpha : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ given by

$$X^\alpha f(x) = x^\alpha f(x)$$

and the most basic Fourier multipliers are the *momentum operators*

$$D^\alpha f(x) = \frac{1}{(2\pi i)^{|\alpha|}} \cdot \partial^\alpha f(x),$$

which have the property that $\widehat{D^\alpha f}(\xi) = \xi^\alpha \widehat{f}(\xi)$. If $m \in C^\infty(\mathbf{R}^d)$ is given, such that m and all it's derivatives are slowly increasing, then we can define a continuous operator $m(X) : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ by setting

$$m(X)f(x) = m(x)f(x).$$

We refer to m as the *symbol* of the operator. Similarly, we can define an operator $m(D) : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ such that

$$\widehat{m(D)f}(\xi) = m(\xi)\widehat{f}(\xi).$$

These give two homomorphisms from the ring of functions m to the ring of continuous operators on $\mathcal{S}(\mathbf{R})$. The family of such operators is very useful in analysis, since families of functions are more amenable to intuition than families of operators; for instance, any constant coefficient differential operator can be represented as $m(D)$ for some polynomial $m(x)$. Thus we have a symbolic calculus for the family of operators $\{X^\alpha\}$ and $\{D^\alpha\}$. Our goal is to find a symbolic calculus which *combines* the family of operators, so that we can consider operator involving both spatial and frequential information.

As a quick utility of this symbolic calculus, given an *elliptic* order k differential operator $L = \sum c_\alpha D^\alpha$, i.e. an operator such that the homogeneous polynomial $\sum_{|\alpha|=k} c_\alpha \xi^\alpha$ has no zeroes except when $\xi = 0$, there exists $R > 0$ such that the associated multiplier $m(\xi) = \sum c_\alpha \xi^\alpha$ satisfies a bound $|m(\xi)| \gtrsim |\xi|^k$ for $|\xi| \geqslant R$. If $\eta(\xi)$ is a smooth cutoff supported on $|\xi| \geqslant R$ and equal to one for $|\xi| \geqslant 2R$, and if we consider the multiplier operator S associated with $m'(\xi) = \eta(\xi)/m(\xi)$, then the symbolic calculus of multiplier operators tells us that $S \circ L - 1$ is a multiplier with a smooth, compactly supported symbol $1 - \eta(\xi)$. This means we have found a *parametrix* for the operator L, i.e. an operator which gives an *approximate inverse* for L, such that for any tempered distribution u,

$$(S \circ L)u(x) = u(x) + (\psi * u)(x),$$

where ψ is a Schwartz distribution. One can imagine this is very useful in the analysis of the PDE L. For instance, this construction implies that for any elliptic operator T, and any compactly supported distribution $u \in \mathcal{E}^*(\mathbf{R}^d)$, $\text{WF}(Tu) = \text{WF}(u)$. One motivation for constructing a symbolic calculus for both families of operators simultaneously is to construct parametrices for elliptic differential operators with *non constant coefficients*.

If these two families commuted jointly, i.e. $X^\alpha D^\beta = D^\beta X^\alpha$ for all α and β, then using standard techniques in the theory of operator algebras, we could define a calculus that associates with each function $a(x,\xi)$ a 'multiplier operator' T_a such that $a \mapsto T_a$ gives a homomorphism between an algebra of functions and an algebra of operators. Unfortunately, $\{X^\alpha\}$ and $\{D^\alpha\}$ do not commute, since we have a commutator relationship

$$[X^i, D^j] = X^i D^j f - D^j X^i = \delta_{ij} \cdot f.$$

Thus the operators do not commute. Nonetheless, the operators X^i and D^j commute 'modulo' more regular operators, i.e. the difference between

$X^i D^j$ and $D^j X^i$, which are differential operators of order one, is a differential operator of order zero. The goal of the theory of pseudodifferential operators is to associate a symbolic calculus $a \mapsto a(x,D)$ which gives a homorphism 'modulo lower order terms'.

We now associate with each function $a(x,\xi)$ an operator $a(x,D)$, such that if $a(x,\xi) = \sum c_\alpha(x)\xi^\alpha$, then $a(x,D)$ is the differential operator $\sum c_\alpha(x)D^\alpha$ (from which the notation $a(x,D)$ comes from). This association will generalize the two families of operators above; if $a(x,\xi) = m(x)$, then $a(x,D)$ is the operator given by multiplication by m, and if $a(x,\xi) = m(\xi)$, then $a(x,D)$ is a Fourier multiplier operator with symbol $m(\xi)$. To get an idea for what this operator should look like, we calculate that if $a(x,\xi) = \sum_\alpha c_\alpha(x)\xi^\alpha$ is the symbol of a differential operator with nonconstant coefficients, then the corresponding differential operator satisfies

$$\begin{aligned} a(x,D)f &= \sum c_\alpha(x) D^\alpha f(x) \\ &= \int_{\mathbf{R}^d} \sum_\alpha c_\alpha(x)\xi^\alpha \widehat{f}(\xi) e^{2\pi i \xi\cdot x}\, d\xi \\ &= \int_{\mathbf{R}^d} a(x,\xi) e^{2\pi i \xi \cdot x} \widehat{f}(\xi)\, d\xi. \end{aligned}$$

This is the integral formula with which we will define a general pseudodifferential operator.

12.1 Basic Definitions

Consider any smooth function $a \in C^\infty(\mathbf{R}^d \times \mathbf{R}^d)$ such that for any $n_1, n_2, m \geqslant 0$, $|\nabla_x^{n_1} \nabla_y^{n_2} \nabla_\xi^m a(x,\xi)| \lesssim_{n,m} \langle \xi \rangle^{t-m}$, where the implicit constant is *uniform* in x. We call such a function a *symbol* of order t, and denote the family of such symbols by $\mathcal{S}^t(\mathbf{R}^d \times \mathbf{R}^d)$. From this function, we can define a continuous operator $T_a : \mathcal{S}(\mathbf{R}^n) \to \mathcal{S}'(\mathbf{R}^n)$ by setting

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

where, since $\widehat{f}$ decays rapidly in ξ, we can interpret the integral in the Lebesgue sense. Any operator T which can be given as T_a for some symbol a is called a *psuedo-differential operator*, or ΨDO for short.

A pseudodifferential operator is uniquely determined by it's symbol. Indeed, if T is a pseudodifferential operator specified by some symbol $a(x,\xi)$, ψ is a compactly supported smooth bump function with $\widehat{\psi}(0) = 1$, and $\psi_R(y) = \psi(y/R)$ then

$$T_a(\text{Mod}_{\xi_0}\psi_R)(x) = R^d \int \widehat{\psi}(R(\xi - \xi_0))a(x,\xi)e^{2\pi i \xi \cdot x}\, d\xi.$$

Thus the Fourier inversion formula implies that

$$a(x,\xi_0) = \lim_{R\to\infty} e^{-2\pi i \xi_0 \cdot x} T_a(\text{Mod}_{\xi_0}\psi_R)(x).$$

Varying ξ_0 gives the complete symbol a.

Unwinding the Fourier transform in the definition of $T_a f$, we can also view such the operator as defined by the kernel

$$K(x,y) = \int a(x,\xi)e^{2\pi i \xi \cdot (x-y)}\, d\xi,$$

where this oscillatory integral must be interpreted distributionally unless a is a symbol of low enough order. From our analysis of wavefront sets, we can simply calculate that

$$\text{WF}'(K) \subset \{(x,\xi;x,\xi) : x \in \text{ess supp}(a) \text{ and } \xi \neq 0\}.$$

Thus general microlocal analysis shows that the kernel K is smooth for $x \neq y$, and that pseudodifferential operators extend to continuous operators from $\mathcal{E}^*(\mathbf{R}^d) \to \mathcal{D}^*(\mathbf{R}^d)$.

pseudodifferential operators can be applied to any compactly supported distribution, and for any such distribution u, $\text{WF}(Tu) \subset \text{WF}(u)$. This is the first instance of the *psuedolocal nature* of psuedodifferential operators; these operators roughly preserve the position of the mass of a function, but with some 'fuzz'. Here is another quantitative estimate on the kernel also showing the pseudolocal nature, which implies that $T_a f$ is a *Schwartz function* whenever f is Schwartz.

Theorem 12.1. *Let K be the kernel given by a pseudodifferential operator of order t, and let $z = x - y$. Then for any non-negative integers n_1, n_2, and N, K is a smooth function away from the origin, and*

$$|\nabla_x^{n_1} \nabla_z^{n_2} K(x,y)| \lesssim_{n_1,n_2,N} \frac{1}{|x-y|^{t+d+n_2+N}},$$

provided $t + d + n_2 + N \geqslant 0$.

Proof. If a is compactly supported in ξ, then taking in absolute values in the integral representation of K gives

$$|\nabla_x^{n_1} \nabla_z^{n_2} K(x,y)| \lesssim 1,$$

so the case is easy. Thus, without loss of generality, in the remainder of the proof we may assume $a(x,\xi) = 0$ for $|\xi| \leqslant 1$. We can then perform a Littlewood-Paley decomposition, i.e. writing

$$a(x,\xi) = \sum_{n=0}^{\infty} a_n(x,y,\xi),$$

where a_n is supported on $|\xi| \sim 2^n$. If K_n is the kernel of the pseudodifferential operator corresponding to a_n, then

$$K(x,y) = \sum_{n=-\infty}^{\infty} K_n(x,y).$$

We claim that

$$|\nabla_x^{n_1} \nabla_z^{n_2} K_n(x,y)| \lesssim_{n,n_1,n_2,N} |x-y|^{-N} 2^{n(t+d+n_2-N)}.$$

This follows from a simple integration by parts, applied to the integral

$$K_n(x,y) = \int \rho(\xi/2^n) a(x,\xi) e^{2\pi i \xi \cdot (x-y)}\, d\xi.$$

Summing up these bounds for sufficiently large N gives the result for $|x-y| \geqslant 1$ (since for these values of x and y bounds for large N imply bounds for small N). For $0 < |x-y| \leqslant 1$, we break the sum into two parts, i.e. writing

$$K(x,y) = \sum_{2^n \leqslant 1/|x-y|} K_n(x,y) + \sum_{2^n > 1/|x-y|} K_n(x,y).$$

For the first sum, we take $N = 0$, and for the second sum, we take $N > t + d + n_2$, which gives the required bounds. □

In addition to studying the behaviour of ΨDOs away from the diagonal, which reflects the pseudolocal behaviour of the distribution, it is also of interest to determine the behaviour of the operator under highly oscillatory, but non-stationary, phenomena. Consider a symbol $a(x,\xi)$, a smooth function $f(y)$, and a smooth phase $\phi(y)$ with $\nabla\phi(y)$ nonvanishing on $\text{supp}_x(a)$. Our goal is to try to determine the asymptotic behaviour of the function $T_a(fe^{2\pi i\lambda\phi})$ as $\lambda \to \infty$. Since T_a is pseudolocal, the value at a point x should be determined to a large degree by the behaviour of $fe^{2\pi i\lambda\phi}$ near x, which, roughly speaking, oscillates near the frequency $\lambda\nabla\phi(x)$. Thus we might expect that

$$T_a\{fe^{2\pi i\lambda\phi}\}(x) \approx a(x,\lambda\nabla\phi(x))f(x)e^{2\pi i\lambda\phi(x)}.$$

This is correct, and in fact, we can obtain a complete asymptotic development as $\lambda \to \infty$. For simplicity, we assume $\text{supp}(a)$ is compact.

Theorem 12.2. *Fix a symbol $a(x,\xi)$ of order t, compactly supported in x, a smooth function $f \in C^\infty(\mathbf{R}^d)$, and a smooth, real-valued function ϕ with $\nabla\phi$ nonvanishing on $\text{supp}_x(a)$. Let $\phi_2(x,y) = \nabla\phi(x)\cdot(x-y)-(\phi(x)-\phi(y))$. Then for any N, we can write*

$$\begin{aligned}&e^{-2\pi i\lambda\phi(x)}T_a\{fe^{2\pi i\lambda\phi}\}(x)\\&= \sum_{|\beta|<N} \frac{1}{\beta!\cdot(2\pi i)^\beta}\cdot D_\xi^\beta a(x,\lambda\nabla\phi(x))\cdot D_y^\beta\{e^{2\pi i\lambda\phi_2(x,y)}f\}\Big|_{y=x} + R_N(x,\lambda),\end{aligned}$$

where $\lambda^{t-\lceil N/2\rceil}R_N \in L^\infty(\mathbf{R}^d\times(0,\infty))$.

In particular, for $N=3$, we find that

$$\begin{aligned}e^{-2\pi i\lambda\phi(x)}T_a\{fe^{2\pi i\lambda\phi}\}(x) = \sum_{|\beta|<2}&\frac{1}{\beta!\cdot(2\pi i)^\beta}D_\xi^\beta a(x,\lambda\nabla\phi(x))D_x^\beta f(x)\\&-(i\lambda/\pi)\sum_{|\beta|=2}D_\xi^\beta a(x,\lambda\nabla\phi(x))D_y^\beta\phi_2(x,x)f(x)\\&+O(\lambda^{t-2}).\end{aligned}$$

Proof. We write

$$e^{-2\pi i\lambda\phi(x)}T_a\{fe^{2\pi i\lambda\phi}\}(x) = \lambda^d\int e^{2\pi i\lambda((\xi-\nabla\phi(x))\cdot(x-y)-\phi_2(x,y))}a(x,\lambda\xi)f(y)\,d\xi\,dy.$$

Since $|\phi_2(x,y)| \lesssim |x-y|^2$, the principal contributions to this integral occurs when $y \approx x$ and $\xi \approx \nabla\phi(x)$. Without loss of generality, we may therefore assume that $a(x,\xi) = 0$ whenever it is not true that $|\xi| \sim \lambda$. We rewrite our integral as

$$\lambda^d \int e^{2\pi i \lambda(\xi\cdot(x-y) - \phi_2(x,y))} a(x, \lambda\nabla\phi(x) + \lambda\xi) f(y)\, d\xi\, dy.$$

Using Taylors formula, we write

$$a(x, \lambda\nabla\phi(x) + \lambda\xi) = \sum_{|\beta|<N} \frac{\lambda^\beta}{\beta!} D_\xi^\beta a(x, \lambda\nabla\phi(x))\xi^\beta + R_{N,\lambda}(x,\xi).$$

We know that $D_\xi^\beta R_{N,\lambda}(x,0) = 0$ for $|\beta| < N$. For any $\beta > 0$, we have $D_\xi^\beta a(x,\xi) \lesssim \lambda^{t-|\beta|}$, and the remainder formula for the Taylor expansion can therefore be used to show that $|D_\xi^\beta R_{N,\lambda}(x,\xi)| \lesssim \lambda^{t-|\beta|}$. Stationary phase thus implies that

$$\left| \lambda^d \int\int e^{2\pi i \lambda((x-y)\cdot\xi + \phi_2(x,y))} R_{N,\lambda}(x,\xi) f(y)\, d\xi\, dy \right| \lesssim \lambda^{t - \lceil N/2 \rceil}.$$

Finally, we note that via an integration by parts,

$$\begin{aligned} &\int e^{2\pi i \lambda(\xi\cdot(x-y) - \phi_2(x,y))} \xi^\beta f(y)\, d\xi\, dy \\ &\quad = (1/\lambda)^{d+\beta} (2\pi i)^{-\beta}\, D_y^\beta \{ e^{-2\pi i \lambda \phi_2(x,y)} f(y) \} \Big|_{y=x} \end{aligned}$$

and substituting them into the formula completes the proof. □

As the order of the symbol a decreases, we expect the behaviour of the corresponding Pseudodifferential operator to become more and more regular. In particular, the operator has order at most $t + d$. In particular, for $t < -d$, we get a distribution which is actually *locally integrable*. If a has order t, we say T_a is an operator of order t, whereas we define *the* order of an operator as the infinum of the order of the symbols that can be used to define a pseudodifferential operator. In this chapter, we will write

$$a \sim \sum_{k=0}^\infty a_k$$

if the order of the Pseudodifferential operator with symbols $a - \sum_{k=0}^N a_k$ tends to $-\infty$ as $N \to \infty$. Since an operator determined by a symbol $a(x,\xi)$ not depending on y is uniquely determined by such a symbol, and we will only deal with symbols of this form after a short introduction, this agrees with the asymptotic notation for symbols introduced previously.

A ΨDO of order $-\infty$ has a kernel K lying in $C^\infty(\mathbf{R}^d \times \mathbf{R}^d)$, and satisfying estimates of the form

$$|\nabla_x^{n_1} \nabla_y^{n_2} K(x,y)| \lesssim_{n,N} \frac{1}{\langle x - y \rangle^N}$$

for any $N \geqslant 0$. We will call such an operator of this form a *smoothing operator*. Any smoothing operator is a ΨDO of order $-\infty$, since if K satisfies these estimates, it is the ΨDO corresponding to the symbol

$$a(x,\xi) = e^{-2\pi i \xi \cdot x} \int K(x,y) e^{2\pi i \xi \cdot y}\, dy,$$

which is a symbol of order $-\infty$. A smoothing operator has the property that Tu is Schwartz for each tempered distribution u. For most purposes, it is convenient to work with pseudodifferential operators *modulo smoothing operators*, since smoothing operators often satisfy all the estimates we are interested in.

One might wish to study a more general family of operators with a *compound symbol* of the form $a(x,y,\xi)$, i.e. an operator of the form

$$T_a f(x) = \int a(x,y,\xi) e^{2\pi i \xi \cdot (x-y)} f(y)\, d\xi\, dy.$$

However, any such operator is already a pseudodifferential operator.

Lemma 12.3. *Given a symbol $a(x,y,\xi)$, we have*

$$a(x,y,\xi) \sim \sum_\beta \frac{1}{\beta!} \frac{1}{(2\pi i)^{|\beta|}} D_\xi^\beta D_y^\beta a(x,x,\xi).$$

Proof. Suppose a has order t. We perform a Taylor expansion, writing

$$a(x,y,\xi) = \sum_{|\beta| \leqslant N} \frac{1}{\beta!} D_y^\beta a(x,x,\xi) \cdot (y-x)^\beta + R_N(x,y,\xi),$$

where $D_y^\beta R_N(x,x,\xi) = 0$ for all $|\beta| \leqslant N$. This means we can find C^∞ functions $b_\beta(x,y,\xi)$, for $|\beta| = N+1$, such that

$$R_N(x,y,\xi) = \sum_{|\beta|=N+1} (2\pi i)^{N+1}(y-x)^\beta b_\beta(x,y,\xi).$$

Now integration by parts shows that

$$\begin{aligned}&\int R_N(x,y,\xi) e^{2\pi i \xi\cdot(x-y)}\, d\xi \\ &\quad = (-1)^{N+1} \sum_{|\alpha|=N+1} \int (D_\xi^\alpha b_\alpha)(x,y,\xi) e^{2\pi i \xi\cdot(x-y)}\, d\xi.\end{aligned}$$

The functions b_β are symbols of order t, so $D_\xi^\alpha b_\beta$ are symbols of order $t-(N+1)$. Thus the Pseudodifferential operator corresponding to R_N has order at most $t-(N+1)$. On the other hand, another integration by parts again shows that

$$\begin{aligned}&\int D_y^\beta a(x,x,\xi)\cdot(y-x)^\beta \cdot e^{2\pi i \xi\cdot(x-y)}\, d\xi \\ &\quad = \frac{1}{(2\pi i)^{|\beta|}} \int D_\xi^\beta D_y^\beta a(x,x,\xi) e^{2\pi i \xi\cdot(x-y)}\, d\xi.\end{aligned}$$

Thus the pseudodifferential operator corresponding to $D_y^\beta a(x,x,\xi)\cdot(y-x)^\beta$ also corresponds to the symbol $1/(2\pi i)^{|\beta|}\cdot D_\xi^\beta D_y^\beta a(x,x,\xi)$, which completes the proof. □

We can use this result to show the adjoint of a pseudodifferential operator $a(x,\xi)$; it is simple to calculate the adjoint is a pseudodifferential operator with compound symbol $(x,y,\xi) \mapsto \overline{a(y,\xi)}$. Nonetheless, the above theorem implies that the adjoint can be given by a symbol $a^*(x,\xi)$ where

$$a^*(x,\xi) \sim \sum_\beta \frac{1}{\beta!} \frac{1}{(2\pi i)^{|\beta|}} \overline{D_\xi^\beta D_x^\beta a(x,\xi)}.$$

In particular, if a is a symbol of order t, then $a^*(x,\xi) - \overline{a(x,\xi)}$ is a symbol of order $t-1$, which we might write as saying that $a^* \approx \overline{a}$, up to lower order terms. In particular, if a is a symbol correspond to a *self adjoint* pseudodifferential operator, then $a \approx \mathrm{Re}(a)$, up to lower order terms.

The choice of (x,ξ) variables is common, but certainly not standard. The association of the pseudodifferential operator with any symbol a in two variables is called the *Kohn-Niremberg quantization*. We could also use the *adjoint Kohn-Niremberg quantization* to associate an operator with every symbol a in two variables, using the (y,ξ) variables instead of the (x,ξ) variables. We find, using the expansion above, that modulo smoothing operators, any symbol in the (y,ξ) variables can be written in the (x,ξ) variables, and moreover,

$$a(y,\xi) \sim \sum_\beta \frac{1}{\beta!} \frac{1}{(2\pi i)^{|\beta|}} D_\xi^\beta D_x^\beta a(x,\xi).$$

In particular, $a(y,\xi) - a(x,\xi)$ corresponds to a pseudodifferential operator of order $t-1$. Thus the operator we get using either approach only matters up to lower order terms.

The family of operators one can describe via the adjoint Kohn-Niremberg quantization is the same as the Kohn-Niernberg quantization, modulo smoothing operators. Thus, in the sequel, there is no harm in sticking with the Kohn-Nirenberg quantization. On the other hand, the symbols representing various operators change. For instance, we previously found that under the Kohn-Nirenberg quantization, the symbol $a(x,\xi) = \sum c_\alpha(x)\xi^\alpha$ corresponded to the differential operator $Lf = \sum c_\alpha D^\alpha f$. Under the adjoint Kohn-Nirenberg quantization, the symbol $a(y,\xi) = \sum c_\alpha(y)\xi^\alpha$ corresponds to the differential operator $Lf = \sum D^\alpha(c_\alpha f)$. If t is the order of these operators, then the difference of these operators is a differential operator of order $t-1$, which reflects the equivalence described above.

Thus we see that, roughly speaking, the operators differ in the order in which they apply spatial and frequency modulation. It is sometimes useful to deal with a quantization that does both in a 'symmetric' manner. To do this, we introduce the *Weyl quantization*, which associates with each symbol a gives the Pseudodifferential operator T with compound symbol $(x,y,\xi) \mapsto a((x+y)/2,\xi)$. The quantization is again equal to the other quantizations, up to lower order terms. This is the approach that works best in a generalization of a functional calculus for any finite family of noncommuting operators (there are notes by Tao which describes this process in detail, but it is beyond the scope of these notes).

Remark. Here, we have worked with symbols satisfying uniform estimates in x. But often one can only work with symbols which *locally* satisfy these estimates in x, i.e. working in the symbol classes $\mathcal{S}^t_{\text{loc}}(\mathbf{R}^d \times \mathbf{R}^d)$. The kernels of operators formed from these symbols satisfy bounds of the form

$$|\nabla_x^{n_1} \nabla_z^{n_2} K(x,y)| \lesssim_{n_1,n_2,N} \frac{1}{|x-y|^{t+d+n_2+N}},$$

where the implicit constant is *locally uniform* in x, and uniform in y. On a related note, such operators can be applied to any compactly supported distribution, and satisfy the microlocalization statement $\text{WF}(Tu) \subset \text{WF}(u)$. On the other hand, unless one has a bound such as

$$|\nabla_x^{n_1} \nabla_y^{n_2} \nabla_\xi^m a(x,\xi)| \lesssim_{n,m} (\langle x \rangle^{k_{1n}} + \langle y \rangle^{k_{2n}}) \cdot \langle \xi \rangle^{k_{nm}},$$

for all n and m, it is not necessarily to apply the operator to Schwartz functions, and tempered distributions. One can consider asymptotics, as long as we work modulo a weaker family of smoothing operators, i.e. those whose kernels lie in $C^\infty(\mathbf{R}^d \times \mathbf{R}^d)$. TODO: CAN WE THINK OF THESE SMOOTHING OPERATORS AS PSEUDODIFFERENTIAL OPERATORS, OR NOT?

12.2 Compositions of ΨDOs, and Parametrices

The composition of a ΨDO of order t and a ΨDO of order s is a ΨDO of order $t+s$, and we have an asymptotic formula for the symbol of such an expansion, reflecting the lack of commutivity between the spatial and frequential variables. In particular, the symbol of the composition is, to first order, the product of the symbols of the two operators.

Theorem 12.4. *Let $a(x,\xi)$ and $b(x,\xi)$ be symbols of order t and s, corresponds to operators T_a and T_b. Then $T_a \circ T_b$ is a ΨDO of order $t+s$, and has symbol*

$$(a \circ b)(x,\xi) \sim \sum_\alpha \frac{1}{\alpha!} \frac{1}{(2\pi i)^{|\alpha|}} D_\xi^\alpha a(x,\xi) \cdot D_x^\alpha b(x,\xi).$$

Proof. We can write

$$\begin{aligned} (T_a \circ T_b) f(x) &= \int a(x,\eta) e^{2\pi i \eta \cdot (x-z)} T_b f(z)\, dz\, d\eta \\ &= \int a(x,\eta) b(z,\xi) e^{2\pi i (\eta - \xi)\cdot(x-z)} e^{2\pi i \xi \cdot (x-y)} f(y)\, dy\, dz\, d\xi\, d\eta. \end{aligned}$$

Thus we see that we can view the composition as a ΨDO with kernel

$$c(x,\xi) = \int\int a(x,\eta)b(z,\xi)e^{2\pi i(\eta-\xi)\cdot(x-z)}\, d\eta\, dz.$$

This is an oscillatory integral, with stationary point when $z = x$ and $\eta = \xi$. Thus we expand power series near this point, i.e. writing

$$a(x,\eta) = \sum_\alpha \frac{1}{\alpha!} D_\xi^\alpha a(x,\xi)(\eta-\xi)^\alpha$$

and

$$b(z,\xi) = \sum_\beta \frac{1}{\beta!} D_x^\beta b(x,\xi)(z-x)^\beta.$$

Using the Fourier inversion formula, we calculate that

$$\begin{aligned}
&\int (\eta-\xi)^\alpha (z-x)^\beta e^{2\pi i(\eta-\xi)\cdot(x-z)}\, d\eta\, dz \\
&\quad = \int \tau^\alpha y^\beta e^{-2\pi i\tau\cdot y}\, d\tau\, dy \\
&\quad = \begin{cases} 0 & \alpha \neq \beta, \\ \alpha!/(2\pi i)^\alpha & \alpha = \beta. \end{cases}
\end{aligned}$$

Working like in our analysis of compound symbols, it suffices to show that if g_1 and g_2 are symbols of order t and s, then

$$f(x,\xi) = \int\int (\eta-\xi)^\alpha (z-x)^\beta g_1(x,\eta)g_2(z,\xi)e^{2\pi i(\eta-\xi)\cdot(x-z)}\, d\eta\, dz$$

is a symbol of order $t+s-M-1$. Applying sufficiently many integration by parts, it actually suffices to show integrals of the form

$$f(x,\xi) = \int\int g_1(x,\eta)g_2(z,\xi)e^{2\pi i(\eta-\xi)\cdot(x-z)}\, d\eta\, dz,$$

have order $t+s$, where g_1 has order t, and g_2 has order s. We write $\lambda = |\xi|$, and $\xi = \lambda\tilde{\xi}$, and write

$$f(x,\xi) = \lambda^d \int\int g_1(x,\lambda\eta)g_2(z,\xi)e^{2\pi i\lambda(\eta-\tilde{\xi})\cdot(x-z)}\, d\eta$$

We can decompose the domain dyadically. For $|\eta| \leqslant 1/2$ and $|x - z| \leqslant 1$, an integration by parts in z gives rapid decay in t. Similarily, we can dyadically sum over the regions where $|\eta| \leqslant 1/2$ and $|x - z| \sim 2^k$ by first integrating in η using integration by parts, then integration in parts in z. This also gives rapid decay in t. Similar arguments give rapid decay in ξ for $|\eta| \sim 2^l$, in fact giving estimates which are summable in l. Thus we are left with giving decay for an integral of the form

$$t^d \int \int g_1(x, t\eta) g_2(z, \xi) \rho(|x - z|) \rho(|\eta| - 1) e^{2\pi i t(\eta - \tilde{\xi}) \cdot (x - z)} \, d\eta \, dz.$$

This domain has a stationary point when $\eta = \xi$ and $z = x$. However, the stationary point is nondegenerate. Thus the integral is $O(\lambda^{t+s} \lambda^{-d})$ and so $|f(x, \xi)| \lesssim \langle \xi \rangle^{t+s}$. Replacing g_1 and g_2 with appropriate derivatives gives a full argument that f is a symbol of order $t + s$. $\square$

A *parametrix* for a pseudodifferential operator T is a pseudodifferential operator S such that $S \circ T$ is the identity operator, modulo smoothing. One useful result of our calculations is that we can easily construct *parametrices* for suitable pseudodifferential. Suppose a is a symbol of order t such that there exists $R > 0$ such that for $|\xi| \geqslant R$,

$$C_1 (1 + |\xi|)^t \leqslant |a(x, \xi)| \leqslant C_2 (1 + |\xi|)^t.$$

We call the operator corresponding to such a symbol *elliptic*. Then there exists a smooth cutoff function $\psi(\xi)$ equal to one for $|\xi| \leqslant R$ such that

$$b_0(x, \xi) = (1 - \psi(\xi)) \cdot \frac{1}{a(x, \xi)}$$

is well defined. It is simple to verify this is a symbol of order $-t$. We note that $a(x, \xi)$ agrees with $(1 - \psi(\xi)) a(x, \xi)$ up to a smoothing operator. By our composition formula, we conclude that there exists a symbol $c_1(x, \xi)$ of order -1 such that $(b_0 \circ a)(x, \xi) = 1 + c_1(x, \xi)$. If we set $b_1(x, \xi) = -(1 - \psi(\xi)) \cdot c_1(x, \xi) / a(x, \xi)$, then b_1 is a symbol order order $-t - 1$, and $((b_0 + b_1) \circ a)(x, \xi) = 1 + c_2(x, \xi)$, where c_2 is a symbol of order -2. Continuing this development, we obtain a sequence of operators $\{b_k\}$, where b_k is a symbol of order $-t - k$. Choosing a symbol

$$b \sim \sum_{k=0}^{\infty} b_k,$$

we conclude that

$$(b \circ a)(x, \xi) \sim 1,$$

which means that, b is a parametrix for a. Similarily, we can construct a symbol c such that $(a \circ c)(x, \xi) \sim 1$. But then

$$b = b \circ (a \circ c) = (b \circ a) \circ c = c$$

so $b = c$, up to a smoothing operator. Thus a left parameterix is automatically a right parameterix.

Remark. The condition that $|a(x, \xi)| \sim \langle \xi \rangle^t$ for large ξ is necessary in order to construct a parametrix of order $-t$. Without loss of generality, this is true since we can replace $a(x, \xi)$ with $a(x, \xi)(1 + |\xi|^2)^{-t/2}$ and $b(x, \xi)$ with $b(x, \xi)(1 + |\xi|^2)^{t/2}$ and still have parametrices. Indeed, if a parameterix $b(x, \xi)$ for $a(x, \xi)$ exists, then, the composition formula implies that $a(x, \xi)b(x, \xi) - 1$ is a symbol of order -1. Thus for large ξ, we know that $|a(x, \xi)b(x, \xi) - 1| \leqslant 1/2$ and $|b(x, \xi)| \lesssim \langle \xi \rangle^{-t}$. But this implies that for large ξ, $|a(x, \xi)| \geqslant 1/2|b(x, \xi)| \gtrsim \langle \xi \rangle^t$.

12.3 Regularity Theory

Let us now discuss the boundedness of certain pseudodifferential operators with respect to various norm spaces. We first note that a differential operator of degree m given by

$$L = \sum c_\alpha(x) D^\alpha,$$

where c_α is bounded, then L maps $H^s(\mathbf{R}^d)$ to $H^{s-m}(\mathbf{R}^d)$ for each s. Thus we might expect the same to be true for pseudodifferential operators of order m. To begin with, we restrict ourselves to pseudodifferential operators of order 0. The kernel of a ΨDO of order zero satisfies estimates of the form

$$|K(x, y)| \lesssim \frac{1}{|x - y|^d}.$$

Thus we focus on obtaining $L^2 \to L^2$ estimates, so that the standard theory of singular integrals gives $L^p \to L^p$ estimates for all $1 < p < \infty$.

Theorem 12.5. *Let T_a be a pseudodifferential operator specified by a symbol a of order zero. Then for any $f \in \mathcal{S}$,*

$$\|T_a f\|_{L^2(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^d)}.$$

Proof. Let us begin by assuming that $\mathrm{supp}_x(a)$ is compact. Then we can apply the Fourier inversion formula, writing

$$a(x,\xi) = \int_{\mathbf{R}^d} \widehat{a}(\lambda,\xi) e^{2\pi i \lambda \cdot x}\, d\lambda$$

where

$$\widehat{a}(\lambda,\xi) = \int_{\mathbf{R}^d} a(x,\xi) e^{-2\pi i \lambda \cdot x}\, dx.$$

Since $\|\nabla_x^n a(x,\xi)\|_{L^\infty(\mathbf{R}^d)} \lesssim_n 1$ for all $n \geqslant 0$ and has uniform compact support in x, we find that

$$|\widehat{a}(\lambda,\xi)| \lesssim_n \langle \lambda \rangle^{-n}$$

for all $n > 0$. For $f \in \mathcal{S}(\mathbf{R}^d)$, we now write

$$\begin{aligned} T_a f(y) &= \int_{\mathbf{R}^d} a(y,\xi) \widehat{f}(\xi) e^{2\pi i \xi \cdot y}\, d\xi \\ &= \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \widehat{a}(\lambda,\xi) \widehat{f}(\xi) e^{2\pi i (\xi+\lambda)\cdot y}\, d\xi\, d\lambda \\ &= \int_{\mathbf{R}^d} T_a^\lambda f(y)\, d\lambda, \end{aligned}$$

where

$$T_a^\lambda f(y) = \int_{\mathbf{R}^d} \widehat{a}(\lambda,\xi) \widehat{f}(\xi) e^{2\pi i (\xi+\lambda)\cdot y}\, d\xi.$$

This is just a Fourier multiplier operator with symbol $m_\lambda(\xi) = \widehat{a}(\lambda,\xi) e^{2\pi i \lambda \cdot y}$. Since $\|m_\lambda\|_{L^\infty(\mathbf{R}^d)} \lesssim_n \langle \lambda \rangle^{-n}$, if $n > d$ we conclude that

$$\|T_a f\|_{L^2(\mathbf{R}^d)} \leqslant \int_{\mathbf{R}^d} \|T_a^\lambda f\|_{L^2(\mathbf{R}^d)}\, d\lambda \lesssim_n \|f\|_{L^2(\mathbf{R}^d)} \int_{\mathbf{R}^d} \langle \lambda \rangle^{-n}\, d\lambda \lesssim \|f\|_{L^2(\mathbf{R}^d)}.$$

Thus the theorem is proved, at least in the case of symbols compactly supported in the spatial domain.

To prove the result for more general symbols, we work with a kernel representation of T_a. Thus we write

$$T_a f(x) = \int_{\mathbf{R}^d} K(x,y) f(x)\, dx,$$

where

$$K(x,y) = \int_{\mathbf{R}^d} a(x,\xi) e^{2\pi i \xi \cdot (x-y)}\, d\xi.$$

We have already shown that the kernel K is C^∞ away from the diagonal, and decays rapidly away from the diagonal. This is one instance of the pseudolocal nature of these operators. Another quantitative result reflecting this nature is that for each $N > 0$ and $x_0 \in \mathbf{R}^d$,

$$\int_{|x-x_0|\leqslant 1} |T_a f(x)|^2\, dx \lesssim_N \int_{\mathbf{R}^d} \frac{|f(x)|^2}{\langle x - x_0 \rangle^N}\, dx.$$

Thus we can 'almost' bound the magnitude of $T_a f$ in a neighbourhood of x_0 by the magnitude of f in a neighbourhood of x_0. We focus on the case $x_0 = 0$, the other cases treated in much the same way. Write $f = f_1 + f_2$, where f_1 is supported on $|x| \leqslant 3$, f_2 is supported on $|x| \geqslant 2$, and $|f_1|, |f_2| \leqslant |f|$. If $\eta(x)$ is a smooth cuttoff supported on $|x| \leqslant 3$, then the symbol $\eta(x) a(x,\xi)$ is compactly supported, and so

$$\begin{aligned}
\int_{|x|\leqslant 1} |T_a f_1(x)|^2\, dx = \int_{|x|\leqslant 1} |T_{\eta a} f_1(x)|^2 &\lesssim \|f_1\|_{L^2(\mathbf{R}^d)}^2 \\
&\lesssim_N \int_{\mathbf{R}^d} \frac{|f_1(x)|^2}{\langle x \rangle^N}\, dx \leqslant \int_{\mathbf{R}^d} \frac{|f(x)|^2}{\langle x \rangle^N}\, dx.
\end{aligned}$$

On the other hand, since $f_2(x)$ is supported on $|x| \geqslant 2$, we find that

$$\begin{aligned}
\int_{|x|\leqslant 1} |T_a f_2(x)|^2\, dx &= \int_{|x|\leqslant 1} \left| \int K(x,y) f_2(y)\, dy \right|^2 dx \\
&\leqslant \int \int_{|x|\leqslant 1} |K(x,y)|^2 |f_2(y)|^2\, dy\, dx \\
&\lesssim_N \int \int_{|x|\leqslant 1} \frac{|f_2(y)|^2}{|x-y|^N}\, dy\, dx \\
&\lesssim \int \int_{|x|\leqslant 1} \frac{|f_2(y)|^2}{\langle y \rangle^N}\, dy \\
&\lesssim \int \frac{|f_2(y)|^2}{\langle y \rangle^N}\, dy.
\end{aligned}$$

But we now find that if $N > d$, then

$$\begin{aligned}
\int |T_a f(x)|^2\, dx &\lesssim \int \int_{|x-y|\leqslant 1} |T_a f(y)|^2\, dy\, dx \\
&\lesssim_N \int \int \frac{|f(y)|^2}{\langle x-y \rangle^N}\, dy\, dx \\
&\lesssim \int |f(y)|^2\, dy,
\end{aligned}$$

which gives L^2 boundedness. □

Sobolev norms follow simply from these bounds. Namely, it follows simply from this that if $a(x,\xi)$ is a symbol of order t, then for $1 < p < \infty$, and any s, we have bounds of the form

$$\|T_a f\|_{L^p_s(\mathbf{R}^d)} \lesssim_{p,s} \|f\|_{L^p_{t+s}(\mathbf{R}^d)}.$$

In particular, for $1 < p < \infty$, if T is an elliptic pseudodifferential operator of order t, then T is *almost* invertible. We cannot quite determine that $\|Tf\|_{L^p_s(\mathbf{R}^d)} \sim_{p,s} \|f\|_{L^p_{t+s}(\mathbf{R}^d)}$. But we can obtain a less quantiative result.

Theorem 12.6. *Let T be an elliptic pseudodifferential operator of order t. If f is a compactly supported distribution, and Tf lies in L^p_s, then f lies in L^p_{s+t}.*

Proof. Since T is elliptic, we can find a parametrix S, which is a ΨDO of order $-t$. Thus there exists a smoothing operator U such that $1 = ST + U$. Since U is smoothing, Uf lies in L^p_{s+t}, and $\|STf\|_{L^p_{s+t}} \lesssim \|Tf\|_{L^p_s}$, which implies $f \in L^p_{s+t}$. $\square$

12.4 Pseudodifferential Operators on Manifolds

It is an important fact that the class of *compactly supported* Pseudodifferential operators are invariant under a change of coordinates, modulo smoothing operators. More precisely, if T is a pseudodifferential operator on an open set $U \subset \mathbf{R}^n$, of order t, compactly supported in x, and $\kappa : U \to V$ is a diffeomorphism, then there is a pseudodifferential operator S of order t, compactly supported on V such that, modulo smoothing operators,

$$S \circ \kappa_* = \kappa_* \circ T.$$

Moreover, if $a(x, \xi)$ is a symbol representing T, and $b(y, \xi)$ is a symbol representing S, then

$$b(y, \eta) - a(\kappa^{-1}(y), D\kappa(x)^{-T} \cdot \eta)$$

is a pseudodifferential operator of order $t - 1$.

Theorem 12.7. *Let U and V be open subsets of Euclidean space, and let $\kappa : U \to V$ be a diffeomorphism. If $a(x, \xi)$ is a symbol of order m, and the kernel of $a(x, D)$ is compactly supported in U, and if*

$$a_\kappa(y, \eta) = e^{-2\pi i \kappa^{-1}(y) \cdot \eta} a(x, D) e^{2\pi i y \cdot \eta}.$$

TODO (Hörmander's book seems to have the most readable discussion)

In light of this, given a manifold M, a continuous operator $T : C_c^\infty(M) \to C^\infty(M)$ is called a *pseudodifferential operator of order t* if whenever (x, U) is a coordinate chart on M, $\psi_0, \psi_1 \in C_c^\infty(U)$, and $\psi_1 = 1$ on a neighborhood of the support of ψ_0, then the operator 'in coordinates', i.e. the operator $T_x : C_c^\infty(x(U)) \to C^\infty(x(U))$ given by

$$T_x f = (\psi_0 \cdot T(\psi_1 \cdot (f \circ x))) \circ x^{-1}$$

is a pseudodifferential operator of order t on $x(U)$. We let $\text{Op}(\mathcal{S}^t(\mathbf{R}^n))$ denote the family of operators of this form. The next Lemma shows this does not enlarge the family of pseudodifferential operators formed from *local symbols* $\mathcal{S}^t_{\text{loc}}(U \times \mathbf{R}^n)$.

Lemma 12.8. *Suppose $T : C_c^\infty(U) \to C^\infty(U)$ is a pseudodifferential operator as above. Then we can find a symbol $a(x,\xi) \in \mathcal{S}^t_{loc}(U \times \mathbf{R}^d)$ such that the kernel of $T = T_a + S$, where S has a kernel in $C^\infty(U \times U)$. The symbol a is uniquely determined up to a symbol in $\mathcal{S}^{-\infty}(\mathbf{R}^d \times \mathbf{R}^d)$.*

Proof. The idea is to work on a partition of unity, which we can sum up appropriately to get a sum over local estimates. The complete proof is supplied in Hormander, Proposition 18.1.19. □

Remark. It suffices to verify the condition defining a pseudodifferential operator for $\{\psi_i\}$ ranging over an atlas of the manifold M. If we assume that the operator has a kernel smooth away from the diagonal, it suffices to only verify it for maps $\psi_1 = \psi_2$ ranging over an atlas.

It is often to discuss operators that can be asymptotically expanded in terms of homogeneous symbols. That is, a symbol a of order t is classical if there exist a sequence of symbols $\{a_k\}$, where $a_k(x,\xi)$ is homogeneous of order $t-k$, and

$$a \sim \sum_{k=0}^{\infty} a_k.$$

We call such operators *classical pseudodifferential operators*, since this was the context first studied by Kohn and Nirenberg. In particular, given a manifold M, we write $\Psi^t_{\text{cl}}(M)$ for the subset of all operators T in Ψ such that T_x is classical for every coordinate system (x,U) on M. The leading term in this asymptotic expansion is invariant under coordinate changes, so for a classical pseudodifferential operator T in $\Psi^t_{\text{cl}}(M)$, we can define the *principal symbol* $a \in C^\infty(T^*M - 0)$ to be the function which gives the leading term in coordinates (for nonclassical operators, there is no canonical choice of a principal symbol, though one can consider an equivalence class of such functions modulo lower order functions). The operator is called *elliptic* if a is nonvanishing.

Given a scalar density $d\omega$, we say a classical pseudodifferential operator T is self adjoint if for any $f,g \in C^\infty(M)$,

$$\int Tf(x)\overline{g(x)}d\omega(x) = \int f(x)\overline{Tg(x)}d\omega(x).$$

We can always choose smooth coordinates around any point such that $d\omega$ agrees with the Lebesgue measure. But this means that if T is a self adjoint elliptic classical pseudodifferential operator, then its principal symbol is real. In particular, it must be either positive everywhere, or negative everywhere.

12.5 Self-Adjoint Pseudo-Differential Operators

A pseudodifferential operator T of order t is *elliptic* if it's symbol $a(x,\xi)$ satisfies an estimate

$$a(x,\xi) \gtrsim \langle|\xi|\rangle^m,$$

where the implicit constant is locally uniform in x.

12.6 Self-Adjoint Elliptic Pseudo-Differential Operators on Compact Manifolds

Let M be a compact manifold, and let T be a elliptic, self-adjoint classical pseudodifferential operator on M, which for simplicity we assume has order one. Without loss of generality, we may assume the principal symbol $a(x,\xi)$ associated with T is non-negative on T^*M. If S is an elliptic operator with principal symbol $a(x,\xi)^{1/2}$, then $T - S^*S$ is an operator of order zero. Thus for $u \in L^2(M)$,

$$\left| \int_M Tu(x)\overline{u}(x)\, dx - \|Su\|_{L^2(M)}^2 \right| = \left| \int_M (T - S^*S)u(x)\overline{u}(x)\, dx \right| \lesssim \|u\|_{L^2(M)}^2.$$

Since S is an operator of order $1/2$, if S' is a parameterix for S, then it has order $-1/2$, which means that for $u \in L^2_{1/2}(M)$,

$$\|u\|_{L^2_{1/2}(M)} \lesssim \|S'Su\|_{L^2_{1/2}(M)} + \|u\|_{L^2(M)}^2 \lesssim \|Su\|_{L^2(M)}^2 + \|u\|_{L^2(M)}^2.$$

Combining these two inequalities, we may find C_1 and C_2 such that

$$\|u\|_{L^2_{1/2}(M)} \leqslant C_1 \left| \int_M Tu(x)\overline{u}(x)\, dx \right| + C_2 \|u\|_{L^2(M)}^2.$$

Then if $c > 0$ is large enough, we conclude that

$$\|u\|^2_{L^2_{1/2}(M)} \leqslant C_1 \int_M (T+c)u(x)\overline{u}(x)\,dx,$$

Thus we have

$$\|(T+c)u\|_{L^2(M)} \sim \|u\|_{L^2_{1/2}(M)}.$$

In particular, $\|(T+c)u\|_{L^2(M)} \sim \|u\|_{L^2_{1/2}(M)}$, since the upper bound follows from Cauchy Schwartz. It follows from the open mapping theorem that $T+c$ is an isomorphism of Banach spaces. In particular, it has a bounded inverse from $L^2(M) \to L^2_{1/2}(M)$, which, after composition with the map $L^2_{1/2}(M) \to L^2(M)$, we will denote by $S : L^2(M) \to L^2(M)$. It follows from the Rellich-Kondrachov embedding theorem that S is a *compact*, self adjoint operator. Thus we can find an orthogonal basis $\{e_i\}$ for $L^2(M)$ and a family of eigenvalues $\{\lambda_i\}$ such that $Se_i = e_i/\lambda_i$. Then $e_i \in L^2_{1/2}(M)$ for each i, and we have $Te_i = (\lambda_i - c)e_i$. Since $T+c$ is positive definite, $\lambda_i > 0$ for all i, so ∞ can be the only limit point, and we can rearrange our eigenfunctions so that the values $\{\lambda_i\}$ are monotonically increasing.

We actually claim that that each of the eigenfunctions is an element of $C^\infty(M)$. But this follows because for each k, because under the assumptions we can use the same argument as above to obtain a bound

$$\|(T+c)^k u\|_{L^2(M)} \sim_k \|u\|_{L^2_k(M)}.$$

TODO. The eigenvalues for $T+c$ are also eigenvalues for $(T+c)^k$, and from this we obtain that

$$\|e_i\|_{L^2_k(M)} \sim_k \lambda_i^k.$$

Thus each of the functions $\{e_i\}$ lies in all of the Sobolev spaces for arbitrarily large k, which implies that the eigenfunctions actually lie in $C^\infty(M)$. In particular, the spectral theorem in this setting implies that $\{e_i\}$ remain orthogonal in $L^2_k(M)$, diagonalize $(T+c)^k$, and so if $f \in L^2_k(M)$,

$$\sum_{i=1}^\infty |(f,e_i)|^2 \lambda_i^{-2k} < \infty.$$

In particular, if $f \in C^\infty(M)$, then for any N, $|(f,e_i)| \lesssim_N \lambda_i^{-N}$.

12.7 The Half Wave Operator

Given a pseudodifferential operator T of order one, we consider the *half-wave equation*

$$\partial/\partial t + 2\pi i T = 0.$$

Studying this equation is useful to understanding the operator T. In particular, it helps us understand the Fourier transform of T, i.e. the family of operators $e^{-2\pi i t T}$, since this family of operators forms a fundamental solution to the half-wave equation. We will focus on the case where T has order one, is self-adjoint, formally positive, in the sense that for any $f \in C_c^\infty(\mathbf{R})$, $(Tf, f) > 0$, and is defined on a compact manifold M.

To understand this operator, our goal is to construct an approximate integral formula for the family of operators $e^{-2\pi i t T}$. That is, we wish to find an operator $S(t)$ such that $S(t) \approx e^{-2\pi i t T}$. Let $a_1(x,\xi)$ be the principal symbol for T. We shall find that, for $|t| \ll 1$, we can find a parametrix $S(t)$ for $e^{-2\pi i t T}$ given by a Fourier integral, i.e. such that for $f \in C_c^\infty(\mathbf{R}^d)$,

$$S(t)f(x) = \int s(t,x,y,\xi) e^{2\pi i [\phi(x,y,\xi) - t a_1(y,\xi)]} f(y)\, dy\, d\xi,$$

where s is a symbol of order zero, ϕ is a symbol of order one, $\phi(x,y,\xi) \approx (x-y)\cdot\xi$ in the sense that for any β,

$$D_\xi^\beta \{\phi(x,y,\xi) - (x-y)\cdot\xi\} \lesssim_\beta |x-y|^2 |\xi|^{1-\beta},$$

and s has small enough support near the diagonal so that $\nabla_\xi \phi(x,y,\xi) \gtrsim |x-y|$ for x and y in the support of s.

Our first tool for studying this operator is to show that for $t = 0$, we actually get a *pseudodifferential operator*. Provided we can compute what this operator looks like in terms of the symbols involved in our expression, this will give constraints that will enable us to determine what s and ϕ should look like.

First, let us understand what conditions ensure that

$$(\partial/\partial t + 2\pi i T) \circ S(t)$$

is a smoothing operator. In local coordinates, if T has symbol $a(x,\xi)$, then

$$\begin{aligned} K(t,x,y) &= \int \frac{d}{dt} \left\{ s(t,x,y,\xi) e^{2\pi i (\phi(x,y,\xi) + t a_1(x,\xi))} \right\} d\xi \\ &\quad + 2\pi i \int\int a(x,\xi) s(t,z,y,\eta) e^{2\pi i (\xi\cdot(x-z) + \phi(z,y,\eta) + t a_1(z,\eta))}\, dz\, d\xi\, d\eta. \end{aligned}$$

If $\Phi(t,z,y,\eta) = \phi(z,y,\eta) - ta_1(y,\eta)$, then we can write the set term in the definition of the kernel as

$$K(t,x,y) = \int (\partial/\partial t + 2\pi i T)\{s(t,\cdot,y,\eta)e^{2\pi i\Phi(t,\cdot,y,\eta)}\}(x)\, d\eta.$$

At the beginning of our notes on pseudodifferential operators, we were able to build up asymptotics for applications of pseudodifferential operators to oscillating quantities. In fact, these asymptotics imply that

$$\begin{aligned} e^{-2\pi i\Phi(t,x,y,\eta)}&(2\pi i T)\{s(t,\cdot,y,\eta)e^{2\pi i\Phi(t,\cdot,y,\eta)}\}(x) \\ &= 2\pi i a_1(x,\nabla_x\Phi(t,x,y,\eta))\cdot s(t,x,y,\eta) \\ &= 2\pi i a_1(x,\nabla_x\phi(x,y,\eta))\cdot s(t,x,y,\eta), \end{aligned}$$

modulo symbols of order -1 in η. Thus we can write the kernel of $(\partial/\partial t + 2\pi i T)S(t)$ as

$$(x,y)\mapsto 2\pi i\int e^{2\pi i\Phi(t,x,y,\eta)}\left\{\Big(a_1(y,\eta)-a_1(x,\nabla_x\phi(t,x,y,\eta))\Big)s(t,x,y,\eta)+R\right\}d\eta,$$

where R is a symbol of order 0. In order for this to be smoothing, it is thus natural to apriori assume that $a_1(y,\eta) = a_1(x,\nabla_x\phi(x,y,\eta))$ for all triples (x,y,η). The theory of Hamilton-Jacobi equations shows that there exists a *unique* $\phi(x,y,\eta)$ with this property such that $\phi(x,y,\eta) = 0$ when $|x-y|\lesssim 1$ and $(x-y)\cdot\eta = 0$, and $\nabla_x\phi(x,y,\eta) = \eta$ when $x = y$. This implies that for $|x-y|\lesssim 1$,

$$\phi(x,y,\eta) = (x-y)\cdot\eta + O(|x-y|^2|\eta|)$$

Thus we have specified our phase function ϕ completely. There is a small inconvenience that ϕ might not be differentiable at the origin. However, this does not cause us a problem, since we can replace ϕ with a phase supported away from the origin by applying a smooth cutoff; this is no problem, because an operator with kernel

$$\int \rho(\eta)e2\pi i\Phi(t,x,y,\eta)s(t,x,y,\eta)\, d\eta,$$

where ρ is a smooth cutoff.

The next Lemma will allow us to obtain a specification of the amplitude defining the operator.

Lemma 12.9. *Consider an operator of the form*

$$Sf(x) = \int a(x,y,\xi) e^{2\pi i \phi(x,y,\xi)} f(y)\, dy\, d\xi,$$

where $a \in S^r$ and vanishes for $|x-y| \gtrsim 1$, $\phi \in S^1$, and is homogeneous of degree one in ξ, $|\nabla_\xi \phi(x,y,\xi)| \gtrsim |x-y|$ on the support of a, and for all $r > 0$,

$$D^\beta_\xi \{\phi(x,y,\xi) - (x-y)\cdot\xi\} \lesssim_\beta |x-y|^2 |\xi|^{1-|\beta|}.$$

Then S is well defined, and is actually a pseudodifferential operator of order r. If T is the pseudodifferential operator with symbol $(x,\xi) \mapsto a(x,x,\xi)$, then $T - S$ is a pseudodifferential operator of order $r-1$.

Conversely, for any *ΨDO T of order r, there exists a symbol a of order r such that for the resulting operator S of order r, $S - T$ is a smoothing operator.*

Proof. Let us first define the operator S. Let $\phi_0(x,y,\xi) = (x-y)\cdot\xi$, $\phi_1(x,y,\xi) = \phi(x,y,\xi)$, $\phi_t = t\phi_1 + (1-t)\phi_0$, and define S_t with the phase function ϕ_t and symbol a. Note that $|\nabla_\xi \phi_t| \gtrsim |x-y|$, uniformly in t. This enables us to compute the kernel $K_t(x,y)$ for $0 \leqslant t \leqslant 1$. For $t = 0$ we have a pseudodifferential operator, and for $t = 1$, we get the kernel $K(x,y)$ we get to compute. It is also simple to see that, since $|\nabla_\xi \phi_t| \gtrsim |x-y|$, that for large N,

$$K(x,y)| \lesssim_N \frac{1}{|x-y|^N},$$

so we already see that S is somewhat pseudolocal.

We have

$$\frac{\partial^N K_t(x,y)}{\partial t^N} = (2\pi i)^N \int (\phi_1(x,y,\xi) - \phi_0(x,y,\xi))^N a(x,y,\xi) e^{2\pi i \phi_t(x,y,\xi)}\, d\xi.$$

Now $(\phi_1 - \phi_0)^N \cdot a$ is a symbol of order $r+N$. But on the other hand, using the fact that $(\phi_1 - \phi_0)^N \lesssim |x-y|^{2N} |\xi|^N$, and thus vanishes to order $2N$ on the diagonal, then combined with the fact that $|\nabla_\xi \phi(x,y,\xi)| \gtrsim |x-y|$, we actually see via an integration by parts $2N$ times in ξ that we can rewrite the integral in terms of a symbol of order $r-N$ and the same phase ϕ_t. Applying Taylor's theorem, we write

$$K(x,y) = K_1(x,y) = \sum_{k=0}^{N-1} \frac{1}{k!} \left.\frac{\partial^k K_t(x,y)}{\partial t^k}\right|_{t=1} + \frac{1}{N!} \int_0^1 t^{N-1} \frac{d^N K_t(x,y)}{dt^N}\, dt.$$

This integral gives an arbitrarily smooth kernel as $N \to \infty$. Thus if we let T be a pseudodifferential operator of order r such that

$$T \sim \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k K_t(x,y)}{\partial t^k} \right|_{t=1},$$

then $T - S$ is a smoothing operator. Now if $\tilde{T}$ is the pseudodifferential operator corresponding to the symbol $a(x,x,\xi)$, then $T - \tilde{T}$, and thus $S - \tilde{T}$, is a pseudodifferential operator of order $r-1$. The converse is similar, working in the opposite direction, i.e. from $t = 1$ to $t = 0$, and is left as an exercise. $\square$

Since I is a ΨDO of order zero, we can find a symbol a of order zero such that if T is the operator with kernel

$$\int a(x,y,\xi) e^{2\pi i \phi(x,y,\xi)}\, d\xi,$$

then $T - I$ is a smoothing operator. To ensure that $S(0) - I$ is a smoothing operator, it is natural to insist that $s(0,x,y,\xi) = a(x,y,\xi)$. We note in particular that since I is a ΨDO with 1 as a symbol, this implies $s(0,x,x,\xi) - 1$ is a symbol of order -1.

Next, we note that, if $b(t,x,y,\eta)$ is a symbol of order k, then modulo symbols of order $k-1$, we have

$$\begin{aligned} e^{-2\pi i \Phi(t,x,y,\eta)} & (2\pi i T)\{s(t,\cdot,y,\eta) e^{2\pi i \Phi(t,\cdot,y,\eta)}\}(x) \\ &= 2\pi i \cdot a(x, \nabla_x \phi(x,y,\eta)) \cdot b(t,x,y,\eta) \\ &\quad + \nabla_\eta a(x, \nabla_x \Phi(t,x,y,\eta)) \cdot \nabla_x b(t,x,y,\eta), \end{aligned}$$

modulo symbols of order $k-1$. Thus if $U(t)$ is the operator with kernel

$$\int b(t,x,y,\xi) e^{2\pi i \Phi(t,x,y,\xi)}\, d\xi,$$

then modulo symbols of order $k-1$ in the integrand, we have

$$\begin{aligned} &(\partial/\partial t + 2\pi i T) \circ U(t) \\ &\quad = \int e^{2\pi i \Phi(t,x,y,\eta)} \left\{ \frac{\partial b}{\partial t}(t,x,y,\eta) + 2\pi i a_0(y,\eta) b(t,x,y,\eta) \right. \\ &\qquad\qquad \left. + \nabla_\xi a(x, \nabla_x \Phi(t,x,y,\eta)) \cdot \nabla_x b(t,x,y,\eta) \right\} \end{aligned}$$

$$
\begin{aligned}
e^{-2\pi i\Phi(t,x,y,\eta)}&(2\pi i T)\{b(t,\cdot,y,\eta)e^{2\pi i\Phi(t,\cdot,y,\eta)}\}(x)\\
&= a(x,\nabla_x\Phi(t,x,y,\eta))b(t,x,y,\eta)\\
&\quad + \frac{1}{2\pi i}\sum_{j=1}^{d} D_\xi^j a(x,\nabla_x\Phi(t,x,y,\eta))D_x^j b(t,x,y,\eta)
\end{aligned}
$$

$$
\begin{aligned}
e^{-2\pi i\Phi(t,x,y,\eta)}&(2\pi i T)\{b(t,\cdot,y,\eta)e^{2\pi i\Phi(t,\cdot,y,\eta)}\}(x)\\
&= \sum_{|\beta|<2}\frac{1}{\beta!(2\pi i)^{|\beta|}}D_\xi^\beta a(x,\nabla_x\Phi(t,x,y,\eta))D_x^\beta b(t,x,y,\eta)\\
&\quad - (i/\pi)\sum_{|\beta|=2} D_\xi^\beta a(x,\nabla_x\Phi(t,x,y,\eta))D_y^\beta \rho(Z)b(t,x,y,\eta)\\
&\quad + R(t,x,y,\eta),
\end{aligned}
$$

where R is a symbol of order $k-2$.

Part III

Calderon-Zygmund Theory

Here, we try and describe the more modern approaches to real-variable harmonic analysis, as developed by the *Calderon-Zygmund school* in the 1960s and 1970s. Almost all of the problems we consider can be phrased as showing some operator is bounded as a map between functions 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 functions 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 the analysis of the operator T, 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 one of a family of *dyadic decomposition techniques*. For such decompositions, it is also possible to generalize our techniques 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 12.10. *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, $\|f_k\|_V \leqslant C \cdot 2^{-\varepsilon k}$. Then $\|f\|_V \lesssim_\varepsilon C$.*

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 $\|f\| \lesssim_\varepsilon C$. ☐

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

Lemma 12.11. *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 applying the divide and conquer approach, splitting up the indices $2^m \leqslant k \leqslant 2^{m+1} - 1$ via a binary tree with depth at most $O(m)$, 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 $s > 1 + \lg(A)$, the previous lemma implies 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 13

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.

13.1 The L^p norms

We begin by introducing the most fundamental monotone, rearrangement invariant norms. For $p \in (0,\infty)$, we define the L^p norm of a measurable function f on a measure space X by

$$\|f\|_{L^p(X)} = \left(\int_X |f(x)|^p \, dx \right)^{1/p}.$$

For $p = \infty$, we define

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

a quantity often called the *essential* supremum. If the measure space X is implicit, these quantities are also denoted by $\|f\|_p$, as we do often in this chapter. 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 in harmonic analysis for $p = 1$ and $p = \infty$ is rather trivial, in the sense that for most operators that occur in practice, it is simple to determine whether these operators are bounded on these spaces; obtaining L^p bounds of an operator for $1 < p < \infty$ reflects a deeper understanding of the quantitative properties of an operator.

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, as $p \to 0$, $L^p(X)$ has more control over the width of functions. It is therefore natural to introduce the space $L^0(X)$ as the space of measurable functions with finite measure support. But there is no natural norm on $L^0(X)$ which can classify the support of functions. After all, such a quantity couldn't be homogenous, since the width of f and αf are the same for each $\alpha \neq 0$.

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 closed set B not containing 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', which manifests in the fact 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 introduce Lorentz norms $L^{p,q}$, which one can think of as differing from the standard L^p norms 'by a logarithmic factor', and the $L^{p,q}$ norm of f will be finite if q is chosen appropriately.

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 sharp singularities, and control of an L^p 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 measure support. Nonetheless, we can conclude that* $\lim_{p\to 0} \|s\|_p^p = \sum |E_n|$.

As $p \to \infty$, the last example shows the width of a function is disregarded completely by the L^p norm, from which it follows that $\|s\|_{L^p(X)} \to \|s\|_{L^\infty(X)}$ as $p \to \infty$. The same is true for more general functions, which we can prove using a density argument.

Theorem 13.1. *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 $\|f\|_0^0 = |\text{supp} f| = |\{x : f(x) \neq 0\}|$, 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 $\|f\|_r \leqslant \|f\|_p^\theta \|f\|_q^{1-\theta}$, a special case of *Hölder's Inequality*.

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. This often introduces some extraneous (though not too inefficient) constants. But if these constants are unnecessary, one can often 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 13.2 (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 13.3. *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 13.3 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 13.4. *If $p \geqslant 1$, then for any functions $f_1, \dots, f_N \geqslant 0$,*

$$(\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \leqslant \|f_1 + \dots + f_N\|_p \leqslant \|f_1\|_p + \dots + \|f_N\|_p. \tag{13.1}$$

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

$$\|f_1\|_p + \dots + \|f_N\|_p \leqslant \|f_1 + \dots + f_N\|_p \leqslant (\|f_1\|_p^p + \dots + \|f_N\|_p^p)^{1/p} \tag{13.2}$$

Proof. The upper bound in (13.1) 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 13.2. 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 $p < 1$. However, for $p < 1$, we do have a *quasi* triangle inequality.

Theorem 13.5. *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,

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

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*

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

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.

One can also often obtain analogous results when dealing with functions which are roughly constant at a scale ε, rather than literally constant at this scale. Basic examples of this include Bernstein's inequality; the Sobolev embeddings are also of this flavor. But this is a topic for another section.

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 right 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 13.6. *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 13.7. *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

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

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 $g(x) = \mathrm{sgn}(f(x))\mathbf{I}_E/|E|$, 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. □

13.2 Decreasing Rearrangements

The measure-theoretic properties of a function's 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 13.8. *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 13.9. *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. The 'mass' of these singularities, of course, is important.

13.3 Weak Norms

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

Theorem 13.10. *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 13.11 (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 $\|f_1 + \cdots + f_N\|_{p,\infty} \lesssim_p 1$. □

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 13.5.

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 $k_1, \dots, k_N > 0$ 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

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

If $f = H_1 \mathbf{I}_{E_1} + H_2 \mathbf{I}_{E_2}$, with $|E_1| = W_1$ and $|E_2| = W_2$, with $H_1 \leqslant H_2$, 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 L^p norm, at least when $p > 1$.

Theorem 13.12. *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 13.13. *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

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

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

13.4 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 13.14. *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 13.15. *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 13.16. *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 $g = f_1 + f_2$,

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

□

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 13.17. *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 decreasing in k, 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. □

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 $f = f_1 + \cdots + f_N$, 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

$$\begin{aligned}\left(\sum_{k\in\mathbf{Z}}\left[H_k W_k^{1/p_2}\right]^{q}\right)^{1/q} &\leqslant \left(\sum_{k\in\mathbf{Z}}\left[H_k W_k^{1/p_2}\right]^{qp_2/p_1}\right)^{(p_1/p_2)/q}\\ &\leqslant \max(H_k)^{1-p_1/p_2}\left(\sum_{k\in\mathbf{Z}}\left[H_k W_k^{1/p_1}\right]^{q}\right)^{(p_1/p_2)/q}\end{aligned}$$

$$\left(\sum_{k\in\mathbf{Z}}\left[H_k W_k^{1/p_1}\right]^q\right)^{1/q} \lessapprox \left(\sum_{k\in\mathbf{Z}}\left[H_k W_k^{1/p_1}\right]^{qp_1/p_2}\right)^{(p_2/p_1)/q}$$
$$\leqslant \left(\frac{1}{\min(H_k)}\right)^{p_2/p_1-1}\left(\sum_{k\in\mathbf{Z}}\left[H_k W_k^{1/p_2}\right]^q\right)^{(p_2/p_1)/q}$$

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 13.18. *If $0 < p_1, p_2, p < \infty$ and $0 < q_1, q_2, q < \infty$ with*

$$1/p = 1/p_1 + 1/p_2 \quad \textit{and} \quad 1/q \geqslant 1/q_1 + 1/q_2,$$

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 $\|f\|_{p_1,q_1} = \|g\|_{p_2,q_2} = 1$ and that $1/q = 1/q_1 + 1/q_2$. 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 13.19. *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

$$\|f\|_{L^1(E)} = \|\mathbf{I}_E f\|_{L^1(X)} \lesssim_{p,q} |E|^{1-1/p} \|f\|_{p,q} < \infty. \qquad \square$$

A consequence of Hölder's inequality is a 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 13.20. *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 $g_k(x) = a_k^{q-p} \text{sgn}(f_k(x)) |f_k(x)|^{p-1}$. 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

$$\sup\left\{\int fg : \|g\|_{p',q'} \leqslant 1\right\}$$

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 13.21. *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

$$E' = \left\{ x : 2^{k-2} < \left| \sum_{j=k-\lg(N)}^{k} f_{1j}(x) + \cdots + f_{Nj}(x) \right| \leqslant 2^{k+1} \right\}$$

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

$$\|f_1 + \cdots + f_N\|_{1,q} \lessapprox_q \left(\sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q}.$$

Applying Minkowski's inequality, we conclude

$$\begin{aligned}
\left(\sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q} &\lesssim \sum_{i=1}^{N} \left(\sum_{k=-\infty}^{\infty} \left(\sum_{j=k-\lg N}^{k} W_{ij2^j} \right)^q \right)^{1/q} \\
&\lessapprox \sum_{i=1}^{N} \left(\sum_{k=-\infty}^{\infty} \sum_{j=k-\lg N}^{k} W_{ij}^q 2^{qj} \right)^{1/q} \\
&\lessapprox \sum_{i=1}^{N} \left(\sum_{j=-\infty}^{\infty} W_{ij}^q 2^{qj} \right)^{1/q} \lesssim 1. \qquad \square
\end{aligned}$$

Here is an interesting weighted inequality whose proof utilizes the layer cake decomposition. I first encountered this inequality in Heo, Nazarov, and Seeger's paper *Radial Fourier Multipliers in High Dimensions*.

Theorem 13.22. *Suppose that $\{f_n\}$ is a family of functions, fix $p_0 < p_\theta < p_1$. Then*

$$\|\sum_n f_n\|_{L^p(X)} \lesssim_{p_0,p,p_1} \left(\sum_n \max_i \left\{2^{n(p-p_i)}\|f_n\|_{p_i,\infty}^{p_i}\right\}\right)^{1/p}$$

In particular, if $\|f_n\|_{p_i,\infty} \leqslant C2^n W_n^{1/p_i}$ for each n, then

$$\|\sum_n f_n\|_{L^p(X)} \lesssim_{p_0,p,p_1} C\left(\sum_n 2^{pn}W_n\right)^{1/p}.$$

This might hold, for instance, if f_n is a sub step function with height 2^n and width W_n for each n.

Proof. Define $f_{nm} = f_n \cdot \mathbf{I}(2^{n+m} \leqslant |f_n| < 2^{n+m+1})$. Then $f_n = \sum_m f_{nm}$. For each fixed m, define $\tilde{f}_m = \sum_n f_{nm}$. Since $f_{nm} \approx 2^{n+m}$, f_m is defined by a sum over different dyadic scales, and so we have a pointwise bound

$$|\sum_n f_{nm}| \sim \max\{2^m : f_{nm} \neq 0\} \sim_p \left(\sum_n |f_{nm}|^p\right)^{1/p}.$$

Thus we find

$$\|\tilde{f}_m\|_{L^p(X)} \lesssim_p \left\|\left(\sum_n |f_{nm}|^p\right)^{1/p}\right\|_{L^p(X)} = \left(\sum_n \|f_{nm}\|_{L^p(X)}^p\right)^{1/p}.$$

Chebyshev's inequality implies that

$$\begin{aligned}\|f_{nm}\|_{L^p(X)}^p &\leqslant 2^{(n+m)p} \cdot \min_i\{2^{-(n+m)p_i}\|f_n\|_{L^{p_i,\infty}(X)}^{p_i}\} \\ &\lesssim \min_i\{2^{(n+m)(p-p_i)}\|f_n\|_{L^{p_i,\infty}(X)}^{p_i}\}.\end{aligned}$$

But this means that if $m \geqslant 0$,

$$\|\tilde{f}_m\|_{L^p(X)} \lesssim_p 2^{-|m|(p_1/p-1)}\left(\sum_n 2^{-n(p_1-p)}\|f_n\|_{L^{p_1,\infty}(X)}^{p_1}\right)^{1/p}$$

and for $m \leqslant 0$,

$$\|\tilde{f}_m\|_{L^p(X)} \lesssim_p 2^{-|m|(1-p_0/p)}\left(\sum_n 2^{n(p-p_0)}\|f_n\|_{L^{p_0,\infty}(X)}^{p_0}\right)^{1/p}$$

Applying the triangle inequality to $\|\sum f_n\|_{L^p(X)} = \|\sum \tilde{f}_m\|_{L^p(X)}$ and summing over m completes the proof. $\square$

13.5 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 $\|f\|_{L^q(Y)L^p(X)}$. We have a duality theory here; for each $1 \leqslant p,q < \infty$ and any f with $\|f\|_{L^p(X)L^q(Y)} < \infty$, the standard L^p and L^q duality gives

$$\|f\|_{L^p(X)L^q(Y)} = \sup \left\{ \int_{X \times Y} f(x,y)h(x,y) \, dx \, dy : \|h\|_{L^{p^*}(X)L^{q^*}(Y)} \leqslant 1 \right\}.$$

It is often important to interchange norms, and we find the biggest quantity obtained by interchanging norms is always obtained with the largest exponents on the inside.

Theorem 13.23. *If $q \geqslant p \geqslant 1$, $\|f\|_{L^p(X)L^q(Y)} \leqslant \|f\|_{L^q(Y)L^p(X)}$.*

Proof. If $p = q$, then the Fubini-Tonelli theorem implies that

$$\|f\|_{L^p(X)L^q(Y)} = \|f\|_{L^q(Y)L^p(X)}.$$

If $p = 1$, then this result is precisely the Minkowski inequality. We now apply complex interpolation to obtain the result in general. In fact, a simple variation of the proof of Riesz-Thorin using the duality established above gives the result. $\square$

Let us consider two special case. Firstly, bounds for the pointwise maxima of functions dominate the maximum L^p norms

$$\sup_n \|f_n\|_{L^p(X)} \leqslant \left\| \sup_n f_n \right\|_{L^p(X)}.$$

Secondly, a special case is the triangle inequality

$$\left\| \sum_n f_n \right\|_{L^p(X)} \leqslant \sum_n \|f_n\|_{L^p(X)}$$

for $p \geqslant 1$.

It turns out that if $q > p$ and $\|f\|_{L^p(X)L^q(Y)} = \|f\|_{L^q(Y)L^p(X)}$, then $|f|$ is a tensor product. Thus switching mixed norms is likely only efficient if the functions we are working with are close to tensor products.

Theorem 13.24. *Suppose $q > p$, f is a function on $X \times Y$, and*

$$\|f\|_{L^p(X)L^q(Y)} = \|f\|_{L^q(Y)L^p(X)} < \infty.$$

Then there exists $f_1(x)$ and $f_2(y)$ such that for any $x \in X$ and $y \in Y$, $|f(x,y)| = |f_1(x)||f_2(y)|$.

Proof. Expanding this equation out, we conclude

$$\left(\int_Y \left(\int_X |f(x,y)|^p \, dx \right)^{q/p} dy \right)^{1/q} = \left(\int_X \left(\int_Y |f(x,y)|^q \, dy \right)^{p/q} dx \right)^{1/p}.$$

Setting $g(x,y) = |f(x,y)|^p$, we see that Minkowski's integral inequality is tight for g, i.e.

$$\left(\int_Y \left(\int_X |g(x,y)| \, dx \right)^{q/p} dy \right)^{p/q} = \left(\int_X \left(\int_Y |g(x,y)|^{q/p} \, dy \right)^{p/q} dx \right).$$

Thus it suffices to show that to show the theorem for $p = 1$ and $q > 1$. Recall the standard proof of Minkowski's inequality, i.e. that by Hölder's inequality

$$\begin{aligned}
\int_Y \left(\int_X |f(x,y)| \, dx \right)^p dy &= \int_X \left[\int_Y |f(x_1,y)| \left(\int_X |f(x_2,y)| \, dx_2 \right)^{p-1} dy \right] dx_1 \\
&\leqslant \int_X \left[\left(\int_Y |f(x_1,y)|^p \, dy \right)^{1/p} \left(\int_Y \left(\int_X |f(x_2,y)| \, dx_2 \right)^{(p-1)p^*} dy \right)^{1/p^*} \right] d \\
&= \left[\int_X \left(\int_Y |f(x_1,y)|^p \, dy \right)^{1/p} dx_1 \right] \left[\int_Y \left(\int_X |f(x_2,y)| \, dx_2 \right)^p \right]^{1/p^*}.
\end{aligned}$$

and rearranging gives Minkowski's inequality. If this inequality is tight, then our application of Hölder's inequality is tight for almost every $x_1 \in X$. Since $\int |f(x_2, y)|\, dx_2 \neq 0$ for all y unless $f = 0$, it follows that there exists $\lambda(x_1)$ for almost every $x_1 \in X$ such that for almost every $y \in Y$,

$$|f(x_1, y)|^p = |\lambda(x_1)| \left(\int_X |f(x_2, y)|\, dx_2 \right)^{p^*(p-1)} = |\lambda(x_1)| \left(\int_X |f(x_2, y)|\, dx_2 \right)^p.$$

Setting $f_1(x) = |\lambda(x)|^{1/p}$ and $f_2(y) = \int_X |f(x,y)|\, dx$ thus completes the proof. □

13.6 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

$$\int \Phi\left(|f(x)|\right)\, dx = 1.$$

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

$$\int \Phi\left(|f(x)|\right)\, dx = 1.$$

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 (often the case, since these norms often occur in probability theory), they are irrelevant.

Lemma 13.25. *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 $\|f\|_{\Phi(L)} \lesssim 1$. □

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 13.26. *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 $\Phi(x) = yx - \Psi(y) \leqslant \Pi(x)$. □

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 13.27. *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. □

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 14

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*.

14.1 Interpolation of Functions

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 $L^{p_0,\infty}(X) \cap L^{p_1,\infty}(X) \subset L^{p_\theta}(X)$.

14.2 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 14.1 (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. $\square$

Remark. The function $e^{-ie^{\pi i s}}$ 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
$$|f(z)| \lesssim |1+z|^\alpha \lesssim (1+|z|)^\alpha. \qquad \square$$

14.3 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.

14.4 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 14.2 (Riesz-Thorin)**.** *Let $p_0, p_1 \in (0,\infty]$ and $q_0, q_1 \in [1,\infty]$. Suppose that*

$$\|Tf\|_{L^{q_0}(Y)} \leqslant A_0 \|f\|_{L^{p_0}(X)} \quad \textit{and} \|Tf\|_{L^{q_1}(Y)} \leqslant A_1 \|f\|_{L^{p_1}(X)}.$$

Then for any $\theta \in (0,1)$, if

$$1/p_\theta = (1-\theta)/p_0 + \theta/p_1 \quad \textit{and} \quad 1/q_\theta = (1-\theta)/q_0 + \theta/q_1,$$

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 14.3 (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. □

14.5 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,

  $$\|T(\mathbf{I}_E)\|_{L^q(Y)} \lesssim |E|^{1/p}.$$

  The equivalence is proven by breaking any sub-step function f with height H and width W into a dyadic sum $\sum_{k=1}^{\infty} H\mathbf{I}_{E_k} 2^{-k}$, where $|E_k| \leqslant W$.
- 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 13.13.

Lemma 14.4. *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 14.5. *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). $\square$

Theorem 14.6 (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 15

Basics of Kernel Operators

We now consider a 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, heuristically at least, by the integral formula

$$(T_K f)(y) = \int_X K(x,y) f(x)\, dx.$$

Our goal is to relate properties of the kernel K to the regularity of the operator T_K with respect to various norms.

Example. *Let $X = Y = \mathbf{R}^d$, equipped with the Lebesgue measure. If we set $K(x,\xi) = e^{2\pi i \xi \cdot x}$, then using this function as a kernel we can obtain an integral operator*

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

In the standard theory of Fourier analysis, we find that if $f \in L^1(\mathbf{R})$, then for any ξ the integral

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

converges absolutely, and is thus well-defined in the sense of a Lebesgue integral. Moreover, for any $f \in L^1(\mathbf{R})$,

$$\|T_K f\|_{L^\infty(\mathbf{R})} \leqslant \|f\|_{L^1(\mathbf{R})}.$$

We also know from the classical Hausdorff-Young inequality that if $1 \leqslant p \leqslant 2$, *then for any* $f \in L^1(\mathbf{R}) \cap L^p(\mathbf{R})$,

$$\|T_K f\|_{L^{p^*}(\mathbf{R})} \leqslant \|f\|_{L^p(\mathbf{R})}.$$

In particular, this means that there exists a unique extension of T_K *to a bounded operator from* $L^p(\mathbf{R})$ *to* $L^{p^*}(\mathbf{R})$*; note, however, that for a general element* $f \in L^p(\mathbf{R})$*, the integral formula*

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

is not well-defined *in the Lebesgue sense. Thus we can only heuristically view the integral formula as defining the integral 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* $A_{ij} = K(j,i)$*. For any* $f : X \to Y$ *we can define a vector* $v \in \mathbf{R}^N$ *by setting* $v_i = f(i)$*, and then*

$$(T_K f)(m) = \sum_{n=1}^N f(n) K(n,m) = \sum_{n=1}^N A_{mn} v_n = (Av)_m.$$

Thus with respect to the standard basis, T_K *is just given by matrix multiplication by* A.

It turns out that if we map *from* $L^1(X)$, or *into* $L^\infty(Y)$, then the conditions on K determining boundedness are trivial to determine for *most* norms. This is one motivation for introduction the intermediate L^p norms, since these norms enable us to extract more features out of the kernel operator K. Before we discuss this, we must first reflect on the fact that without even qualitative knowledge of the kernel K besides it's measurability, it is difficult to know how one might interpret the integral formula defining the operator. A natural trick to begin with is to introduce the sublinear analogue of the kernel operator, i.e. the operator S_K defined by setting

$$(S_K f)(y) = \int_X |K(x,y)||f(x)|; dx$$

The flexibility of the theory of non-negative Lebesgue integrals means this operator is well defined for *any* measurable f (though $S_k f(y)$ may be infinite for particular values of y). Moreover, if we are to interpret the integral formula for $(T_K f)(y)$ in the Lebesgue sense, it is necessary and sufficient that $(S_K f)(y) < \infty$.

Theorem 15.1. *Fix $q \geqslant 1$. Then*

$$\|S_k f\|_{L^q(Y)} \leqslant \|K\|_{L^q(Y)L^\infty(X)} \|f\|_{L^1(X)}.$$

Thus $T_k f(y)$ are well defined by a Lebesgue integral for almost every $y \in Y$, and

$$\|T_k f\|_{L^q(Y)} \leqslant \|K\|_{L^q(Y)L^\infty(X)} \|f\|_{L^1(X)}.$$

Proof. The proof is just a simple consequence of Minkowski's inequality, i.e.

$$\begin{aligned}
\|S_K f\|_{L^q(Y)} &= \|Kf\|_{L^1(X)L^q(Y)} \\
&\leqslant \|Kf\|_{L^q(Y)L^1(X)} \\
&= \int \left(\int |K(x,y)|^q \, dy \right)^{1/q} |f(x)| \, dx \\
&\leqslant \|K\|_{L^q(Y)L^\infty(X)} \|f\|_{L^1(X)}.
\end{aligned}$$

□

Remark. In a great many situations, this constant is tight. For isntance, suppose

$$K = \sum_{i=1}^{N} \sum_{j=1}^{M} a_{ij} \mathbf{I}_{E_i \times F_j}$$

where $E_1, \ldots, E_N$ and $F_1, \ldots, F_N$ are disjoint finite measure sets. Then there exists $i \in \{1, \ldots, N\}$ 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 $T_K f = \sum_{j=1}^{M} a_{ij} \mathbf{I}_{F_j}$, so

$$\|T_K f\|_{L^q(Y)} = \left(\sum_{j=1}^{M} |a_{ij}|^q |F_j| \right)^{1/q} = \|K\|_{L^q(Y)L^\infty(X)} \|f\|_{L^1(X)}.$$

Thus we conclude that for a certain 'dense' family of K, the inequality above is tight, which gives a strong heuristic that the inequality above is tight for a great many operators K, which trivializes the analysis of $L^1(X) \to L^q(Y)$ estimates.

A dual statement trivializes the analysis of bounds from $L^p(X)$ to $L^\infty(Y)$.

Theorem 15.2. *Suppose* $1 \leqslant p \leqslant \infty$. *Then*

$$\|S_K f\|_{L^\infty(Y)} \leqslant \|K\|_{L^{p^*}(X)L^\infty(Y)} \|f\|_{L^p(X)}.$$

Thus if $\|K\|_{L^{p^*}(X)L^\infty(Y)} < \infty$, *then* $T_K f(y)$ *is well defined for almost every* $y \in Y$, *and*

$$\|T_K f\|_{L^\infty(Y)} \leqslant \|K\|_{L^{p^*}(X)L^\infty(Y)} \|f\|_{L^p(X)}.$$

Proof. One option to proving this bound is to take the adjoint of the kernel operator T_K and rely on previous estimates, but we can work more directly. Applying Hölder's inequality, we conclude that

$$\|S_K f\|_{L^\infty(Y)} = \|Kf\|_{L^1(X)L^\infty(Y)} \leqslant \|K\|_{L^{p^*}(X)L^\infty(Y)} \|f\|_{L^p(X)}. \qquad \square$$

Though trivial, the two kernel bounds can often be applied together with an interpolation argument to give more complicated bounds.

Theorem 15.3 (Schur's Test). *Suppose that* $\|K\|_{L^1(X)L^\infty(Y)} \leqslant A$ *and* $\|K\|_{L^1(Y)L^\infty(X)} \leqslant B$. *Then for every* $1 \leqslant p \leqslant \infty$ *and* $f \in L^p(X)$, $(T_K f)(y)$ *is well defined by an absolutely convergent integral, and*

$$\|T_K f\|_{L^p(Y)} \leqslant A^{1-1/p} B^{1/p} \|f\|_{L^p(X)}.$$

Proof. The previous two results imply that S_K is bounded from $L^1(X)$ to $L^1(Y)$ and from $L^\infty(X)$ to $L^\infty(Y)$. Real interpolation (since S_K is sublinear) shows that S_K is bounded from $L^p(X)$ to $L^p(Y)$ for all $1 \leqslant p \leqslant \infty$. Thus the operator T_K is well defined by Lebesgue integrals for $f \in L^p(X)$. Applying the Riesz-Thorin interpolation theorem to T_K, which satisfies the bounds $\|T_K f\|_{L^1(Y)} \leqslant A\|f\|_{L^1(X)}$ and $\|T_K f\|_{L^\infty(Y)} \leqslant B\|f\|_{L^\infty(X)}$, we obtain the required result. $\square$

For $1 < p < \infty$, we do not expect Schur's test to be sharp in general. But a good heuristic is that it is sharp provided that

- For all $y \in Y$,
$$\int |K(x,y)|\, dx \approx A$$
and for all $x \in X$,
$$\int |K(x,y)|\, dy \approx B.$$
- There is little oscillation in the kernel K.

Assuming X and Y have finite measure, from the first property we conclude that
$$A|Y| \approx B|X|.$$
Thus if we set $f = \mathbf{I}_X$, then $Tf(y) \approx A$ for all $y \in Y$, hence
$$\|Tf\|_{L^p(Y)} \approx A|Y|^{1/p} \approx A^{1-1/p}B^{1/p}|X|^{1/p}.$$
Thus we have tightness. If the second property remains true, but the first property fails, Schur's lemma still may remain sharp if we consider a weighted inequality, or alternatively, if we decompose the operator into components on which the marginal is approximately constant.

In some senses, if we are allowed to work with arbitrary weights, and if $K \geqslant 0$, Schur's test is always sharp. Suppose that
$$\|T_K f\|_{L^p(Y)} \leqslant A\|f\|_{L^p(X)}$$
for all $f \in L^p(X)$, and this inequality is sharp for some particular function f_0. We may assume without loss of generality that $\|f_0\|_{L^p(X)} = 1$ and, since K is non-negative, that $f \geqslant 0$. Thus
$$\int_Y (T_K f_0(y))^p\, dy = A^p.$$
An application of Lagrangian multipliers and basic calculus of variations then shows that there exists a scalar λ such that
$$T_K^*((T_K f)^{p-1})(x) = \lambda f(x)^{p-1}.$$
But this means that
$$\begin{aligned}\lambda &= \int \lambda f(x)^p\, dx \\ &= \int f(x) T_K^*((T_K f)^{p-1})(x)\, dx \\ &= \int (T_K f)^p(x) = A^p.\end{aligned}$$

Thus if we set $w(x) = f(x)$ and $v(y) = (T_K f(y))^{p-1}$, then

$$\int_X K(x,y)w(x)\, dx = v(y)^{1/(p-1)}$$

and

$$\int_Y K(x,y)v(y)\, dy = A^p w(x)^{p-1}.$$

Using these estimates, a weighted variant of Schur's lemma gives the bound $\|T_K f\|_{L^p(X)} \leqslant A\|f\|_{L^p(X)}$, which shows that the two weighted identities above contain as much information as the original bound.

TODO: Fill in rest of Tao's notes.

15.1 Localization In Space

Localization is a fundamental technique in anlaysis, since it enables us to isolate certain parts of a function or operator. If we understand these localized parts, one can then often recover results about the original result using a partition of unity.

Chapter 16

Maximal Averages

This chapter is about exploring the behaviour of basic averaging operators. A classical example, given a function $f \in L^1_{\text{loc}}(\mathbf{R})$, are the averaging operators

$$A_\delta f(x) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} f(y)\, dy.$$

If $f \in C(\mathbf{R})$, then for each $x \in \mathbf{R}$, $\lim_{\delta \to 0} A_\delta f(x) = f(x)$. This fact is fundamentally connected to differentiation under the integral sign; if we define the function

$$F(x) = \int_0^x f(y)\, dy$$

then for each $x \in \mathbf{R}$,

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{1}{h} \int_x^{x+h} f(y)\, dy = f(x).$$

Our main goal will be study whether pointwise convergence of the averages $A_\delta f$ hold for a more general family of functions or equivalently, studying whether a kind of fundamental theorem of calculus holds for a more general family of measurable functions, which are not necessarily continuous.

The classical family of averaging operators are defined for $\delta > 0$, $f \in L^1_{\text{loc}}(\mathbf{R}^d)$, and $x \in \mathbf{R}^d$ by setting

$$A_\delta f(x) = \frac{1}{|B(x,\delta)|} \int_{B(x,\delta)} f(y)\, dy,$$

where $B(x,\delta)$ is the ball of radius δ centred at x. A simple application of Schur's lemma shows that $\|A_\delta f\|_{L^p(\mathbf{R}^d)} \leqslant \|f\|_{L^p(\mathbf{R}^d)}$ for all $1 \leqslant p \leqslant \infty$, uniformly in p. This uniform bound in δ is strong enough, together with the density of compactly supported continuous functions is enough to conclude that for any $f \in L^p(\mathbf{R}^d)$, for $1 \leqslant p < \infty$, $A_\delta f$ converges to f in L^p norm. This implies that for any $f \in L^p(\mathbf{R}^d)$, there exists a sequence δ_i converging to zero such that $A_{\delta_i} f$ converges to f pointwise almost everywhere. In this chapter, we would like to show $A_\delta f$ converges to f pointwise almost everywhere *without taking a subsequence of values* δ_i.

Hardy and Littlewood introduced a powerful technique to study such pointwise convergence problems, known as the *method of maximal functions*. For each $f \in L^1_{\text{loc}}(X)$, we define

$$Mf(x) = \sup_{\delta>0} A_\delta |f|(x) = \sup_{\delta>0} \frac{1}{|B(x,\delta)|} \int_{B(x,\delta)} |f(y)|\, dy.$$

The next theorem indicates why obtaining bounds on a maximal operator gives pointwise convergence results.

Theorem 16.1. *Let V be a quasinorm space, let $0 < q < \infty$, and consider a family of bounded operators $T_t : V \to L^{q,\infty}(X)$. Then we can define the pointwise maximal operator*

$$T_* f(x) = \sup_t |T_t f(x)|.$$

Suppose that for every $f \in L^p(X)$,

$$\|T_* f\|_{L^{q,\infty}(X)} \lesssim \|f\|_V.$$

Then for any bounded operator $S : V \to L^{q,\infty}(X)$, the set

$$\{f \in V : \lim_{t\to\infty} T_t f(y) = Sf(y) \text{ for a.e } y\}$$

is closed in V.

Proof. Fix a sequence $\{u_n\}$ in V converging to $u \in V$, and suppose for each n,

$$\lim_{t\to\infty} (T_t u_n)(x) = Su_n(x)$$

holds for almost every $x \in X$. For each $\lambda > 0$, we find

$$\begin{aligned}
&|\{x \in X : \limsup_{t \to \infty} |T_t u(x) - S u(x)| > \lambda\}| \\
&\qquad \leqslant |\{x \in X : \limsup_{t} |T_t(u - u_n)(x) - S(u - u_n)(x)| > \lambda\}| \\
&\qquad \leqslant |\{x \in X : |T_*(u - u_n)(x)| > \lambda/2\}| + |\{x : |S(u - u_n)(x)| > \lambda/2\}| \\
&\qquad \lesssim_{p,q} \frac{\|u - u_n\|_V^q}{\lambda^q} + \frac{\|u - u_n\|_V^p}{\lambda^p}.
\end{aligned}$$

as $n \to \infty$, this quantity tends to zero. Thus for all $\lambda > 0$,

$$|\{x : \limsup_{t \to \infty} |T_t u(x) - S u(x)| > \lambda\}| = 0$$

Taking $\lambda \to 0$ gives that $\limsup_t |T_t u(x) - S u(x)| = 0$ for almost every $x \in X$. But this means precisely that $T_t u(x) \to S u(x)$ for almost every $x \in X$. □

Taking $t = \delta$, $T_t = A_\delta$, and S the identity map, the theorem above implies that one way to obtain almost everywhere convergence for the averages we consider is via bounding the maximal operator M. Thus we consider a bound of the form

$$\left\| \sup_{\delta > 0} A_\delta f \right\|_{L^{q,\infty}(\mathbf{R}^d)} \lesssim \|f\|_V$$

for an appropriate norm $\|\cdot\|_V$ and $0 < q < \infty$. We have already obtained a bound

$$\sup_{\delta > 0} \|A_\delta f\|_{L^{q,\infty}(\mathbf{R}^d)} \leqslant \sup_{\delta > 0} \|A_\delta f\|_{L^q(\mathbf{R}^d)} \leqslant \|f\|_{L^q(\mathbf{R}^d)}$$

but moving the supremum inside the L^q norm is nontrivial. One way to think about the difference between the two bounds is that the latter uniformly controls the height and width of the functions $A_\delta f$, whereas the former inequality shows that the main contribution to the height and widths of the functions $A_\delta f$ are uniformly supported in similar regions of space.

16.1 Covering Methods

The bound $\|Mf\|_{L^\infty(\mathbf{R}^d)} \leqslant \|f\|_{L^\infty(\mathbf{R}^d)}$ from a direct calculation. Thus there are trivial techniques of bounding the height of the function Mf in terms of the height of the function f. The difficult part is obtaining control of the width of Mf in terms of the width of f. This can only be obtained up to a certain degree, because unless $f = 0$, Mf is non-vanishing on the entirety of $\mathbf{R}^d$ so the width of f 'explodes'. A slightly more technical calculation shows that we cannot even have a bound of the form $\|Mf\|_{L^1(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}$. In fact, $\|Mf\|_{L^1(\mathbf{R}^d)} = \infty$ for any nonzero $f \in L^1(\mathbf{R}^d)$.

Example. *Fix $f \in L^1(\mathbf{R}^d)$. By rescaling, we may assume without loss of generality that $\|f\|_{L^1(\mathbf{R}^d)} = 2$. Then, for suitably large $R \geqslant 1$,*

$$\int_{B_R(0)} |f(x)|\, dx \geqslant 1.$$

For each $x \in \mathbf{R}^d$, $B_R(0) \subset B_{|x|+R}(x)$ and so

$$Mf(x) \geqslant \fint_{B_{|x|+R}(x)} |f(y)|\, dy \gtrsim \frac{1}{(|x|+R)^d} \gtrsim \frac{1}{|x|^d}$$

But this means that

$$\int_{\mathbf{R}^d} |Mf(x)| \gtrsim \int_{\mathbf{R}^d} \frac{1}{|x|^d} = \infty.$$

If we are more careful, we can even find examples of $f \in L^1(\mathbf{R}^d)$ such that Mf is not even locally integrable. If $f(x) = 1/|x|\log|x|^2$, then the fact that for $x \geqslant 0$

$$\begin{aligned} \frac{1}{2h}\int_{x-h}^{x+h} \frac{dy}{|y|\log|y|^2} &= \frac{1}{2h}\left(\frac{1}{\log(x-h)} - \frac{1}{\log(x+h)}\right) \\ &= \frac{1}{2x\log x} + O\left(\frac{h}{\log x}\right) \end{aligned}$$

implies that

$$Mf(x) \geqslant \frac{1}{2x\log x}.$$

Thus Mf isn't integrable about the origin. Note however, that Mf is on the border *of integrability, which hints at the fact that we have a weak type $(1,1)$ bound.*

The last example shows that $|Mf(x)| \gtrsim |x|^{-d}$. Note, however, that $|x|^{-d}$ is only *barely* nonintegrable. We will also show that Mf is barely nonintegrable by obtaining a bound

$$\|Mf\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}.$$

Interpolation thus shows that $\|Mf\|_{L^p(\mathbf{R}^d)} \lesssim_{d,p} \|f\|_{L^p(\mathbf{R}^d)}$ for all $1 < p \leqslant \infty$. The standard real-variable technique of obtaining this bound is geometric, applying a covering argument. To obtain the weak-type bound, we must show that the set

$$E_\lambda = \{x \in \mathbf{R}^d : |Mf(x)| > \lambda\}$$

is small. If $|Mf(x)| > \lambda$, there is a ball B around x such that

$$\int_B |f(y)|\, dy > \lambda |B|.$$

Clearly $B \subset E_\lambda$. If we could find a large family of *disjoint balls* $B_1, \dots, B_N$ such that this inequality held, such that $\sum |B_i| \gtrsim_d |E_\lambda|$, then we would conclude that

$$\|f\|_{L^1(\mathbf{R}^d)} \geqslant \sum_{i=1}^N \int_{B_i} |f(y)|\, dy > \lambda \sum_{i=1}^N |B_i| \gtrsim_d \lambda |E_\lambda|$$

which would show $|E_\lambda| \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}/\lambda$, which would show $\|Mf\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}$. This intuition is true, and the process through which we obtain the family of disjoint balls $B_1, \dots, B_N$ is through the *Vitali covering lemma*.

This particular technique has been shown to generalize to a wide variety of situations including the maximal ball average. All that is really required for the basic theory is a basic 'covering type argument' that holds in a great many situations. In particular, we can generalize this argument to a *space of homogenous type*. We consider a locally compact topological space X together with a nonzero Radon measure. For each $x \in X$ and $\delta > 0$, we fix an open, precompact set $B(x,\delta)$, which we assume to be monotonically increasing in δ. The fundamental property we require of these sets is that there is $c > 0$ such that for any $x \in X$ and $\delta > 0$, if we set

$$B^*(x,\delta) = \bigcup \{B(x',\delta) : B(x,\delta) \cap B(x',\delta) \neq \varnothing\},$$

then $|B^*(x,\delta)| \leqslant c|B(x,\delta)|$. In the case of balls in $\mathbf{R}^d$, $B^*(x,\delta) \subset B^*(x,3\delta)$, and so $|B^*(x,\delta)| \leqslant 3^d |B(x,\delta)|$, so $c = 3^d$. More generally, if we are working in any metric space X, where $B(x,\delta)$ are the balls of radius δ in this metric space, and our measure satisfies a *doubling condition*

$$|B(x,3\delta)| \lesssim |B(x,\delta)|$$

for all $x \in X$ and $\delta > 0$, then our assumption holds. We also assume the following two technical assumptions

- For any $x \in X$,

$$\bigcap_{\delta > 0} \overline{B}(x, \delta) = \{x\} \quad \text{and} \quad \bigcup_{\delta > 0} B(x, \delta) = X$$

- For any open set $U \subset X$ and $\delta > 0$, the function

$$x \mapsto |B(x, \delta) \cap U|$$

  is a continuous function of x.

These are fairly easily verifiable in any particular instance. It follows from these technical assumptions that $|B(x, \delta)| > 0$ for each $x \in X$ and $\delta > 0$, and moreover, for each $\delta > 0$, and $f \in L^1_{\text{loc}}(X)$, the averaged function $A_\delta f$ given by setting

$$A_\delta f(x) = \frac{1}{|B(x, \delta)|} \int_{B(x,\delta)} f(y)\, dy,$$

is measurable.

Lemma 16.2. *If $f \in L_1^{loc}(X)$, then $A_\delta f$ is a measurable function.*

Proof. If $f = a_1 \mathbf{I}_{U_1} + \dots + a_N \mathbf{I}_{U_N}$ is a simple function, where $U_1, \dots, U_N$ are open sets, then

$$A_\delta f(x) = a_1 \frac{|B(x, \delta) \cap U_1|}{|B(x, \delta)|} + \dots + a_N \frac{|B(x, \delta) \cap U_N|}{|B(x, \delta)|}$$

is a continuous function by our technical assumptions. Next, if $f \geqslant 0$ is a step function, then there exists a monotonically decreasing family of simple functions $\{f_n\}$ such that $f_n \to f$ pointwise, then the monotone convergence theorem implies that $A_\delta f_n \to A_\delta f$ pointwise, so $A_\delta f$ is measurable. Finally, decomposing any measurable function into the difference of non-negative measurable functions and then considering pointwise limits of step functions completes the proof. $\square$

It also follows from our technical assumptions that for any open set U containing x, there exists δ_0 such that for $\delta \leqslant \delta_0$, $\overline{B(x, \delta)} \subset U$. It follows that for any $f \in C(X)$ and $x \in X$,

$$\lim_{\delta \to 0} A_\delta f(x) = f(x). \tag{16.1}$$

If $Mf = \sup_{\delta>0} A_\delta f$, then we will show

$$\|Mf\|_{L^{1,\infty}(X)} \lesssim_c \|f\|_{L^1(X)}.$$

In particular, this shows that for any $f \in L^1(X)$,

$$\lim_{\delta\to 0} A_\delta f(x) = f(x)$$

for almost every $x \in X$. Since this result is a *local result*, it is easy to verify that the result also holds for any $f \in L^1_{\text{loc}}(X)$, i.e. it also holds for any $f \in L^p(X)$ for $1 \leqslant p \leqslant \infty$.

Lemma 16.3 (Vitali Covering Lemma)**.** *If $B_1, \dots, B_n$ is a finite collection of balls in X, then there is a disjoint subcollection $B_{i_1}, \dots, B_{i_M}$ such that*

$$\left|\bigcup_{i=1}^N B_i\right| \leqslant c \sum_{j=1}^M |B_{i_j}|.$$

Proof. Consider the following greedy selection procedure. Let B_{i_1} be the ball in our collection of maximal radius. Given that we have selected $B_{i_1}, \dots, B_{i_k}$, let $B_{i_{k+1}}$ be the ball of largest radius not intersecting previous balls selected if possible. Continue doing this until we cannot select any further balls. If B_j is any ball not chosen by this procedure, it must intersect a ball with radius at least as big as B_j itself. But this means that

$$\bigcup_{i=1}^N B_i \subset \bigcup_{j=1}^M B_{i_j}^*.$$

Thus

$$\left|\bigcup_{i=1}^N B_i\right| \leqslant \sum_{j=1}^M |B_{i_j}^*| \leqslant c \sum_{j=1}^M |B_{i_j}|. \qquad \square$$

We have already indicated our proof strategy for proving a weak type bound for the maximal operator, but let us now do things more rigorously.

Theorem 16.4. *For any $f \in L^1(X)$,*

$$\|Mf\|_{L^{1,\infty}(X)} \leqslant c\|f\|_{L^1(X)}.$$

Proof. Set

$$E_\lambda = \{x \in \mathbf{R}^d : Mf(x) > \lambda\}.$$

Fix a compact subset K of E_λ of finite measure. Then K is covered by finitely many balls $B_1, \dots, B_N$ such that on each ball B_i,

$$\int_{B_i} |f(y)|\, dy > \lambda |B_i|.$$

Using the Vitali lemma, extract a disjoint subfamily $B_{i_1}, \dots, B_{i_M}$ with

$$\left| \sum_{j=1}^{M} B_{i_j} \right| \leqslant c \sum_{j=1}^{M} |B_{i_j}|.$$

Then

$$\|f\|_{L^1(X)} > \lambda \sum_{j=1}^{M} |B_{j_i}| \geqslant \frac{\lambda}{c} \left| \bigcup_{j=1}^{M} B_{j_i} \right| \geqslant \frac{\lambda |K|}{c}.$$

Rearranging gives

$$|K| \leqslant \frac{c\|f\|_{L^1(X)}}{\lambda}.$$

Since K was arbitrary, inner regularity gives

$$|E_\lambda| \leqslant \frac{c\|f\|_{L^1(X)}}{\lambda}.$$

Since λ was arbitrary, the proof is complete. □

Remark. The same covering-type argument also gives the boundedness of the *uncentered* Hardy-Littlewood maximal function

$$M'f(x) = \sup_{x \in B} \frac{1}{|B|} \int_B |f(y)|\, dy$$

where B ranges over all balls $\{B(x', \delta) : x' \in X, \delta > 0\}$. Since the supremum is over more balls, we have $Mf(x) \leqslant M'f(x)$ for each $x \in X$. If we assume a stronger condition than we did previously, that if $B(x', \delta) \cap B(x, \delta) \neq \varnothing$, then $B(x', \delta) \subset B(x, c\delta)$ (so that $B^*(x, \delta) \subset B(x, c\delta)$), and that $|B(x, c\delta)| \leqslant c'|B(x, \delta)|$, then we also find $M'f(x) \leqslant c'Mf(x)$. Thus in these situations, M and M' are roughly equivalent operators. We shall find these assumptions are also useful for generalizing the Calderon-Zygmund type decompositions that come up in the real-variable analysis of singular integrals.

Remark. We can exploit the ordering of the real line to show that for any family of intervals $\{I_\alpha\}$ covering a compact set K, there is a subcover $I_1,\dots,I_N$ such that any point in $\mathbf{R}$ is contained in at most two of the intervals. A modification of the argument above shows this gives the slightly better bound $\|Mf\|_{L^{1,\infty}(\mathbf{R})} \leqslant 2\|f\|_{L^1(\mathbf{R})}$, rather than the bound $\|Mf\|_{L^{1,\infty}(\mathbf{R})} \leqslant 3\|f\|_{L^1(\mathbf{R})}$.

16.2 Dyadic Methods

There are many different techniques for showing the boundedness of the maximal operator. Let us consider some *dyadic methods* for proving the inequality. Recall that the set of dyadic cubes is

$$\{Q_{n,k} : n \in \mathbf{Z}, k \in 2^n \mathbf{Z}^d\}$$

where $Q_{n,k}$ is the cube $[k_1, k_1 + 2^n] \times \cdots \times [k_d, k_d + 2^n]$. We note that dyadic cubes nest within one another much more easily than balls do (cubes are either nested or disjoint). In particular, if $Q_1,\dots,Q_N$ is any collection of dyadic cubes, there exists an almost disjoint subcollection $Q_{i_1},\dots,Q_{i_k}$ with $Q_{i_1} \cup \cdots \cup Q_{i_k} = Q_1 \cup \cdots \cup Q_N$. In particular, this operates as a Vitali-type covering lemma with a constant independant of d, so if we define the *dyadic* Hardy-Littlewood maximal operator

$$M_\Delta f(x) = \sup_{x \in Q} \frac{1}{|Q|} \int_Q |f(y)|\, dy$$

then we easily obtain the bound $\|M_\Delta f\|_{L^{1,\infty}(\mathbf{R}^d)} \leqslant \|f\|_{L^1(\mathbf{R}^d)}$, with no implicit constant depending on d. The bound $\|M_\Delta f\|_{L^\infty(\mathbf{R}^d)} \leqslant \|f\|_{L^\infty(\mathbf{R}^d)}$ is easy, so interpolation gives $\|M_\Delta f\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}$ for all $1 < p \leqslant \infty$, with a constant now *independant of dimension*.

Two families of sets $\{B(x,\delta) : x \in \mathbf{R}^d, \delta > 0\}$ and $\{B'(x,\delta) : x \in \mathbf{R}^d, \delta > 0\}$ are *equivalent* if there exists c_1, c_2 such that

$$B'(x, c_1\delta) \subset B(x,\delta) \subset B(x, c_2\delta).$$

It follows that the resultant maximal averages from the two sets pointwise differ from one another by a universal constant. This allows us to obtain bounds for maximal averages over cubes centred at a point, and ellipses with bounded eccentricity, etc. If for $2^k \leqslant \delta \leqslant 2^{k+1}$, we set $B(x,\delta)$ to be a dyadic cube with sidelength $1/2^k$, then the family $\{B(x,\delta)\}$ is *not* equivalent to the usual family of cubes, but below we show that bounds on the two maximal operators are still equivalent.

If Q is a dyadic cube, then it is contained in a ball B with $|Q| \lesssim_d B$. It follows that for any function f and $x \in \mathbf{R}^d$,

$$M_\Delta f(x) \lesssim_d Mf(x).$$

Thus bounds on M automatically give bounds on M_Δ. The opposite pointwise inequality is unfortunately, *not true*. For instance, if f is the indicator function on $[0,1]$. Then $M_\Delta f$ is supported on $[0,1]$, but Mf is positive on the entirety of $\mathbf{R}$. To reduce the study of M to the study of M_Δ, we must instead rely on the $1/3$ *translation trick* of Michael Christ.

Lemma 16.5. *Let $I \subset [0,1]$ be an interval. Then there exists an interval J, which is either a dyadic interval, or a dyadic interval shifted by $1/3$, such that $I \subset J$ and $|J| \lesssim |I|$.*

Proof. Let $I = [a,b]$. Perform a binary expansion of a and b, writing

$$a = 0.a_1 a_2 \dots \quad \text{and} \quad b = b_1 b_2 \dots.$$

Let n be the first value where $a_n \neq b_n$. Then $a_n = 0$ and $b_n = 1$. Then $[a,b]$ is contained in the dyadic interval

$$Q_1 = \left[0.a_1 \dots a_{n-1}, 0.a_1 \dots a_{n-1} + 1/2^{n-1}\right]$$

which has length $1/2^{n-1}$. Find $0 \leqslant i < \infty$ such that

$$a = 0.a_1 \dots a_{n-1} 01^i 0 \dots$$

and $0 \leqslant j < \infty$ such that

$$b = 0.a_1 \dots a_{n-1} 10^j 1.$$

If no such j exists, then $b = 0.a_1 \dots a_{n-1} 1$, and so $[a,b]$ is contained in the rational interval

$$Q_2 = \left[0.a_1 \dots a_{n-1} 01^i, 0.a_1 \dots a_{n-1} 01^i + 1/2^{n+i}\right]$$

and $b-a \geqslant 1/2^{n+i+1}$, so $|Q_2| \leqslant 2(b-a)$. Now if $i \leqslant 5$ or $j \leqslant 5$, then $b-a \geqslant 1/2^{n+5}$, so $|Q_1| \leqslant 2^5(b-a)$. On the other hand, if $i \geqslant 5$ and $j \geqslant 5$, we find $b-a \geqslant 1/2^{n+\min(i,j)}$. Then we can find a dyadic interval Q_3 and $2 \leqslant r \leqslant 5$ such that

$$1/3+Q_3 = \left[0.a_1\dots a_{n-1}01^{\min(i,j)-r}1010\dots, 0.a_1\dots a_{n-1}01^{i-r}1010\dots + 1/2^{n+\min(i,j)-r}\right]$$

and so $1/3+Q_3$ contains $[a,b]$ and $|Q_3| = 1/2^{n+\min(i,j)-r} \leqslant 2^5(b-a)$. □

It follows that for each $x \in \mathbf{R}^d$, and any function f,

$$Mf(x) \lesssim_d (M_\Delta f)(x) + (M_\Delta \mathrm{Trans}_{1/3} f)(x).$$

Since the L^p norms are translation invariant, this implies that the dyadic maximal operator and the maximal operator satisfy equivalent bounds, with operator norms differing by a constant depending on n. Since we independently obtained bounds on M_Δ, this section provides an alternate proof to the boundedness of M.

There is an alternate way to view the operator M_Δ. For each integer n, we let $\mathcal{B}(n)$ denote the family of all sidelength $1/2^n$ dyadic cubes. Thus $\mathcal{B}(n)$ gives a decomposition of $\mathbf{R}^d$ into an almost disjoint union of cubes. If we define the conditional expecation operators

$$E_n f(x) = \sum_{Q \in \mathcal{B}(n)} \left(\frac{1}{|Q|} \fint_Q f \right) \cdot \mathbf{I}_Q$$

then $M_\Delta f = \sup_{n \in \mathbf{Z}} E_n f$. In particular, it is easy to see from the bounds on M_Δ that for any $f \in L^1_{\text{loc}}(\mathbf{R}^d)$, $\lim_{n \to \infty} E_n f(x) = f(x)$ holds for almost every $x \in \mathbf{R}^d$. It is simple to conclude from this result a very useful technique, known as the *Calderón-Zygmund decomposition*.

Theorem 16.6. *Given $f \in L^1(\mathbf{R}^d)$ and $\lambda > 0$, we can write $f = g + b$, where $\|g\|_{L^\infty(\mathbf{R}^d)} \lesssim_d \lambda$, and there is an almost disjoint family of dyadic cubes $\{Q_i\}$ such that g is supported on $\bigcup_i Q_i$,*

$$\sum_i |Q_i| \leqslant \frac{\|f\|_{L^1(\mathbf{R}^d)}}{\lambda},$$

and for each i,

$$\int_{Q_i} f(y)\, dy = 0.$$

We also have $\|g\|_{L^1(\mathbf{R}^d)}, \|b\|_{L^1(\mathbf{R}^d)} \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}$.

Proof. Write $E = \{x : M_\Delta f(x) > \lambda\}$. By the dyadic Hardy-Littlewood maximal inequality,

$$|E| \leqslant \frac{\|f\|_{L^1(\mathbf{R}^d)}}{\lambda}.$$

Because f is integrable, $E \neq \mathbf{R}^d$. Thus we can write E as the almost disjoint union of dyadic cubes $\{Q_i\}$, such that for each i,

$$\int_{Q_i} |f(x)|\, dx > \lambda |Q_i|,$$

and also, if R_i is the parent cube of Q_i,

$$\int_{R_i} |f(x)|\, dx \leqslant \lambda |R_i|.$$

This can be done by a greedy strategy, taking the union of dyadic cubes of largest sidelength contained in E. This means

$$\int_{Q_i} |f(x)|\, dx \leqslant \int_{R_i} |f(x)|\, dx \leqslant \lambda |R_i| \leqslant 2^d \lambda |Q_i|.$$

Define

$$g(x) = \begin{cases} f(x) & : x \notin E, \\ \frac{1}{|Q_i|} \int_{Q_i} f(x)\, dx & : x \in Q_i \text{ for some } i. \end{cases}$$

For almost every $x \in E^c$, $|f(x)| \leqslant \lambda$, since $E_n f(x) \leqslant \lambda$ for each n, and $E_n f(x) \to f(x)$ as $n \to \infty$ for almost every x. Conversely, if $x \in Q_i$ for some i, then

$$\left| \frac{1}{|Q_i|} \int_{Q_i} f(x)\, dx \right| \leqslant \frac{1}{|Q_i|} \int_{Q_i} |f(x)|\, dx \leqslant 2^d \lambda.$$

Thus $\|g\|_{L^\infty(\mathbf{R}^d)} \lesssim_d \lambda$. If we define $b = f - g$, then b is supported on $\bigcup Q_i = E$, and for each i,

$$\int_{Q_i} b(x)\, dx = \int_{Q_i} \left(f(x) - \frac{1}{|Q_i|} \int_{Q_i} f(y)\, dy \right) dx = 0. \qquad \square$$

The Calderon-Zygmund theorem will be very useful to us in the sequel, especially when we analyze the theory of singular integrals.

16.3 Lebesgue Density Theorem

If E is a measurable subset of $\mathbf{R}^d$, and $x \in \mathbf{R}^d$, we say x is a point of *Lebesgue density* of E, or has *full metric density* if

$$\lim_{\delta \to 0} \frac{|B(x,\delta) \cap E|}{|B(x,\delta)|} = 1$$

This means that for any $\varepsilon > 0$, the inequality $|B(x,\delta) \cap E| \geqslant (1-\varepsilon)|B(x,\delta)|$ holds for suitably small δ, so E asymptotically contains as large a fraction of the local points around x as is possible. Since $\chi_E \in L^1_{\text{loc}}(\mathbf{R}^d)$, we can apply the Lebesgue differentiation theorem to immediately obtain an interesting result.

Theorem 16.7 (Lebesgue Density Theorem)**.** *If E is a measurable subset, then almost every point in E is a point of Lebesgue density, and almost every point not in E is not a point of Lebesgue density.*

The fact that a point is a point of Lebesgue density implies the existence of large sets of rigid patterns in E. A simple corollary is that any set of positive Lebesgue measure contains arbitrarily long arithmetic progressions.

Theorem 16.8. *Let $E \subset \mathbf{R}^d$ be a set of nonzero Lebesgue measure. Then for any non-zero $a_1, \dots, a_N \in \mathbf{R}$ there exists $x \in E$ and $c \in \mathbf{R}$ such that*

$$a_1 x + c, \dots, a_N x + c \in E.$$

Proof. Without loss of generality, by translation we may assume 0 is a point of Lebesgue density of E. We then claim that we can set $c = 0$. It is simple to see that if t_0 is a point of Lebesgue density for a set E and a set F, then it is also a point of Lebesgue density for $E \cap F$. In particular 0 is a point of Lebesgue density for $E \cap a_1^{-1} E \cap \dots \cap a_N^{-1} E$, which means the set is nonempty. If $y \in E \cap a^{-1} E \cap \dots \cap a_N^{-1} E$, then $y, a_1 y, \dots, a_N y \in E$. □

If f is locally integrable, the *Lebesgue set* of f consists of all points $x \in \mathbf{R}^d$ such that $f(x)$ is finite and

$$\lim_{\delta \to 0} \frac{1}{|B_\delta|} \int_{B(x,\delta)} |f(y) - f(x)| \, dy = 0.$$

If f is continuous at x, it is obvious that x is in the Lebesgue set of f, and if x is in the Lebesgue set of f, then $A_\delta f(x) \to f(x)$ as $\delta \to 0$.

Theorem 16.9. *If $f \in L^1_{loc}(\mathbf{R}^d)$, almost every point is in the Lebesgue set of f.*

Proof. For each rational number p, the function $|f - p|$ is measurable, so that there is a set E_p of measure zero such that for $x \in E_p^c$,

$$\lim_{\delta \to 0} \fint_{B(x,\delta)} |f(y) - p| \; dy \to |f(x) - p|.$$

Taking unions, we conclude that $E = \bigcup E_p$ is a set of measure zero. Suppose $x \in E^c$, and $f(x)$ is finite. For any $\varepsilon > 0$, there is a rational p such that $|f(x) - p| < \varepsilon$, and we know the equation above holds, so

$$\lim_{\delta \to 0} \fint_{B(x,\delta)} |f(y) - f(x)| \; dy \leqslant \limsup_{\delta \to 0} \fint_{B(x,\delta)} (|f(y) - p| + |p - f(x)|) \; dy \leqslant 2\varepsilon.$$

We can then let $\varepsilon \to 0$. Since $f(x)$ is finite for almost all x, this completes the proof. $\square$

It is interesting to note that for any $f \in L^1_{\text{loc}}(\mathbf{R}^d)$, there is $g \in L^1_{\text{loc}}(\mathbf{R}^d)$ such that $f = g$ almost everywhere, and the Lebesgue set of g is maximal. One choice is to define

$$g(x) = \limsup_{\delta \to 0} \fint_{B(x,\delta)} f(y) \; dy.$$

This gives us a way to identify a single pointwise defined function from an equivalence class of locally integrable functions defined distributionally. However, this isn't often done, because it doesn't really help in the analysis of integrable functions. Often the Lebesgue set of f is defined to be the Lebesgue set of g, when once to think of the Lebesgue set as a distributional invariant of f.

16.4 Ergodic Averages

One consequence of the more general setting to homogenous spaces is that if we define, for $f : \mathbf{Z} \to \mathbf{C}$,

$$Mf(n) = \sup_{N > 0} \sum_{m=1}^{N} |f(n+m)|$$

then $\|Mf\|_{L^{1,\infty}(\mathbf{Z})} \leqslant 2\|f\|_{L^1(\mathbf{Z})}$ and $\|Mf\|_{L^p(\mathbf{Z})} \lesssim_p \|f\|_{L^p(\mathbf{Z})}$ for $1 < p \leqslant \infty$. A consequence of this inequality is a pointwise convergence result that emerges from ergodic theory. We recall that a *measure preserving system* is a probability space X together with a measure preserving transformation $T : X \to X$.

Theorem 16.10. *Let X and T form a measure preserving transformation. Then for all $f \in L^1(X)$ and almost every $x \in X$, the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f(x)$$

exists.

Proof. Fix $N_0 > 0$ and $f \in L^1(X)$, and define a measurable function F on $X \times [2N_0]$ by defining

$$F(x,n) = T^n f(x).$$

Let

$$MF(x,n) = \sup_{1\leqslant N\leqslant N_0} \frac{1}{N} \sum_{m=1}^{N} T^{n+m} f(x).$$

Then the integer-valued maximal inequality implies that

$$\|MF\|_{l^{1,\infty}[N_0]} \lesssim \|F\|_{l^1[2N_0]}$$

and integrating in X, that

$$\|MF\|_{l^{1,\infty}[N_0]L^1(X)} \lesssim \|F\|_{l^1[2N_0]L^1(X)} = \|F\|_{L^1(X\times[2N_0])} = 2N_0\|f\|_{L^1(X)}.$$

TODO FINISH THIS. $\square$

16.5 Approximations to the Identity

We now switch to the study of how we can approximate functions by convolutions of concentrated functions around the origin. In this section we define the various classes of such functions which give convergence results, to various degrees of strength. We say a family of integrable functions $\{K_\alpha : \alpha > 0\}$ in $\mathbf{R}^d$ is a *good kernel* if it is bounded in the L^1 norm, for every $\alpha > 0$,

$$\int K_\alpha(x)\, dx = 1$$

and if for every $\delta > 0$,

$$\lim_{\alpha \to 0} \int_{|x| \geqslant \delta} |K_\alpha(x)|\, dx \to 0.$$

It requires only basic analysis to verify good kernel convergence.

Theorem 16.11. *If $\{K_\alpha\}$ is a good kernel, then for any absolutely integrable function f, $\lim_{\alpha \to 0} f * K_\alpha = f$ in $L^1(\mathbf{R}^d)$, and for any continuity point x of f, $\lim_{\alpha \to 0} (f * K_\alpha)(x) = f(x)$.*

Proof. Note that

$$\begin{aligned}
\|(f * K_\alpha) - f\|_1 &= \int |(f * K_\alpha)(x) - f(x)|\, dx \\
&= \int \left| \int K_\alpha(y)[f(x-y) - f(x)]\, dy \right| dx \\
&\leqslant \int |K_\alpha(y)| \|T_y f - f\|_1\, dy
\end{aligned}$$

where $(T_y f)(x) = f(x-y)$. We know that $\|T_y f - f\|_1 \to 0$ as $y \to 0$. Thus, for each ε, we can pick δ such that if $|y| < \delta$, $\|T_y f - f\|_1 \leqslant \varepsilon$, and if we pick α large enough that $\int_{|y| \geqslant \delta} |K_\alpha(y)|\, dy \leqslant \varepsilon$, and then

$$\|(f * K_\alpha) - f\|_1 \leqslant \varepsilon \int_{|y| < \delta} |K_\alpha(y)|\, dy + 2\|f\|_1 \int_{|y| \geqslant \delta} |K_\alpha(y)|\, dy \leqslant \varepsilon[\|K_\alpha\|_1 + 2\|f\|_1]$$

Since $\|K_\alpha\|_1$ is universally bounded over α, we can let $\varepsilon \to 0$ to obtain convergence. If x is a fixed point of continuity, and for a given $\varepsilon > 0$, we pick $\delta > 0$ with $|f(y) - f(x)| \leqslant \varepsilon$ for $|y - x| < \delta$, then

$$\begin{aligned}
|(f * K_\alpha)(x) - f(x)| &= \left| \int_{-\infty}^{\infty} f(y) K_\alpha(x-y)\, dy - f(x) \right| \\
&= \left| \int_{-\infty}^{\infty} [f(y) - f(x)] K_\alpha(x-y)\, dy \right| \\
&= \left| \int_{-\delta}^{\delta} [f(y) - f(x)] K_\alpha(x-y)\, dy \right| \\
&\quad + \left| \int_{|y| \geqslant \delta} [f(y) - f(x)] K_\alpha(x-y)\, dy \right| \\
&\leqslant \varepsilon \|K_\alpha\|_1 + [\|f\|_1 + f(x)] \int_{|y| \geqslant \delta} |K_\alpha(y)|\, dy
\end{aligned}$$

If $\|K_\alpha\|_1 \leqslant M$ for all α, and we choose α large enough that $\int_{|y| \geqslant \delta} |K_\alpha(y)| \leqslant \varepsilon$, then we conclude the value about is bounded by $\varepsilon[M + \|f\|_1 + f(x)]$, and we can then let $\varepsilon \to 0$. $\square$

To obtain almost sure pointwise convergence of $f * K_\alpha$ to f, we must place stronger conditions on our family. We say a family $K_\delta \in L^1(\mathbf{R}^d)$, is an *approximation to the identity* if $\int K_\delta = 1$, and

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

where the constant bound is independent of x and δ. These assumptions are stronger than being a good kernel, because if K_δ is an approximation to the identity, then

$$\int_{|x| \geqslant \varepsilon} |K_\delta(x)| \leqslant \int_\varepsilon^\infty \int_{S^{d-1}} \frac{C\delta}{r}\, d\sigma dr = C\delta|S^{n-1}| \int_\varepsilon^\infty \frac{dr}{r} \leqslant \frac{C\delta|S^{n-1}|}{\varepsilon}$$

which converges to zero as $\delta \to 0$. Combined with

$$\int_{|x| < \varepsilon} |K_\delta(x)| \leqslant C \int_0^\varepsilon \int_{S^{d-1}} \frac{r^{d-1}}{\delta^d} d\sigma dr = \frac{C\varepsilon^d |S^{n-1}|}{d\delta^d}$$

This calculation also implies

$$\|K_\delta\|_1 \leqslant C|S^{n-1}| \left[\frac{\delta}{\varepsilon} + \frac{\varepsilon^d}{\delta^d} \right]$$

Setting $\varepsilon = \delta$ optimizes this value, and gives a bound

$$\|K_\delta\|_1 \leqslant 2C|S^{n-1}|$$

So an approximation to the identity is a stronger version of a good kernel.

Example. *If φ is a bounded function in $\mathbf{R}^d$ supported on the closed ball of radius one with $\int \varphi(x)\, dx = 1$, then $K_\delta(x) = \delta^{-d}\varphi(\delta^{-1}x)$ is an approximation to the identity, because by a change of variables, we calculate*

$$\int_{\mathbf{R}^d} \frac{\varphi(\delta^{-1}x)}{\delta^d} = \int_{\mathbf{R}^d} \varphi(x) = 1$$

Because φ is bounded, we find

$$|K_\delta(x)| \leqslant \frac{\|\varphi\|_\infty}{\delta^d}$$

Now K_δ is supported on a disk of radius δ, this bound also shows

$$|K_\delta(x)| \leqslant \frac{\delta\|\varphi\|_\infty}{|x|^{d+1}}$$

and so K_δ is an approximation to the identity. If φ is an arbitrary integrable function, then K_δ will only be a good kernel.

Example. *The Poisson kernel in the upper half plane is given by*

$$P_y(x) = \frac{1}{\pi}\frac{y}{x^2+y^2}$$

where $x \in \mathbf{R}$, and $y > 0$. It is easy to see that

$$P_y(x) = y^{-1}P_1(xy^{-1})$$

And

$$\int \frac{1}{1+x^2} = \arctan(\infty) - \arctan(-\infty) = \pi$$

We easily obtain the bounds

$$|P_y(x)| \leqslant \frac{\|P_1\|_\infty}{y} \qquad |P_y(x)| \leqslant \frac{y}{\pi|x|^2}$$

so the Poisson kernel is an approximation to the identity.

Example. *The heat kernel in $\mathbf{R}^d$ is defined by*

$$H_t(x) = \frac{e^{-|x|^2/4t}}{(4\pi t)^{d/2}}$$

where $\delta = t^{1/2} > 0$. Then $H_t(x) = \delta^{-d}H_1(x\delta^{-1})$, and

$$\int e^{-|x|^2/4} = \frac{1}{2^d}\int e^{-|x|^2} = \frac{|S^{n-1}|}{2^d}\int_0^\infty r^{d-1}e^{-r^2}\,dr$$

Example. *The Poisson kernel for the disk is*

$$\frac{P_r(x)}{2\pi} = \begin{cases} \frac{1}{2\pi}\frac{1-r^2}{1-2r\cos x + r^2} & : |x| \leqslant \pi \\ 0 & : |x| > \pi \end{cases}$$

where $0 < r < 1$*, and* $\delta = 1 - r$*.*

Example. *The Féjer kernel is*

$$\frac{F_N(x)}{2\pi} = \begin{cases} \frac{1}{2\pi N}\frac{\sin^2(Nx/2)}{\sin^2(x/2)} \end{cases}$$

where $\delta = 1/N$*.*

As $\delta \to 0$, we may think of the K_δ as 'tending to the unit mass' Dirac delta function δ at the origin. δ may be given a precise meaning, either in the theory of Lebesgue-Stieltjes measures or as a 'generalized function', but we don't need it to discuss the actual convergence results of the functions K_δ.

Theorem 16.12. *If* $\{K_\delta\}$ *is an approximation to the identity, and* f *is integrable on* $L^1(\mathbf{R}^d)$*, then* $(f * K_\delta)(x) \to f(x)$ *for every* x *in the Lebesgue set of* f*, and* $f * K_\delta$ *converges to* f *in the* L^1 *norm.*

Proof. We rely on the fact that if x is in the Lebesgue set, then the function

$$A(r) = \frac{1}{r^d}\int_{|y| \leqslant r} |f(x-y) - f(x)|\, dy$$

is a bounded continuous function of $r > 0$, converging to 0 as $r \to 0$. This means that if $\Delta(y) = |f(x-y) - f(x)||K_\delta(y)|$, then

$$\int \Delta(y)\, dy = \int_{|y| \leqslant \delta} \Delta(y) + \sum_{k=0}^{\infty} \int_{2^k\delta \leqslant |y| \leqslant 2^{k+1}\delta} \Delta(y)$$

The first term is easily upper bounded by $CA(\delta)$, and the k'th term of the sum by $C'2^{-k}A(2^{k+1}\delta) \leqslant C''2^{-k}$ for constants C', C'' that do not depend on δ. Letting $\delta \to 0$ gives us the convergence result. □

16.6 The Strong Maximal Function

TODO

16.7 The Tangential Poisson Maximal Function

TODO

Chapter 17

Aside: Differentiability of Measurable Functions

A simple consequence of our results is a kind of fundamental theorem of calculus. If $f \in L^1(\mathbf{R})$, then we can define $F \in C(\mathbf{R})$ by setting

$$F(t) = \int_{-\infty}^{t} f(s)\, ds.$$

It follows from the maximal functions bounds we've established that F is differentiable almost everywhere, and $F'(t) = f(t)$ for almost every t. Thus the fundamental theorem of calculus holds in this setting, i.e.

$$F(t) = \int_{-\infty}^{t} F'(s)\, ds.$$

Let us now consider *what* conditions we can assume on a measurable function f such that f is differentiable almost everywhere, such that $f' \in L^1(\mathbf{R})$, and such that

$$f(t) = \int_{-\infty}^{t} f'(s)\, ds$$

holds for almost every $t \in \mathbf{R}$. Clearly this is equivalent to finding which functions are expressed as the indefinite integral of an integrable function.

We shall find that if f has *bounded variation*, then most of these problems are answered. If f is a complex valued function on $[a,b]$, and P is a partition, we can consider it's variation on a partition $P = a \leqslant t_0 < \cdots <$

$t_n \leqslant b$ to be

$$V(f,P) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|$$

we say f has *bounded variation* if there is a constant M such that for any partition P, $V(f,P) \leqslant M$. This implies that, since the net $P \mapsto V(f,P)$ is increasing, the net converges to a value $V(f) = V(f,a,b)$, the *total variation* of f on $[a,b]$.

The problem of the variation of a function is very connected to the problem of the *rectifiability of curves*. If $x : [a,b] \to \mathbf{R}^d$ parameterizes a continuous curve in the plane, then, for a given partition $P = a \leqslant t_0 \leqslant \dots \leqslant t_n$, we can consider an approximate length

$$L_P(x) = \sum_{k=1}^{n} |x(t_i) - x(t_{i-1})|$$

If x has a reasonable notion of length, then the straight lines between $x(t_{i-1})$ and $x(t_i)$ should be shorter than the length of x between t_{i-1} and t_i. It therefore makes sense to define the *length* of x as

$$L(x) = \sup L_P(x)$$

The triangle inequality implies that the map $P \mapsto L_P(x)$ is an increasing net, so L is also the limit of the meshes as they become finer and finer. If $L(x) < \infty$, we say x is a *rectifiable curve*. One problem is to determine what analytic conditions one must place on x in order to guarantee regularity, and what further conditions guarantee that, if x_i is differentiable almost everywhere,

$$L(x) = \int_a^b \sqrt{x_1'(t)^2 + \dots + x_n'(t)^2}\, dt$$

Considering rectifiable curves leads directly to the notion of a function with bounded variation.

Theorem 17.1. *A curve x is rectifiable iff each x_i has bounded variation.*

Proof. We can find a universal constants $A, B > 0$ such that for any $x, y \in \mathbf{R}^d$,

$$A \sum |x_i - y_i| \leqslant |x - y| \leqslant B \sum |x_i - y_i|$$

This means that if P is a partition of $[a,b]$, then

$$A\sum_{ij}|x_j(t_i)-x_j(t_{i-1})| \leqslant \sum|x(t_i)-x(t_{i-1})| \leqslant B\sum_{ij}|x_j(t_i)-x_j(t_{i-1})|$$

So $A\sum V(x_i,P) \leqslant L_P(x) \leqslant B\sum V(x_i,P)$ gives the required result. $\square$

Example. *If f is a real-valued, monotonic, increasing function on $[a,b]$, then f has bounded variation, and one can verify that $V(f) = f(b) - f(a)$.*

Example. *If f is differentiable at every point, and f' is bounded, then f has bounded variation. The mean value theorem implies that if $|f'| \leqslant M$, then for all $x, y \in [a,b]$,*

$$|f(x) - f(y)| \leqslant M|x-y|$$

This implies that $V(f,P) \leqslant M(b-a)$ for all partitions P.

Example. *Consider the functions f defined on $[0,1]$ with*

$$f(x) = \begin{cases} x^a \sin(x^{-b}) & : 0 < x \leqslant 1 \\ 0 & : x = 0 \end{cases}$$

Then f has bounded variation on $[0,1]$ if and only if $a > b$. The function oscillates from increasing to decreasing on numbers of the form $x = (n\pi)^{-1/b}$, so the total variation is described as

$$V(f) = 1 + \sum_{n=1}^{\infty}(n\pi)^{-a/b} + ((n+1)\pi)^{-a/b}$$

This sum is finite precisely when $a/b > 1$. Thus functions of bounded variation cannot oscillate too widely, too often.

The next result is a decomposition theorem for bounded variation functions into bounded increasing and decreasing functions. We define the *positive variation* of a real valued function f on $[a,b]$ to be

$$P(f,a,b) = \sup_P \sum_{f(t_i) \geqslant f(t_{i-1})} f(t_i) - f(t_{i-1})$$

The *negative variation* is

$$N(f,a,b) = \sup_P \sum_{f(t_i) \leqslant f(t_{i-1})} -[f(t_i) - f(t_{i-1})]$$

Note that for each partition P, the sums of the two values above add up to the variation with respect to the partition.

Lemma 17.2. *If f is real-valued and has bounded variation on $[a,b]$, then for all $a \leqslant x \leqslant b$,*

$$f(x) - f(a) = P(f,a,x) - N(f,a,x)$$
$$V(f) = P(f,a,b) + N(f,a,b)$$

Proof. Given ε, there exists a partition $a = t_0 < \cdots < t_n = x$ such that

$$\left| P(f,a,x) - \sum_{f(t_i) \geqslant f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| < \varepsilon$$

$$\left| N(f,a,x) + \sum_{f(t_i) \leqslant f(t_{i-1})} f(t_i) - f(t_{i-1}) \right| < \varepsilon$$

It follows that

$$|f(x) - f(a) - [P(f,a,x) - N(f,a,x)]| < 2\varepsilon$$

and we can then take $\varepsilon \to 0$. The second identity follows the same way. $\square$

A real function f on $[a,b]$ has bounded variation if and only if f is the difference of two increasing bounded functions, because if f has bounded variation, then

$$f(x) = [f(a) + P(f,a,x)] - N(f,a,x)$$

is the difference of two bounded increasing functions. On the other hand, the difference of two bounded increasing functions is clearly of bounded variation. A complex function has bounded variation if and only if it is the linear combination of four increasing functions in each direction.

Theorem 17.3. *If f is a continuous function of bounded variation, then*

$$x \mapsto V(f,a,x) \qquad x \mapsto V(x,b)$$

are continuous functions.

Proof. $V(f,a,x)$ is an increasing functin of x, so for continuity on the left it suffices to prove that for each x and ε, there is $x_1 < x$ such that $V(f,a,x_1) \geqslant V(f,a,x) - \varepsilon$. If we consider a partition

$$P = \{a = t_0 < \cdots < t_n = x\}$$

where $|V(f,P) - V(f,a,x)| \leqslant \varepsilon$, then by continuity of f at x, there is $t_{n-1} < x_1 < x$ with $|f(x) - f(x_1)| < \varepsilon$, and then if we modify P to obtain Q by swapping t_n with x_1, we find

$$\begin{aligned} V(f,a,x_1) \geqslant V(f,Q) &= V(f,P) - |f(x) - f(t_{n-1})| + |f(x_1) - f(t_{n-1})| \\ &\geqslant V(f,P) - \varepsilon \geqslant V(f,a,x) - \varepsilon \end{aligned}$$

A similar argument gives continuity on the right, and the continuity as the left bound of the interval changes. □

To obtain the differentiation theorem for functions of bounded variation, we require a lemma of F. Riesz.

Lemma 17.4 (Rising Sun lemma)**.** *If f is real-valued and continuous on $\mathbf{R}$, and E is the set of points x where there exists $h > 0$ such that $f(x+h) > f(x)$, then, provided E is non-empty, it must be open, and can be written as a union of disjoint intervals (a_n, b_n), where $f(b_n) = f(a_n)$. If f is continuous on $[a,b]$, then E is still an open subset of $[a,b]$, and can be written as the disjoint union of countably many intervals, with $f(b_n) = f(a_n)$ except if $a_n = a$, where we can only conclude $f(a_n) \leqslant f(b_n)$.*

Proof. The openness is clear, and the fact that E can be broken into disjoint intervals follows because of the characterization of open sets in $\mathbf{R}$. If

$$E = \bigcup (a_n, b_n)$$

Then $f(a_n + h) \leqslant f(a_n)$ and $f(b_n + h) \leqslant f(b_i)$, implying in particular that $f(b_n) \leqslant f(a_n)$, If $f(b_n) < f(a_n)$, then choose $f(b_n) < c < f(a_n)$. The intermediate value theorem implies there is x with $f(x) = c$, and we may choose the largest $x \in [a_n, b_n]$ for which this is true. Then since $x \in (a_n, b_n)$, there is $y \in (x, b_i)$ with $f(x) < f(y)$, and by the intermediate value theorem, since $f(b_n) < f(x) < f(y)$, there must be $x' \in (y, b_n)$ with $f(x') = c$, contradicting that x was chosen maximally. The proof for closed intervals operates on the same principles. □

Theorem 17.5. *If f is increasing and continuous on $[a,b]$, then f is differentiable almost everywhere. That is,*

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

exists for almost every $x \in [a,b]$, f' is measurable, and

$$\int_a^b f'(x) \leqslant f(b) - f(a)$$

In particular, if f is bounded on $\mathbf{R}$, then f' is integrable on $\mathbf{R}$.

Proof. the theorem in the It suffices to assume that f is increasing, and we shall start by proving case assuming f is continuous. We define

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

and the four *Dini derivatives*

$$D_+ f(x) = \liminf_{h \downarrow 0} \Delta_h f(x) \quad D^+ f(x) = \limsup_{h \downarrow 0} \Delta_h f(x)$$

$$D_- f(x) = \liminf_{h \uparrow 0} \Delta_h f(x) \quad D^- f(x) = \limsup_{h \uparrow 0} \Delta_h f(x)$$

Clearly, $D_+ f \leqslant D^+ f$ and $D_- f \leqslant D^- f$, It suffices to show $D^+ f(x) < \infty$ for almost every x, and $D^+ f(x) \leqslant D_- f(x)$ for almost every x, because if we consider the function $g(x) = -g(-x)$, then we obtain $D^- f(x) \leqslant D_+ f(x)$ for almost every x, so

$$D^+ f(x) \leqslant D_- f(x) \leqslant D^- f(x) \leqslant D_+ f(x) \leqslant D^+ f(x) < \infty$$

for almost every x, implying all values are equal, and that the derivative exists at x.

For a fixed $\gamma > 0$, consider $E_\gamma = \{x : D^+ f(x) > \gamma\}$. Since each $\Delta_h f$ is continuous, the supremum of the $\Delta_h f$ over any index set is lower semicontinuous, and since

$$D^+ f(x) = \lim_{h \to 0} \sup_{0 \leqslant s \leqslant h} \Delta_h f(x+s)$$

can be expressed as the countable limit of these lower semicontinuous functions, $D^+ f$ is measurable, hence E_γ is measurable. Now consider the shifted function $g(x) = f(x) - \gamma x$. If $\bigcup (a_i, b_i)$ is the set obtainable from g from the rising sun lemma, then $E_\gamma \subset \bigcup (a_i, b_i)$, for if $D^+ f(x) > \gamma$, then there is $h > 0$ arbitrarily small with $\Delta_h f(x) > \gamma$, hence $f(x+h) - f(x) > \gamma h$, hence $g(x+h) > g(x)$. We know that $g(a_k) \leqslant g(b_k)$, so $f(b_k) - f(a_k) \geqslant \gamma(b_k - a_k)$, so

$$|E_\gamma| \leqslant \sum(b_k - a_k) \leqslant \frac{1}{\gamma}\sum f(b_k) - f(a_k) \leqslant \frac{f(b) - f(a)}{\gamma}$$

Thus $|E_\gamma| \to 0$ as $\gamma \downarrow 0$, implying $D^+ f(x) = \infty$ only on a set of measure zero.

Now for two real numbers $r < R$, we will now show

$$E = \{a \leqslant x \leqslant b : D^+ f(x) > R \quad D_- f(x) < r\}$$

is a set of measure zero. Letting r and R range over all rational numbers establishes that $D^+ f(x) \leqslant D_- f(x)$ almost surely. We assume $|E| > 0$ and derive a contradiction. By regularity, we may consider an open subset U in $[a,b]$ containing E such that $|U| < |E|(R/r)$. We can write U as the union of disjoint intervals I_n. For a fixed I_N, apply the rising sun lemma to the function $rx - f(-x)$ on the interval $-I_N$, yielding a union of intervals (a_n, b_n). If we now apply the rising sun lemma to the function $f(x) - Rx$ on (a_n, b_n), we get intervals (a_{nm}, b_{nm}), whose union we denote U_N. Then

$$R(b_{nm} - a_{nm}) \leqslant f(b_{nm}) - f(a_{nm}) \qquad f(b_n) - f(a_n) \leqslant r(b_n - a_n)$$

then, because f is increasing,

$$\begin{aligned} |U_N| = \sum_{nm}(b_{nm} - a_{nm}) &\leqslant \frac{1}{R}\sum_{nm}(f(b_{nm}) - f(a_{nm})) \\ &\leqslant \frac{1}{R}\sum f(b_n) - f(a_n) \leqslant \frac{r}{R}\sum_n (b_n - a_n) \leqslant \frac{r}{R}|I_N| \end{aligned}$$

Now $E \cap I_N$ is contained in U_N, because if $x \in E \cap I_N$, then $D^+ f(x) > R$ and $D_- f(x) < r$, so we can sum in N to conclude that

$$|E| \leqslant \sum \frac{r}{R}|I_N| = \frac{r}{R}|U_N| < |E|$$

a contradiction proving the claim. □

Corollary 17.6. *If f is increasing and continuous, then f' is measurable, non-negative, and*

$$\int_a^b f'(x)\, dx \leqslant f(b) - f(a)$$

Proof. The fact the f' is measurable and non-negative results from the fact that the functions $g_n(x) = \Delta_{1/n} f(x)$ are non-negative and continuous, and $g_n \to f'$ almost surely. We know

$$\begin{aligned}\int_a^b f'(x) &\leqslant \liminf_{n\to\infty} \int_a^b g_n(x) = \liminf_{n\to\infty} n \int_a^b [f(x+1/n) - f(x)] \\ &= \liminf_{n\to\infty} n \left[\int_b^{b+1/n} f(x) - \int_a^{a+1/n} f(x) \right] = f(b) - f(a)\end{aligned}$$

where the last equality follows because f is continuous. $\square$

Even for increasing continuous functions, the inequality in the theorem above need not be an equality, as the next example shows, so we need something stronger to obtain our differentiation theorem.

Example. *The Cantor-Lebesgue function is a continuous increasing function f from $[0,1]$ to itself, with $f(0) = 0$, and $f(1) = 1$, but with $f'(x) = 0$ almost everywhere. This means*

$$\int_0^1 f'(x) = 0 < 1 = f(1) - f(0)$$

so we cannot obtain equality in general. To construct f, consider the Cantor set $C = \bigcap C_k$, where C_k is the disjoint union of 2^k closed intervals. Set $f_0 = 0$, and $f_1(0) = 0$, $f_1(x) = 1/2$ on $[1/3, 2/3]$, $f_1(1) = 1$, and f linear between $[0, 1/3]$ and $[2/3, 1]$. Similarily, set $f_2(0) = 0$, $f_2(x) = 1/4$ on $[1/9, 2/9]$, $f_2(x) = 1/2$ on $[1/3, 2/3]$, $f_2(x) = 3/4$ on $[7/9, 8/9]$, and $f_2(1) = 1$. The functions f_i are increasing and cauchy in the uniform norm, so they converge to a continuous function f called the Cantor function. *f is constant on each interval in the complement of the cantor set, so $f'(x) = 0$ almost everywhere.*

To obtain equality in the integral formula, we require additional conditions on our increasing functions, provided by absolute continuity.

17.1 Absolute Continuity

A function $f : [a, b] \to \mathbf{R}$ is *absolutely continuous* if for every $\varepsilon > 0$, there is $\delta > 0$ such that whenever $(a_1, b_1), \dots, (a_n, b_n)$ are disjoint intervals with

$\sum(b_i - a_i) < \delta$, $\sum|f(b_i) - f(a_i)| < \varepsilon$. Thus the function should be 'essentially constant' over every set of zero measure. It is easy to see from this that absolutely continuous functions must be uniformly continuous, and have bounded variation. Thus f has a decomposition into the difference of two continuous increasing functions, and one can see quite easily that these functions are also absolutely continuous. Most promising to us, if f is a function defined by $f(x) = \int_a^x g(t)\,dt$, where $g \in L^1[a,b]$, then f is absolutely continuous. This shows that absolute continuity is necessary in order to hope for the integral formula

$$\int_a^b f'(x)\,dx = f(b) - f(a)$$

The Cantor function is *not* absolutely continuous, since it is constant except on the Cantor set, and we can cover the Cantor set by intervals with total length $(2/3)^n$ for each n. Thus it is impossible for the Cantor function to satisfy the fundamental theorem of calculus.

Theorem 17.7. *If $g \in L^1(\mathbf{R})$, and*

$$f(x) = \int_a^x g(t)\,dt$$

then f is absolutely continuous.

Proof. Fix $\varepsilon > 0$. We claim that there is δ such that if $|E| < \delta$, then $\int_E |g| < \varepsilon$. Otherwise there are sets E_n with $|E_{n+1}| \leqslant |E_n|/3$ and with $\int_{E_n} |g| \geqslant \varepsilon$. Thus if we define the sets $E'_m = E_m - \bigcup_{n>m} E_n$ then the E'_m and we have $|E_m| \sim |E_m|'$. Since g is integrable, we must have $\sum \int_{E'_n} |g| < \infty$, so we conclude that as $N \to \infty$,

$$\sum_{n \geqslant N} \int_{E'_n} |g| \to 0$$

Yet for any N,

$$\sum_{n \geqslant N} \int_{E'_n} |g| = \int_{E_N} |g| \geqslant \varepsilon$$

which is an impossibility. Thus such a δ exists for every ε, and so if we have disjoint intervals (a_n, b_n) with $\sum(b_n - a_n) < \delta$, then

$$\sum |f(b_n) - f(a_n)| = \sum \left| \int_{a_n}^{b_n} g(t) \right| \leqslant \sum \int_{a_n}^{b_n} |g| = \int_{\bigcup(a_n,b_n)} |g| < \varepsilon$$

which shows the function is absolutely continuous. $\square$

To prove the differentiation theorem, we require a covering estimate not unlike that used to prove the Lebesgue differentiation theorem. We say a collection of balls is a *Vitali covering* of a set E if for every $x \in E$ and every $\eta > 0$, there is a ball B in the cover containing x with radius smaller than η. Thus every point is covered by an arbitrary small ball.

Lemma 17.8. *If E is a set of finite measure, and $\{B_\alpha\}$ is a Vitali covering of E, then there eixsts a disjoint family of cubes $\{B_\beta\}$ in the covering such that*

$$\left| E - \bigcup_\beta B_\beta \right| = 0.$$

Proof. Fix $\delta > 0$. We claim we can find a disjoint family of cubes $B_1, \dots, B_N$ in the Vitali cover such that

$$\left| E - \bigcup_{i=1}^N B_i \right| \leqslant \delta.$$

By inner regularity, pick a compact set $K \subset E$ with $|K| \geqslant |E| - \delta/2$. Then K is covered by finitely many balls of radius less than η in the covering $\{B_\alpha\}$, and the elementary Vitali covering lemma gives a disjoint subcollection of balls $B_1, \dots, B_{n_0}$ with

$$|K| \leqslant \left| \bigcup B_\alpha \right| \leqslant 3^d \sum |B_k|$$

so $\sum |B_k| \geqslant 3^{-d}|K|$. If $\sum |B_k| \geqslant |K| - \delta/2$, we're done. Otherwise, define $E_1 = K - \bigcup \overline{B_k}$. Then

$$|E_1| \geqslant |K| - \sum |\overline{B_k}| = |K| - \sum |B_k| > \delta/2$$

If we pick a compact set $K_1 \subset E_1$ with $|K_1| \geqslant \delta/2$, then if we remove all sets in the Vitali covering which intersect $B_1, \dots, B_{n_0}$, then we still obtain a Vitali covering for K_1, and we can repeat the argument above to find a disjoint collection of open sets $B_1^1, \dots, B_{n_1}^1$ with $\sum |B_k^1| \geqslant 3^{-d}|K_1|$. Then $\sum |B_k| + \sum |B_k^1| \geqslant 2(3^{-d}\delta)$. If $\sum |B_k| + \sum |B_k^1| < |K| - \delta/2$, we repeat the same process, finding a disjoint family for $K_2 \subset E_2$, where $E_2 = K_1 - \bigcup \overline{B_k^1}$. If this process repeats itself k times, then we obtain a family of open sets with total measure greater than or equal to $k(3^{-d}\delta)$. But then if we eventually have $k \geqslant (|E| - \delta)3^d/\delta$, then we obtain the required bound.

We now construct our final cover inductively. Given E, we can find finitely many balls $B_{11}, \dots, B_{1n_1}$ such that

$$\left| E - \bigcup_{i=1}^{n_1} B_i \right| \leqslant 1/2.$$

Set $E_1 = E - \bigcup_{i=1}^{n_1} B_i$. If we remove all balls from the Vitali cover that intersect the balls $B_1, \dots, B_{n_1}$, we still have a Vitali cover of E_1. Inductively, we can then find a disjoint family of balls $B_{k1}, \dots, B_{kn_k}$ which are disjoint from all previously selected balls such that

$$\left| E - \bigcup_{k=1}^{k_0} \bigcup_{i=1}^{n_k} B_{ki} \right| \leqslant 1/2^{k_0}.$$

Taking $k_0 \to \infty$ gives an infinite family of disjoint balls which cover E up to a set of zero Lebesgue measure. □

Theorem 17.9. *If $f : [a,b] \to \mathbf{R}$ is absolutely continuous, then f' exists almost everywhere, and if $f'(x) = 0$ almost surely, then f is constant.*

Proof. It suffices to prove that $f(a) = f(b)$, since we can then apply the theorem on any subinterval. Let $E = \{x \in (a,b) : f'(x) = 0\}$. Then $|E| = b - a$. Fix $\varepsilon > 0$. Since for each $x \in E$, we have

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = 0$$

This implies that the family of intervals (x,y) such that the inequality $|f(y) - f(x)| \leqslant \varepsilon(y - x)$ holds forms a Vitali covering of E, and we may therefore select a family of disjoint intervals $I_i = (x_i, y_i)$ with

$$\sum |I_i| \geqslant |E| - \delta = (b-a) - \delta$$

But $|f(y_i) - f(x_i)| \leqslant \varepsilon(y_i - x_i)$, so we conclude

$$\sum |f(y_i) - f(x_i)| \leqslant \varepsilon(b-a)$$

The complement of I_i is a union of intervals $J_i = (x_i', y_i')$ of total length $\leqslant \delta$. Applying the absolute continuity of f, we conclude

$$\sum |f(y_i') - f(x_i')| \leqslant \varepsilon$$

so applying the triangle inequality,

$$|f(b) - f(a)| \leqslant \sum |f(y_i') - f(x_i')| + \sum |f(y_i) - f(x_i)| \leqslant \varepsilon(b - a + 1)$$

We can then let $\varepsilon \to 0$ to obtain equality. □

Theorem 17.10. *Suppose f is absolutely continuous on $[a,b]$. Then f' exists almost every and is integrable, and*

$$f(b) - f(a) = \int_a^b f'(y)\, dy$$

so the fundamental theorem of calculus holds everywhere. Conversely, if $f \in L^1[a,b]$, then there is an absolutely continuous function g with $g' = f$ almost everywhere.

Proof. Since f is absolutely continuous, we can write f as the difference of two continuous increasing functions on $[a,b]$, and this easily implies f is differentiable almost everywhere and is integrable on $[a,b]$. If $g(x) = \int_a^x f'(x)$, then g is absolutely continuous, hence $g - f$ is also absolutely continuous. But we know that $(g-f)' = g' - f' = 0$ almost everywhere, so the last theorem implies that g differs from f by a constant. Since $g(a) = 0$, $g(x) = f(x) - f(a)$. The converse was proved exactly in our understanding of differentiating integrals. □

We now dwell slightly longer on the properties of absolutely continuous functions, which enables us to generalize other properties of integrals found in the calculus. We begin by noting that it is easy to verify that if f and g are absolutely continuous functions, then fg is also absolutely continuous. We know f', g', and $(fg)'$ exist almost everywhere. But when all three exist simultaneously, the product rule gives $(fg)' = f'g + fg'$. The absolute continuity implies that

$$\int_a^b f'g + fg' = \int_a^b (fg)' = f(b)g(b) - f(a)g(a)$$

Thus one can integrate a pair of absolutely continuous functions by parts. Next, we shall show that monotone absolutely continuous functions are precisely those we can use to change variables. One important thing to note is that even if f is a continuous function, and g is measurable, $g \circ f$ need not be measurable. The easy reason to see this is that the inverse image of every open set in g is measurable, so in order to guarantee $g \circ f$ is measurable we need the inverse image of every measurable set under f be measurable.

Example. *Consider the function $f(x) = \int_0^x \chi_E(x)\, dx$, where E is a thick Cantor set. Then f is absolutely continuous, strictly increasing on $[0,1]$, and maps E to a set of measure zero. This is because $E = \lim E_n$, where E_n is a family of intervals with $|E_n| \downarrow |E|$. Then $f(E_n)$ has total length $|E_n - E|$, so as $n \to \infty$, we see $\lim f(E_n) = f(E)$ has measure zero. This means that $f(X)$ is measurable for any subset X of E, and in particular, if X is non-measurable, then $f^{-1}(f(X))$ cannot be measurable, even though $f(X)$ is measurable. Note that f is strictly increasing even though it's derivatives vanish on a set of positive measure.*

The next lemmas will show that even though $g \circ f$ may not be Lebesgue measurable when f is absolutely continuous and g is Lebesgue measurable, this does not really bother us too much when changing variables.

Lemma 17.11. *If f is absolutely continuous, then it maps sets of measure zero to sets of measure zero.*

Proof. Let E be a set of measure zero. Then for each $\delta > 0$, E is coverable by a family of open intervals with total length δ. But if δ is taken small enough, this means that $f(E)$ is coverable by a family of open intervals with total length bounded by ε, for any ε. $\square$

This property of absolutely continuous functions is independant of the properties of the Euclidean domain as it's domain, and is used in the generalization of absolute continuity to more general domains, or even to measures. If f is absolutely continuous, then the image of every interval is an interval, and since $f(\bigcup K_n) = \bigcup f(K_n)$, this implies that the image of a F_σ set is measurable. But since every measurable set of $\mathbf{R}$ differs from a F_σ set by a set of measure zero, the image of every Lebesgue measurable set is Lebesgue measurable. The reverse is almost true.

Lemma 17.12. *If f is absolutely continuous, and E measurable, then the set*

$$f^{-1}(E) \cap \{x : f'(x) > 0\}$$

is measurable.

Proof. If E is an open set, then

$$|E| = \int_{f^{-1}(E)} f'(x)\, dx$$

It suffices to prove this when E is an interval, and then this is just the theorem of differentiation for absolutely continuous functions. But then applying the dominated convergence theorem shows that this equation remains true if E is an G_δ set. Furthermore, this means the theorem is true if E is a closed set, and so by applying the monotone convergence theorem, the theorem is true if E is an F_σ set. But if E is an arbitrary measurable set, then for every ε there are F_σ and G_δ sets $K \subset E \subset U$ with $|U - K| = 0$. But

$$\alpha |f^{-1}(U-K) \cap \{f' \geqslant \alpha\}| \leqslant \int_{f^{-1}(U-K)} f'(x)\, dx = |U - K| = 0$$

Thus $f^{-1}(U-K) \cap \{f' \geqslant \alpha\}$ is a set of measure zero, and so in particular by completeness, every set contained in this set is measurable, in particular $f^{-1}(U-E) \cap \{f' \geqslant \alpha\}$ is measurable. But now this means

$$\{f' \geqslant \alpha\} - f^{-1}(U-E) \cap \{f' \geqslant \alpha\} = f^{-1}(E) \cap \{f' \geqslant \alpha\}$$

is measurable. Taking $\alpha \downarrow 0$ completes the proof. $\square$

Because of this, even though $g \circ f$ is not necessarily measurable, $(g \circ f) f'$ is always measurable if f is absolutely continuous. Thus the expression $\int (g \circ f) f'$ makes sense, and thus we can always interpret the change of variables formula.

Theorem 17.13. *If f is absolutely continuous, and g is integrable, then*

$$\int g(f(x)) f'(x)\, dx = \int g(y)\, dy$$

Proof. Using the notation in the last proof, if E is measurable, then

$$|K| = \int_{f^{-1}(K)} f'(x)\, dx \leqslant \int_{f^{-1}(E)} f'(x)\, dx \leqslant \int_{f^{-1}(U)} f'(x)\, dx = |U|$$

and $|U| = |K| = |E|$, so that for any measurable set E,

$$|E| = \int_{f^{-1}(E)} f'(x)\, dx$$

This imples the theorem we need to prove is true whenever g is the characteristic function of any measurable set. But then by linearity, it is true for any simple function. By monotone convergence, it is then true for any non-negative function, and then by partitioning g into the sum of simple functions, we obtain the theorem in general. □

17.2 Differentiability of Jump Functions

We now consider the differentiability of not necessarily continuous monotonic functions. Set f to be an increasing function on $[a,b]$, which we may assume to be bounded. Then the left and right limits of f exist at every point, which we will denote by $f(x-)$ and $f(x+)$. Of course, we have $f(x-) \leqslant f(x) \leqslant f(x+)$. If there is a discontinuity, this means we are forced to have a 'jump discontinuity' where f skips an interval. This implies that f can only have countably many such discontinuities, because a family of disjoint intervals on $\mathbf{R}$ is at most countable. Now define the jump function $\Delta f(x) = f(x^+) - f(x-)$, with $\theta(x) \in [0,1]$ defined such that $f(x_n) = f(x_n-) + \theta(x)\Delta f(x)$. If we define the functions

$$j_y(x) = \begin{cases} 0 & : x < y \\ \theta(y) & : x = y \\ 1 & x > y \end{cases}$$

then we can define the *jump function* associated with f by

$$J(x) = \sum_x \Delta f(x) j_n(x)$$

Since f is bounded on $[a,b]$, we make the final observation that

$$\sum_{x\in[a,b]} \Delta f(x) \leqslant f(b) - f(a) < \infty$$

so the series defining J converges absolutely and uniformly.

Lemma 17.14. *If f is increasing and bounded on $[a,b]$, then J is discontinuous precisely at the values x with $\Delta f(x) \neq 0$ with $\Delta J(x) = \Delta f(x)$. The function $f - J$ is continuous and increasing.*

Proof. If x is a continuity point of f, then j_y is continuous at x, and hence, because $\sum_y \Delta f(y) j_y(x) \to J(x)$ uniformly, so we conclude that J is continuous at x. On the other hand, for each y, $j_y(y-) = 0$ and $j_y(y+) = 1$, and if we label the points of discontinuity of f by $x_1, x_2, \dots$, then

$$J(x) = \sum_{i=1}^{k} \Delta f(x_i) j_{x_i} + \sum_{i=k+1}^{\infty} \Delta f(x_i) j_{x_i}$$

The right hand partial sums are continuous at x_k, whereas the left hand sum has a jump discontinuity of the same order as f at x_k, we conclude J also has this discontinuity. But this means that

$$(f-J)(x_k+) - (f-J)(x_k-) = 0$$

so $f - J$ is continuous at every point. $f - J$ is increasing because of the inequality

$$J(y) - J(x) \leqslant \sum_{x < x_n \leqslant y} \alpha_n \leqslant f(y) - f(x)$$

which follows because J is just the sum of jump discontinuities, and the right hand side because f can decrease and increase outside of the jump discontinuities. $\square$

Since $f - J$ is continuous and increasing, it is differentiable almost everywhere. It therefore remains to analyze the differentiability of the jump function J.

Theorem 17.15. *J' exists and vanishes almost everywhere.*

Proof. Fix $\varepsilon > 0$, and consider

$$E = \left\{ x \in [a,b] : \limsup_{h \to 0} \frac{J(x+h) - J(x)}{h} > \varepsilon \right\}$$

Then E is measurable, because we can take the lim sup over rational numbers because J is increasing. We want to show it has measure zero. Suppose $\delta = |E|$. Consider $\eta > 0$ to be chosen later, and find n such that $\sum_{k=n}^{\infty} \alpha_k < \eta$. Write

$$J_0(x) = \sum_{n > N} \alpha_n j_n$$

Then $J_0(b) - J_0(a) < \eta$. Now E differs from the set

$$E' = \left\{ x \in [a,b] : \limsup_{h \to 0} \frac{J_0(x+h) - J_0(x)}{h} > \varepsilon \right\}$$

by finitely many points. Using inner regularity, find a compact set $K \subset E'$ with $|K| \geqslant \delta/2$. For each $x \in K$, we can find intervals (α_x, β_x) upon which $J_0(\beta_x) - J_0(\alpha_x) \geqslant \varepsilon|\beta_x - \alpha_x|$. But applying the elementary Vitali covering lemma, we can find a disjoint family of such intervals with $\sum(\beta_{x_i} - \alpha_{x_i}) \geqslant |K|/3 \geqslant \delta/6$. But now we find

$$J_0(b) - J_0(a) \geqslant \sum J_0(\beta_{x_i}) - J_0(\alpha_{x_i}) \geqslant \varepsilon\delta/6$$

This means $\delta \leqslant 6\eta/\varepsilon$, and by letting $\eta \to 0$, we can conclude $\delta = 0$. $\square$

This concludes our argument that *every* function of bounded variation has a derivative almost everywhere, because every such function can be uniquely written (up to a shift in the range of the functions) as the sum of a continuous function and a jump function. If f is a function with bounded variation, then the function

$$F(x) = \int_0^x f'(x)$$

is absolutely continuous, and $f - F$ is a continuous function with derivative zero almost everywhere. The fact that this decomposition is unique up to a shift as well (which can easily be seen in the case of an increasing function, from which the general case follows) leads us to refer to this as the *Lebesgue decomposition* of a function of bounded variation on the real line.

17.3 Rectifiable Curves

We now consider the validity of the length formula

$$L = \int_a^b (x'(t)^2 + y'(t)^2)^{1/2} \, dt$$

where L is the length of the curve parameterized by (x, y) on $[a, b]$. We cannot always expect this formula to hold, because if x and y are both the Cantor devil staircase function, then the formula above gives a length of zero, whereas we know the curve traces a line between 0 and 1, hence has length at least $\sqrt{2}$.

Theorem 17.16. *If a curve is parameterized by absolutely continuous functions x and y on $[a, b]$, then the curve is rectifiable, and has length*

$$\int_a^b (x'(t) + y'(t))^{1/2} \, dt$$

Proof. This proof can be reworded as saying if f is complex-valued and absolutely continuous, then it's total variation can be expressed as

$$V(f, a, b) = \int_a^b |f'(t)| \, dt$$

If $P = \{a \leqslant t_1 < \dots < t_n \leqslant b\}$ is a partition, then

$$\sum |f(t_{n+1}) - f(t_n)| = \sum \left| \int_{t_n}^{t_{n+1}} f'(t) \, dt \right| \leqslant \sum \int_{t_n}^{t_{n+1}} |f'(t)| \, dt \leqslant \int_a^b |f'(t)| \, dt$$

so $V(f, a, b) \leqslant \int_a^b |f'(t)| \, dt$. To prove the converse inequality, fix $\varepsilon > 0$, and find a step function g with $f' = g + h$, with $\|h\|_1 \leqslant \varepsilon$. If $G(x) = \int_a^x g(t) \, dt$ and $H(x) = \int_a^x h(t) \, dt$, then $F = G + H$, and $V(f, a, b) \geqslant V(G, a, b) - V(H, a, b) \geqslant V(G, a, b) - \varepsilon$, and if we partition $[a, b]$ into $a = t_0 < \dots < t_N$, where G is constant on each (t_n, t_{n+1}), then

$$\begin{aligned} V(G, a, b) &\geqslant \sum |G(t_n) - G(t_{n-1})| = \sum \left| \int_{t_{n-1}}^{t_n} g(t) \, dt \right| \\ &= \sum \int_{t_{n-1}}^{t_n} |g(t)| \, dt = \int_a^b |g(t)| \, dt \geqslant \|f'\|_1 - \varepsilon \end{aligned}$$

Letting $\varepsilon \to 0$ now gives the result. □

It is interesting to note that any rectifiable curve has a special *parameterization by arclength*, i.e. a parameterization $(x(t), y(t))$ such that if L is the length function associated to the parameterization, then $L(A,B) = B - A$. This is obtainable by inverting the length function.

Theorem 17.17. *If $z = (x,y)$ is a parameterization of a rectifiable curve by arclength, then x and y are absolutely continuous, and $|z'| = 1$ almost everywhere.*

Proof. For any $s < t$,

$$t - s = L(s,t) = V(f,s,t) \geqslant |z(t) - z(u)|$$

so it follows immediately that $|z|$ is an absolutely continuous function, and $|z'| \leqslant 1$ almost surely. But now we know that

$$\int_a^b |z'(t)| = b - a$$

and this equality can now only hold if $|z'(t)| = 1$ almost surely. □

17.4 Bounded Variation in Higher Dimensions

Since the higher dimensional Euclidean domains do not have an ordering, it is impossible to define their length by partitioning their domain, and the meaning of a jump discontinuity is no longer clear. However, there are properties equivalent to having bounded variation which are more extendable to higher dimensions.

Theorem 17.18. *The following properties of $f : \mathbf{R} \to \mathbf{R}$ are equivalent, for some fixed finite constant A.*

- *f can be modified on a set of measure zero so that it has bounded variation not exceeding A.*
- $\int |f(x+h) - f(x)| \leqslant A|h|$ *for all* $h \in \mathbf{R}$.
- *For any C^1 function φ with compact support,* $\left|\int f(x)\varphi'(x)\right| \leqslant A\|\varphi\|_\infty$.

Proof. If $V(f) = A$, where $A < \infty$, then we can write $f = f^+ - f^-$, where f^+ and f^- are both increasing functions, and with $V(f) = V(f^+) + V(f^-)$. It then follows that $|f(x+h) - f(x)| \leqslant (f^+(x+h) - f^+(x)) + |f^-(x+h) - f^-(x)|$, so it suffices to prove the second property by assuming f is increasing. But then by the monotone convergence theorem, assuming $h > 0$ without loss of generality,

$$\int |f(x+h) - f(x)| = \lim_{y \to \infty} \int_{-y}^{y} f(x+h) - f(x) = \lim_{y \to \infty} \int_{y}^{y+h} f(x) - \int_{-y-h}^{-y} f(x)$$

The first term of the limit converges to $hV(f)$, and the second to zero, completing the first part of the theorem. Now assuming the second point, we prove the third point. Then using the second point, we find

$$\begin{aligned} \left| \int f(x)\varphi'(x) \right| &= \left| \lim_{h \to 0} \int f(x) \frac{\varphi(x+h) - \varphi(x)}{h} \right| \\ &= \left| \lim_{h \to 0} \int \frac{f(x-h) - f(x)}{h} \varphi(x) \right| \leqslant A \|\varphi\|_\infty \end{aligned}$$

Finally, we consider the third point being true. The set of all partitions with rational points is countable. Suppose that for each rational $P = \{t_0 < \cdots < t_N\}$ there is a set E_P of measure zero for each rational partition P such that

$$\sum_{n=1}^{N} \sup_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E_P}} f(x) - \inf_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E_P}} f(x) \leqslant A$$

Then the union of E_P over all rational P has measure zero. We can modify f on E_P by setting $f(x) = \liminf_{y \to 0} f(x+y)$, and then $V(f, P) \leqslant A$ for all rational partitions P. If Q is now any partition, we can find a rational partition P with $V(f, P) \geqslant V(f, Q) - \varepsilon$, and so $V(f, P) \leqslant A - \varepsilon$. Taking $\varepsilon \to 0$ completes the argument. Thus if f cannot be modified to have finite variation A, there exists a rational partition P such that for any set E of measure zero,

$$\sum_{n=1}^{N} \sup_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E}} f(x) - \inf_{\substack{x \in [t_{n-1}, t_n] \\ x \notin E}} f(x) > A$$

Thus for any ε, there exists $E_n^+, E_n^- \subset [t_{n-1}, t_n]$ of positive measure such that

$$\sum_{n=1}^{N} \inf_{x \in E_n^+} f(x) - \sup_{x \in E_n^-} f(x) > A$$

If we consider the polygonal function ϕ which □

17.5 Minkowski Content

Given a set $K \in \mathbf{R}^n$, we let K^δ denote the open set consisting of points x with $d(x,K) < \delta$. The m dimensional *Minkowski content* of K is defined to be

$$\lim_{\delta \to 0} \frac{1}{\alpha(n-m)} \frac{|K^\delta|}{\delta^m}$$

where $\alpha(d)$ is the volume of the unit ball in d dimensions. When this limit exists, we denote it by $M^m(K)$. In this section, we mainly discuss the one dimensional Minkowski content in two dimensions, i.e. the values of

$$\lim_{\delta \to 0} \frac{|K^\delta|}{2\delta^m}$$

and it's relation the length of curves. Since we now only care about the one dimensional Minkowski content, we let $M(K)$ denote the one dimension Minkowski content.

Lemma 17.19. *If $\Gamma = \{z(t) : a \leqslant t \leqslant b\}$ is a curve, and Δ is the distance between the endpoints of the curve, then $|\Gamma^\delta| \geqslant 2\delta\Delta$.*

Proof. By rotating, we may assume that both endpoints of the curve lie on the x axis, so $z(a) = (A,0)$, $z(b) = (B,0)$ with $A < B$, so $\Delta = B - A$. If $\Delta = 0$, the theorem is obvious. Otherwise, for each point $x \in [A,B]$ there is $t(x)$ such that if $z_1(t(x)) = x$, and so Γ^δ contains $x \times [z_2(t(x)) - \delta, z_2(t(x)) + \delta]$, which has length 2δ. Thus by Fubini's theorem,

$$|\Gamma^\delta| = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi_{\Gamma^\delta}(x,y)\, dx\, dy \geqslant \int_A^B 2\delta = 2\delta\Delta$$

so the theorem is proved. □

Theorem 17.20. *If* $\Gamma = \{z(t) : a \leqslant t \leqslant b\}$ *is a quasi-simple curve (simple except at finitely many points), then the Minkowski content of* Γ *exists if and only if* Γ *is rectifiable, and in this case* $M^1(\Gamma)$ *is the length of the curve* L.

Proof. To prove the theorem, we consider the upper and lower Minkowski contents

$$M^*(\Gamma) = \limsup_{\delta \to 0} \frac{|\Gamma|^\delta}{\alpha(n-1)\delta} \quad M_*(\Gamma) = \liminf_{\delta \to 0} \frac{|\Gamma|^\delta}{\alpha(n-1)\delta}$$

First, we prove that $M^*(\Gamma) \leqslant L$. Consider a partition P of $[a,b]$, and let L_P be the length of the polygonal approximation to the curve. By refining the partition, we may assume that Γ is simple, with the repeated points at the boundaries of the intervals. For each interval I_n in the partition, we select a closed subinterval $J_n = [t_n, u_n]$ such that Γ is simple on $\bigcup J_n$, and

$$\sum |z(u_n) - z(t_n)| \geqslant L_P - \varepsilon$$

Since the intervals J_n are disjoint, for suitably small δ the sets J_n^δ are disjoint. Applying the previous lemma, we conclude that

$$|\Gamma^\delta| \geqslant \sum |J_n^\delta| \geqslant 2\delta \sum |z(u_n) - z(t_n)| = 2\delta(L_p - \varepsilon)$$

First, by letting $\varepsilon \to 0$ and then $\delta \to 0$, we conclude that $M_*(\Gamma) \geqslant \lim_P L_P$. In particular, this shows that if Γ has Minkowski content one, then the curve is rectifiable. Conversely, we consider the functions

$$F_n(s) = \sup_{0 < |h| < 1/n} \left| \frac{z(s+h) - z(s)}{h} - z'(s) \right|$$

Because z is continuous, this supremum can be considered over a countable, dense subset, and so each F_n is measurable. Since $F_n(s) \to 0$ for almost every s, we can apply Egorov's theorem to show that this limit is uniform except on a singular set E with $|E| < \varepsilon$, so that for some large N, for $s \notin E$ and $|h| < 1/N$, $|z(s+h) - z(s) - hz'(s)| < \varepsilon h$. We now split the interval $[a,b]$ into consecutive intervals $I_1, \dots, I_{M+1}$, with each interval but I_{M+1} having length $1/N$. We let Γ_n denote the section of the curve travelled along the interval I_n. Thus $|\Gamma^\delta| \leqslant \sum |\Gamma_n^\delta|$. If an interval I_n contains an element of E^c, we say I_n is a 'good' interval. Then we can pick an element $x_n \in I_n$ for which for any $x \in I_n$,

$$|z(x) - z(x_n) - (x - x_n)z'(x_n)| < \varepsilon|x - x_n| < \varepsilon/N$$

Thus Γ_n is covered by a ε/N thickening of a length $1/N$ line J_n in $\mathbf{R}^2$ through $z(x_n)$ with slope $z'(x_n)$. Thus if $\varepsilon \leqslant 1$, we conclude

$$|\Gamma_n^\delta| \leqslant J_n^{\varepsilon/N+\delta} \leqslant (1/N + 2\varepsilon/N + 2\delta)(2\varepsilon/N + 2\delta)$$
$$\leqslant 2\delta/N + O(\delta\varepsilon/N + \delta^2 + \varepsilon/N^2)$$

Since $M \leqslant NL$, if we take the sum of $|\Gamma_n^\delta|$ over all 'good' intervals we obtain an upper bound of

$$NL\left(2\delta/N + O(\delta\varepsilon/N + \delta^2 + \varepsilon/N^2)\right) = 2\delta L + O(\delta\varepsilon + \delta^2 N + \varepsilon/N)$$

On the other hand, if I_n is contained within E, or if $n = M + 1$, we say I_n is a bad interval. Since E has total measure bounded by ε, there can be at most $\varepsilon N + 1$ bad intervals. On these intervals we use the crude estimate $|z(t) - z(u)| \leqslant |t - u|$ (true because z is an arclength parameterization) to show Γ_n is contained in a rectangle with sidelengths $1/N$, so we obtain that $|\Gamma_n^\delta| \leqslant (1/N + 2\delta)^2 = O(1/N^2 + \delta^2)$. Thus the sum of $|\Gamma_n^\delta|$ over the 'bad intervals' is bounded by

$$O(\varepsilon/N + 1/N^2 + \varepsilon N\delta^2 + \delta^2)$$

In particular, the sum of the two bounds gives

$$|\Gamma^\delta| \leqslant 2\delta L + O(\delta\varepsilon + \delta^2 N + \varepsilon/N + 1/N^2)$$

Or

$$\frac{|\Gamma^\delta|}{2\delta} \leqslant L + O(\varepsilon + \delta N + \varepsilon/N + 1/N^2)$$

If we choose $N \geqslant 1/\delta$, we get that

$$\frac{|\Gamma^\delta|}{2\delta} \leqslant L + O(\varepsilon + \delta N + \delta\varepsilon) = L + O(\varepsilon + \delta N)$$

Letting $\delta \downarrow 0$, we conclude that $M^*(\Gamma) \leqslant L + O(\varepsilon)$, and we can then let $\varepsilon \downarrow 0$ to conclude $M^*(\Gamma) \leqslant L$. This completes the proof that if Γ is rectifiable, then Γ has one dimensional Minkowski content, and $M(\Gamma) = L$. $\square$

If Γ is rectifiable, it is parameterizable by a Lipschitz map (the arclength parameterization). If we instead consider a curve parameterizable by a map z which is Lipschitz of order α, which may no longer be absolutely continuous, but still has a decay very similar to the Minkowski dimension decay.

Theorem 17.21. *If z is a planar curve which is Lipschitz of order $\alpha > 1/2$, then it's trace Γ satisfies $|\Gamma^\delta| = O(\delta^{2-1/\alpha})$.*

Proof. Since $|z(t) - z(s)| \leqslant |t-s|^\alpha$, we can cover z by $O(N)$ radius $1/N^\alpha$ balls, so $|\Gamma| \lesssim N^{1-2\alpha}$, and so $|\Gamma^\delta| \lesssim N(1/N^\alpha + \delta)^2$. Setting $N = \delta^{-1/\alpha} + O(1)$ gives $|\Gamma^\delta| \lesssim \delta^{2-\alpha-1/\alpha}$. □

17.6 The Isoperimetric Inequality

We now use our Minkowski content techniques to prove the isoperimetric inequality, which asks us to find the region in the plane with largest area whose boundary has a bounded length L. We suppose Ω is a bounded region of the plane, whose boundary $\partial\Omega$ is a rectifiable curve with length L. In particular, we shall find the region with the largest area whose boundary has a fixed length are balls. A key inequality used in the proof is the Brun Minkowski inequality, which lowers bounds the measure of $A+B$ in terms of A and B. If we hope for an estimate $|A+B|^\alpha \gtrsim |A|^\alpha + |B|^\alpha$, then taking $B = \alpha A$, where A is convex and, for which $A + \alpha A = (1+\alpha)A$, we find $(1+\alpha)^{d\alpha} \gtrsim (1 + \alpha^{d\alpha})$. Thus $\alpha \geqslant 1/d$.

Lemma 17.22. *If A, B, and $A+B$ are measurable, $|A+B|^{1/d} \geqslant |A|^{1/d} + |B|^{1/d}$.*

Proof. Suppose first that A and B are rectangles with side lengths x_n and y_n. Then the Minkowski inequality becomes

$$\left(\prod (x_n + y_n)\right)^{1/d} \geqslant \left(\prod x_n\right)^{1/d} + \left(\prod y_n\right)^{1/d}$$

Replacing x_n with $\lambda_n x_n$ and y_n with $\lambda_n y_n$, we find that we may assume $x_n + y_n = 1$, and so we must prove that for any $x_n \leqslant 1$,

$$\left(\prod x_n\right)^{1/d} + \left(\prod (1-x_n)\right)^{1/d} \leqslant 1$$

But this inequality is an immediate consequence of the arithmetic geometric mean inequality. Thus the case is proved. Next, we suppose A and B are unions of disjoint closed rectangles, and we prove the inequality by induction on the number of rectangles. Without loss of generality, by symmetry in A and B, we may assume that A has at least two rectangles R_1 and R_2. Since the inequality is translation invariant separately in A and B, and R_1 and R_2 is disjoint, hence separated by a coordinate axis, we may assume there exists an index j such that every element x of R_1 has $x_1 < 0$ and every element x of R_2 has $x_1 > 0$. Let $A^+ = A \cap \{x_j \leqslant 0\}$ and $A^- = A \cap \{x_j \geqslant 0\}$. Next, we translate B such that if $B^\pm$ are defined similarily, then

$$\frac{|B^\pm|}{|B|} = \frac{|A^\pm|}{|A|}$$

Note that $A+B$ contains the union of $A^+ + B^+$ and $A^- + B^-$, and this union is disjoint. Thus by induction,

$$\begin{aligned}
|A+B| &\geqslant |A^+ + B^+| + |A^- + B^-| \\
&\geqslant (|A^+|^{1/d} + |B^+|^{1/d})^d + (|A^-|^{1/d} + |B^-|^{1/d})^d \\
&= |A^+| \left(1 + \left(\frac{|B|^+}{|A|^+}\right)^{1/d}\right)^d + |A^-| \left(1 + \left(\frac{|B|^-}{|A|^-}\right)\right)^d \\
&= (|A|^{1/d} + |B|^{1/d})^d
\end{aligned}$$

Thus the proof is completed for unions of rectangles. The proof then passes to open sets by approximating open sets by closed rectangles contained within. Then we can pass to where A and B are compact sets, since then $A+B$ is compact, and so if we consider the open thickenings A^ε, B^ε, and $(A+B)^\varepsilon$, then

$$|A| = \lim |A^\varepsilon| \quad |B| = \lim |B^\varepsilon| \quad |A+B| = \lim |(A+B)^\varepsilon|$$

and $(A+B)^\varepsilon \subset A^\varepsilon + B^\varepsilon \subset (A+B)^{2\varepsilon}$. Finally, we can use inner regularity to obtain the theorem in full. □

Theorem 17.23. *For any region* Ω, $4\pi|\Omega| \leqslant L^2$.

Proof. For $\delta > 0$, consider

$$\Omega_+(\delta) = \{x : d(x,\Omega) < \delta\} \quad \Omega_-(\delta) = \{x : d(x,\Omega^c) \geqslant \delta\}$$

Then we have a disjoint union $\Omega_+(\delta) = \Omega_-(\delta) + \Gamma^\delta$, where Γ is the boundary curve of Ω. Furthermore, $\Omega_+(\delta)$ contains $\Omega + B_\delta$, and Ω contains $\Omega_-(\delta) + B_\delta$. Applying the Brun Minkowski inequality, we conclude

$$|\Omega_+(\delta)| \geqslant (|\Omega|^{1/2} + \pi^{1/2}\delta)^2 \geqslant |\Omega| + 2\pi^{1/2}\delta|\Omega|^{1/2}$$

$$|\Omega| \geqslant (|\Omega_-(\delta)|^{1/2} + \pi^{1/2}\delta)^2 \geqslant |\Omega_-(\delta)| + 2\pi^{1/2}\delta|\Omega_-(\delta)|^{1/2}$$

But

$$|\Gamma^\delta| = |\Omega_+(\delta)| - |\Omega_-(\delta)| \geqslant 2\pi^{1/2}\delta\left(|\Omega|^{1/2} + |\Omega_-(\delta)|^{1/2}\right)$$

Dividing by 2δ and letting $\delta \to 0$, we conclude $L \geqslant 2\pi^{1/2}|\Omega|^{1/2}$. This is precisely the inequality we need. □

Using some Fourier analysis, we can prove that the only smooth curves which make this inequality tight are circles. Indeed, if a closed C^1 curve $\Gamma = \{z(t) : a \leqslant t \leqslant b\}$ is given, then Green's theorem implies the area of its interior is given by

$$\frac{1}{2}\left|\int_\Gamma x\,dy - y\,dx\right| = \frac{1}{2}\left|\int_a^b x(t)y'(t) - y(t)x'(t)\right|$$

We then take a Fourier series in x and y.

Theorem 17.24. *The only curves Γ with rectifiable boundary such that $A = \pi(L/2)^2$ are circles.*

Proof. By normalizing, we may assume z is an arcline parameterization, and Γ has length 2π, so $z : [0, 2\pi] \to \mathbf{R}^2$, and z is absolutely continuous. If $x(t) \sim \sum a_n e^{nit}$ and $y(t) \sim \sum b_n e^{int}$, then $x'(t) \sim \sum ina_n e^{nit}$ and $y(t) \sim \sum inb_n e^{nit}$. Parseval's equality implies

$$\int_0^{2\pi} x(t)y'(t) - y(t)x'(t) = 2\pi i \sum n(b_n\overline{a_n} - a_n\overline{b_n})$$

Thus the area of the curve is precisely

$$\pi\left|\sum n(b_n\overline{a_n} - a_n\overline{b_n})\right| \leqslant \pi\sum 2n|b_n a_n| \leqslant \pi\sum |n|(|a_n|^2 + |b_n|^2)$$

On the other hand, the length constraint implies that, since $|z'(t)| = 1$,

$$1 = \frac{1}{2\pi}\int_0^{2\pi} x'(t)^2 + y'(t)^2 = \sum |n|^2(|a_n|^2 + |b_n|^2)$$

If $A = \pi$, then

$$\sum |n|(|a_n|^2 + |b_n|^2) \geqslant 1 = \sum |n|^2(|a_n|^2 + |b_n|^2)$$

This means we cannot have $|n| < |n|^2$ whenever a_n or b_n is nonzero. Thus the Fourier support of x and y is precisely $\{-1, 0, 1\}$. Since x is real valued, $a_1 = \overline{a_{-1}} = a$, $b_1 = \overline{b_{-1}}$. We thus have $2(|a_1|^2 + |b_1|^2) = 1$, and since we must have a a scalar multiple of b so the Cauchy Schwarz inequality application becomes an equality, we must have $|a_1| = |b_1| = 1/2$. If $a_1 = e^{i\alpha}/2$ and $b_1 = e^{i\beta}/2$, the fact that $1 = 2|a_1\overline{b_1} - \overline{a_1}b_1|$ implies $|\sin(\alpha - \beta)| = 1$, hence $\alpha - \beta = k\pi/2$, where k is an odd integer. Thus $x(s) = \cos(\alpha + s)$, and $y(s) = \cos(\beta + s)$, which parameterizes a circle. $\square$

Chapter 18

Singular Integral Operators

Let us now consider some kernel operators

$$Tf(y) = \int K(x,y) f(x)\, dx$$

where $K(x,y)$ is singular for $x = y$. The prototypical example is the Hilbert transform on $\mathbf{R}$, i.e.

$$Hf(y) = \int \frac{f(x-y)}{y}\, dy.$$

One cannot even interpret the right hand side in the Lebesgue sense because the function $1/y$ is not Lebesgue integrable. We can proceed

Chapter 19

Fourier Multiplier Operators

Our aim in this chapter is to study the boundedness of *Fourier multiplier operators*. Given a function $m : \mathbf{R}^d \to \mathbf{C}$, known as a *symbol*, we want to associate a multiplier operator T, sometimes denoted $m(D)$, which when applied to a function $f : \mathbf{R}^d \to \mathbf{C}$ should be formally given by the equation

$$Tf(x) = \int_{\mathbf{R}^d} m(\xi)\widehat{f}(\xi)e^{2\pi i\xi\cdot x}\, d\xi.$$

In maximum generality, for any tempered distribution m on $\mathbf{R}^d$ we can define T as a continuous operator from $\mathcal{S}(\mathbf{R}^d)$ to itself. But often times we will consider much more regular symbols m, i.e. those which are locally integrable functions, and our goal will be to obtain stronger continuity statements for the associated multiplier operators. If K is the tempered distribution which is the Fourier transform of m, then $Tf = K * f$. Thus Fourier multiplier operators are precisely the same as the class of convolution operators formed by tempered distributions. In any case, the map $m \mapsto m(D)$ gives an injective *algebra homomorphism* from the family of all tempered distributions to the family of continuous operators on $\mathcal{S}(\mathbf{R}^d)$ (another injective homomorphism is obtained by considering the multiplier operators $m \mapsto m(X)$, where $m(X)$ is the multiplier operator $m(X)f = m \cdot f$, the familty of *spatial* multiplier operators, which have a much simpler theory). The main goal, of course, is to determine what properties of the symbol or it's Fourier transform imply boundedness properties of the operator T.

Remark. In engineering these operators are known as *filters*, and occur in a variety of contexts. Due to the presence of error the regularity of these operators are of utmost importance. The function m is known as the *system-transfer function*, *optical-transfer function*, or *frequency response*, depending on the context, and the function K is known as the *point-spread function*.

Example. *Over* $\mathbf{R}$, *consider a rough cutoff* $\mathbf{I}_{[-1,1]}$. *Then we calculate explicitly that*

$$K(x) = \int_{-1}^{1} e^{2\pi i \xi \cdot x} = \frac{\sin(2\pi x)}{\pi x}.$$

Thus convolution by K acts by cutting off higher frequency parts of the function. In engineering this operator is called a low pass filter.

Example. *Over* $\mathbf{R}$, *we consider the Fourier multiplier*

$$m(\xi) = -i \cdot sgn(\xi).$$

Then $m(D)$ is the Hilbert transform.

Example. *In* $\mathbf{R}^d$, *we consider the Fourier multiplier*

$$m_R(\xi) = \mathbf{I}(|\xi| \leqslant R).$$

The operator $m_R(D)$ is known as the ball multiplier operator. *More generally, given any compact set S we can consider the Fourier multiplier $\mathbf{I}_S(D)$. In the engineering literature these multipliers are called* ideal low pass filters.

Example. *In this chapter, it is natural to renormalize the differentiation operators $D^\alpha : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ so that for $f \in \mathcal{S}(\mathbf{R}^d)$,*

$$\widehat{D^\alpha f} = \xi^\alpha \widehat{f}.$$

In particular, this implies that if $m(\xi) = \xi_i^\alpha$, then $m(D) = D^\alpha$. More generally, if $m(\xi) = \sum_{|\alpha| \leqslant k} c_\alpha \xi^\alpha$, then

$$m(D) = \sum_{|\alpha| \leqslant k} c_\alpha D^\alpha.$$

Thus the family of Fourier multiplier operators contains all constant coefficient differential operators.

Fourier multiplier operators have proved essential to our study of classical Fourier analysis. In particular, we have used Fourier multiplier operators to prove a great many results; the convolution operator by the Poisson kernel is a Fourier multiplier given by the symbol $e^{-|x|}$, and the heat kernel is a Fourier multiplier with symbol $e^{-\pi|x|^2}$. This is no coincidence. It is a general heuristic that any well-behaved translation invariant operator is given by convolution with an appropriate function.

We have already seen in our study of distributions that any translation invariant continuous linear operator $T : C_c^\infty(\mathbf{R}^d) \to C^\infty(\mathbf{R}^d)$ is given by convolution with a distribution. If the distribution is tempered, we can take the Fourier transform to conclude that the operator is a Fourier multiplier operator. In fact, if $1 \leqslant p, q \leqslant \infty$ and a translation invariant operator T satisfies a bound of the form

$$\|Tf\|_{L^q(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

for any $f \in \mathcal{S}(\mathbf{R}^d)$, then T is a Fourier multiplier operator. To prove this, we rely on an elementary Sobolev embedding.

Lemma 19.1. *Suppose $1 \leqslant p, q \leqslant \infty$. If $f \in L^p(\mathbf{R}^d)$ has a strong derivative $D^\alpha f$ in $L^p(\mathbf{R}^d)$ for all $|\alpha| \leqslant d+1$, then $f \in C(\mathbf{R}^d)$, and*

$$\|f\|_{L^\infty(\mathbf{R}^d)} \lesssim_{d,p} \sum_{|\alpha| \leqslant d+1} \|D^\alpha f\|_{L^p(\mathbf{R}^d)}.$$

Proof. Let us first suppose $p = 1$. Then

$$|\widehat{f}(x)| \lesssim \frac{\sum_{|\alpha| \leqslant d+1} |x^\alpha \widehat{f}(x)|}{(1+|x|)^{d+1}}.$$

Since $1/(1+|x|)^{d+1} \in L^1(\mathbf{R}^d)$, we can integrate both sides of the equation to conclude that

$$\|\widehat{f}\|_{L^1(\mathbf{R}^d)} \lesssim \sum_{|\alpha| \leqslant d+1} \|D^\alpha f\|_{L^1(\mathbf{R}^d)}.$$

It follows by the Fourier inversion formula that $f \in C(\mathbf{R}^d)$, and moreover,

$$\|f\|_{L^\infty(\mathbf{R}^d)} \leqslant \sum_{|\alpha| \leqslant d+1} \|D^\alpha f\|_{L^1(\mathbf{R}^d)},$$

which completes the proof for $p = 1$.

For $p > 1$, any compactly supported bump function ϕ, and any multi-index α with $|\alpha| \leqslant d+1$,

$$\|D^\alpha(\phi f)\|_{L^1(\mathbf{R}^d)} \leqslant \sum_{\beta \leqslant \alpha} \|D^\beta \phi \cdot D^{\alpha-\beta} f\|_{L^1(\mathbf{R}^d)} \lesssim_\phi \sum_{\beta \leqslant d+1} \|D^\beta f\|_{L^p(\mathbf{R}^d)}.$$

It follows from the previous case that $\phi f \in C(\mathbf{R})$, and

$$\|\phi f\|_{L^\infty(\mathbf{R}^d)} \lesssim_\phi \sum_{\beta \leqslant d+1} \|D^\beta f\|_{L^p(\mathbf{R}^d)}.$$

These bounds hold uniformly over translates of ϕ, and taking advantage of this shows that $f \in C(\mathbf{R})$, and that

$$\|f\|_{L^\infty(\mathbf{R}^d)} \lesssim \sum_{\beta \leqslant d+1} \|D^\beta f\|_{L^p(\mathbf{R}^d)}. \qquad \square$$

Theorem 19.2. *Suppose $1 \leqslant p, q \leqslant \infty$, and $T : \mathcal{S}(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ is a linear map commuting with translations and satisfies*

$$\|Tf\|_{L^q(\mathbf{R}^d)} \leqslant \|f\|_{L^p(\mathbf{R}^d)}$$

for all $f \in \mathcal{S}(\mathbf{R}^d)$. Then T is a Fourier multiplier operator.

Proof. For any $f \in \mathcal{S}(\mathbf{R}^d)$, $Tf \in W^{q,n}(\mathbf{R}^d)$ for any $n > 0$. To see this, we note that for any $h > 0$ and $k \in \{1, \dots, d\}$, and

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

Then $\Delta_h(Tf) = T(\Delta_h f)$ because T is translation invariant. Since f is a Schwartz function, $\Delta_h f$ converges to $D^k f$ in $L^p(\mathbf{R}^d)$. Thus by continuity of f, Tf has a strong derivative $T(D^k f)$ in $L^q(\mathbf{R}^d)$. Induction shows Tf has strong derivatives of all orders. The last lemma shows that $Tf \in C(\mathbf{R}^d)$, and

$$\begin{aligned}
\|Tf\|_{L^\infty(\mathbf{R}^d)} &\lesssim \sum_{|\alpha| \leqslant n+1} \|D^\alpha(Tf)\|_{L^q(\mathbf{R}^d)} \\
&= \sum_{|\alpha| \leqslant n+1} \|T(D^\alpha f)\|_{L^q(\mathbf{R}^d)} \\
&\lesssim \sum_{|\alpha| \leqslant n+1} \|D^\alpha f\|_{L^q(\mathbf{R}^d)}.
\end{aligned}$$

The map $f \mapsto Tf(0)$ is thus a continuous operator on $\mathcal{S}(\mathbf{R}^d)$, and therefore defines a tempered distribution Λ. Translation invariance shows that $Tf = \Lambda * f$, and setting $m = \widehat{\Lambda}$ completes the proof. $\square$

Remark. It follows from this argument that if $T : \mathcal{S}(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ is a linear operator commuting with translations satisfying a bound

$$\|Tf\|_{L^q(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)},$$

then $Tf \in C^\infty(\mathbf{R}^d)$ and is slowly increasing, as is all of it's derivatives.

19.1 Frequency Localization

One use of Fourier multipliers is to localize the support of the Fourier tranform of a function to a portion of space. Given a compactly supported function m supported on some set S, and any function f, $\widehat{m(D)f} = m\widehat{f}$ is supported on S. Frequency localization has the additional feature that $m(D)f$ is smooth (actually analytic), since it's Fourier transform is compactly supported. One major advantage of localizing in frequency is that it enables us to control the smoothness of a function; if m is a cutoff function supported ξ_0, then $\widehat{m(D)f}$ is supported near ξ_0, and so we might expect that $\partial_\alpha(m(D)f) \approx \xi_0^\alpha f$. This makes frequency useful especially useful in problems involving derivatives. Another application of this principle results from uncertainty principle heuristics; if m is supported on a cube centred at the origin with sidelengths $R_1, \dots, R_d$, then $m(D)f$ will be roughly speaking, locally constant on 'dual rectangles' with sidelength $1/R_1, \dots, 1/R_d$. More generally, if $\widehat{f}$ is supported on a cube centred at ξ_0, then f acts roughly like a constant multiple of $e^{2\pi i \xi \cdot x}$ on dual rectangles. This trick comes up all over the place, e.g. in the study of wave packet decompositions of functions.

Let us consider some examples. If we consider the Fourier multiplier $\mathbf{I}_{[-R,R]}$ in $\mathbf{R}^1$, then this multiplier corresponds to the kernel

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

i.e. a sinc function. If f is supported on an interval I, then it follows from a simple estimate that

$$|(K * f)(x)| \lesssim \frac{\|f\|_{L^1(\mathbf{R})}}{d(x,I)}.$$

Thus the frequency localization is no longer supported on I, but decays away from this interval at a rate of $1/x$. Unfortunately,this is often not enough to obtain useful estimates. In $\mathbf{R}^d$, if we write $K = K_1 \otimes \dots \times K_d$ be the kernel associated with the rectangle multiplier, then we have a similar $1/x$ decay estimate in each direction.

We can do better if we use a smooth cutoff function, i.e. we choose a smooth non-negative function m compactly supported on I which equals one on the inner third of I, and satisfies $\|\partial_\alpha m\|_{L^\infty(\mathbf{R}^d)} \lesssim_\alpha |I|^{-|\alpha|}$ for all multi-indices α, and then consider the Fourier multiplier $m(D)$. Then $m(D)$ corresponds to the convolution kernel

$$K(x) = \int e^{2\pi i \xi \cdot x} m(\xi)\, d\xi.$$

Integrating by parts gives that for all $n > 0$, $|K(x)| \lesssim_n |I|^{1-n}|x|^{-n}$, which decreasing away from the origin much faster than for a rough cutoff. This implies that we have the decay estimates

$$|(K * f)(x)| \lesssim_n \frac{|I|^{1-n}\|f\|_{L^1(\mathbf{R}^d)}}{d(x,J)^n}.$$

for f supported on an interval J, and all $n > 0$. This is often much more useful than the estimates above, and indicates that such operators 'almost' preserve spatial localization as well.

19.2 L^p Regularity

We now wish to study conditions on m which guarantee bounds of the form

$$\|m(D)f\|_{L^q(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}.$$

for all $f \in \mathcal{S}(\mathbf{R}^d)$. This question was of course prominant throughout the study of Harmonic analysis, but the explicit problem of studying general multipliers was pushed forwards by analysts like Hörmander in the 1960s. Littlewood's principle tells us that the only interesting case occur with 'the larger exponent on the left'.

Theorem 19.3. *Fix $1 \leqslant q < p \leqslant \infty$, and suppose m is a tempered distribution on $\mathbf{R}^d$ satisfying a uniform bound*

$$\|m(D)f\|_{L^q(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

for all $f \in \mathcal{S}(\mathbf{R}^d)$. Then $m = 0$.

Proof. Suppose $m \neq 0$ and $q < p$. Then there is $f_0 \in \mathcal{S}(\mathbf{R}^d)$ with $m(D)f_0 \neq 0$. Thus $m(D)f_0$ lies in $C^\infty(\mathbf{R}^d) \cap L^q(\mathbf{R}^d)$. Fix a large integer N and pick $x_1, \dots, x_N \in \mathbf{R}^d$ separated far enough apart that

$$\left\| \sum_{n=1}^N \text{Trans}_{x_n} f_0 \right\|_{L^p(\mathbf{R}^d)} \gtrsim N^{1/p} \|f_0\|_{L^p(\mathbf{R}^d)}$$

and

$$\left\| \sum_{n=1}^N \text{Trans}_{x_n} m(D) f_0 \right\|_{L^q(\mathbf{R}^d)} \sim N^{1/q} \|m(D)f_0\|_{L^q(\mathbf{R}^d)} \lesssim N^{1/q} \|f_0\|_{L^p(\mathbf{R}^d)}.$$

Translation invariance of convolution shows $N^{1/q} \lesssim N^{1/p}$, which is impossible for suitably large N. Thus $m = 0$. $\square$

In general, a characterization of the tempered distributions which give bounded convolution operators is unknown except in a few very particular situations. For each $1 \leqslant p \leqslant q \leqslant \infty$, we let $\|m\|_{M^{p,q}(\mathbf{R}^d)}$ denote the operator norm of the multiplier operator $m(D)$ from $L^p(\mathbf{R}^d)$ to $L^q(\mathbf{R}^d)$, i.e. the smallest quantity such that

$$\|m(D)f\|_{L^q(\mathbf{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbf{R}^d)} \|f\|_{L^p(\mathbf{R}^d)}$$

for all $f \in \mathcal{S}(\mathbf{R}^d)$. We let $M^{p,q}(\mathbf{R}^d)$ be the set of tempered distributions for which the bound is finite. For simplicity, we also let $M^p(\mathbf{R}^d)$ denote $M^{p,p}(\mathbf{R}^d)$. These all form Banach spaces. By symmetries of the Fourier transform, it is easy to check that translations, modulations, and dilations all preserve the space $M^{p,q}(\mathbf{R}^d)$, though dilations need not preserve the $M^{p,q}$ norm unless $p = q$.

Example. *We claim that $M^{2,2}(\mathbf{R}^d) = L^\infty(\mathbf{R}^d)$, in the sense that the two spaces consist of the same family of distributions, and the norms on both spaces are equal, i.e. $\|m\|_{M^{2,2}(\mathbf{R}^d)} = \|m\|_{L^\infty(\mathbf{R}^d)}$. To see this, let*

$$\Phi(x) = e^{-\pi|x|^2}$$

be the Gaussian distribution. Then Φ is a Schwartz function, and satisfies

$$\widehat{m(D)\Phi} = \Phi \cdot m.$$

Since $\Phi \in L^2(\mathbf{R}^d)$, $m(D)\Phi \in L^2(\mathbf{R}^d)$, and so $\Phi \cdot m \in L^2(\mathbf{R}^d)$. But since $1/\Phi \in L^\infty_{loc}(\mathbf{R}^d)$, this implies that $m \in L^1_{loc}(\mathbf{R}^d)$. But now the result follows by Parseval's inequality, since a bound

$$\|m(D)f\|_{L^2(\mathbf{R}^d)} \leqslant C\|f\|_{L^2(\mathbf{R}^d)}$$

holds for all $f \in L^2(\mathbf{R}^d)$ if and only if

$$\|m \cdot g\|_{L^2(\mathbf{R}^d)} \leqslant C\|g\|_{L^2(\mathbf{R}^d)},$$

so we have reduced the study of boundedness of a multiplier on $L^2(\mathbf{R}^d)$ to the study of Hölder's inequality.

For any tempered distribution m and $f, g \in \mathcal{S}(\mathbf{R}^d)$, the fact that the Fourier transform is self adjoint implies that

$$\begin{aligned}\langle m(D)f, g\rangle &= \langle m \cdot \widehat{f}, \widehat{g}\rangle \\ &= \langle \widehat{f}, \overline{m} \cdot \widehat{g}\rangle \\ &= \langle f, \overline{m}(D)g\rangle.\end{aligned}$$

Thus we have an adjoint relation $m(D)^* = \overline{m}(D)$, which gives a natural duality theory for Fourier multiplier operators.

Theorem 19.4. *For any $1 \leqslant p, q \leqslant \infty$ and any tempered distribution m,*

$$\|m\|_{M^{p,q}(\mathbf{R}^d)} = \|m\|_{M^{q^*,p^*}(\mathbf{R}^d)}.$$

Proof. Using the adjoint relation, if

$$\|m(D)f\|_{L^q(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

then

$$\|m^*(D)f\|_{L^{p^*}(\mathbf{R}^d)} \lesssim \|f\|_{L^{q^*}(\mathbf{R}^d)}$$

But it is easy to calculate that if we set $[Ru](x) = u(-x)$, then for any $x \in \mathbf{R}^d$,

$$[m^*(D)f](x) = [m(D)(Rf^*)(-x)]^*$$

and so $\|m^*(D)f\|_{L^{p^*}(\mathbf{R}^d)} = \|m(D)f\|_{L^{p^*}(\mathbf{R}^d)}$. $\square$

In particular, if $1 \leqslant p \leqslant \infty$ and $m \in M^p(\mathbf{R}^d)$, then also $m \in M^{p^*,p^*}(\mathbf{R}^d)$ and so Riesz-interpolation implies $m \in M^{2,2}(\mathbf{R}^d)$. Thus if we are studying L^p to L^p boundedness for any $1 \leqslant p \leqslant \infty$, we may restrict our attention to bounded Fourier multipliers.

Example. *The only remaining space which we can completed characterize are the spaces $M^{1,q}(\mathbf{R}^d) = M^{p,\infty}(\mathbf{R}^d)$, where $1 \leqslant p,q \leqslant \infty$, which we can write as $\widehat{L^q(\mathbf{R}^d)}$ for $q > 1$, and the set $M(\mathbf{R}^d)$ of all finite Borel measures for $q = 1$. Moreover, in the case $q > 1$ we have*

$$\|m\|_{M^{1,q}(\mathbf{R}^d)} = \|\widehat{m}\|_{L^q(\mathbf{R}^d)}$$

and for $q = 1$,

$$\|m\|_{M^{1,1}(\mathbf{R}^d)} = \|\widehat{m}\|_{M(\mathbf{R}^d)}.$$

Given such an $m \in M^{1,q}(\mathbf{R}^d)$, if $K = \widehat{m}$, and $\{\phi_\varepsilon\}$ is a family of standard mollifiers, then

$$\|K * \phi_\varepsilon\|_{L^q(\mathbf{R}^d)} = \|m(D)(\phi_\varepsilon)\|_{L^q(\mathbf{R}^d)} \lesssim \|\phi_\varepsilon\|_{L^1(\mathbf{R}^d)} = 1.$$

Applying the Banach-Alaoglu theorem, the family $\{K * \phi_\varepsilon\}$ has a weak $*$ convergent subsequence in $L^q(\mathbf{R}^d)^{**}$. Thus there is $\lambda \in L^q(\mathbf{R}^d)^{**}$ and $\varepsilon_i \to 0$ such that $K * \phi_{\varepsilon_i} \to \lambda$. But this means that $K * \phi_{\varepsilon_i}$ converges to λ distributionally. It also converges to K distributionally, so $K = \lambda \in (L^q(\mathbf{R}^d))^{**}$. If $q > 1$, then $(L^q(\mathbf{R}^d))^{**} = L^q(\mathbf{R}^d)$, with the same norm, and $(L^1(\mathbf{R}^d))^{**} = M(\mathbf{R}^d)$.

For other values of p and q, finding a simple characterization of $M^p(\mathbf{R}^d)$ is a much more subtle task. For instance, it still remains an open question to determine the values of p and δ for which the multiplier

$$m^\delta(\xi) = \max((1 - |\xi|^2)^\delta, 0)$$

lies in $M^p(\mathbf{R}^d)$, a problem known as the *Bochner-Riesz conjecture*.

The difficulty here is that m^δ is singular on the boundary of the unit sphere, which is a large, curved set. Intuition suggests that a smooth Fourier multiplier would have a rapidly decaying Fourier transform, which would therefore be well posed as a convolution operator. One basic instance of this phenomenon occurs if $m \in \mathcal{S}(\mathbf{R}^d)$. Then $\widehat{m} \in \mathcal{S}(\mathbf{R}^d)$, hence integrable, and by Young's convolution inequality, it follows that for any $p \leqslant q$, $\|m\|_{M^{p,q}(\mathbf{R}^d)} \lesssim 1$, and the induced map from $\mathcal{S}(\mathbf{R}^d)$ to $M^{p,q}(\mathbf{R}^d)$ is continuous. Another example is the following calculation: if $\Omega \subset \mathbf{R}^d$ is bounded and open, C is a fixed constant, and $L : \mathbf{R}^d \to \mathbf{R}^d$ is an invertible linear map, then we say ϕ is a *bump function adapted to* $L(\Omega)$ if ϕ is smooth and supported in $L(\Omega)$, and $\nabla^k(\phi \circ L)(x) \leqslant C$ for all $x \in \Omega$. If m is a bump function adapated to $L(\Omega)$, then it follows that $\|m\|_{M^{p,q}(\mathbf{R}^d)} \lesssim_{C,\Omega} 1$.

If the multiplier m is only singular on a small set, we can likely still obtain some interesting estimates. For instance, the Hilbert transform, corresponding to the Fourier multiplier $m(\xi) = i\text{sgn}(\xi)$, is singular at $\xi = 0$, but still satisfies the bounds

$$\|Hf\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}$$

for all $f \in L^p(\mathbf{R}^d)$ and $1 < p < \infty$. One can see this from the Calderon-Zygmund theory of singular integrals, or via the Hörmander-Mikhlin theory we will develop shortly. It follows from translation symmetry that for $1 < p < \infty$ and any (possibly unbounded interval) I,

$$\|\mathbf{I}_I\|_{M^p(\mathbf{R}^d)}, \|\mathbf{I}_I\|_{M^p(\mathbf{R}^d)} \lesssim_p 1.$$

Now suppose $m \in L^\infty(\mathbf{R}^d)$ has *bounded variation*, which means the quantity

$$V(m) = \sup_{\xi_1 < \dots < \xi_N} \sum_{i=1}^{N-1} |m(\xi_{i+1}) - m(\xi_i)|.$$

is finite. Then m has countably many discontinuities, and the variation prevents too much nonsmoothness. In this case, we can obtain an operator norm bound.

Theorem 19.5. *For any symbol m on $\mathbf{R}$, and any $1 < p < \infty$,*

$$\|m\|_{M^p(\mathbf{R})} \lesssim_p \|m\|_{L^\infty(\mathbf{R})} + V(m).$$

Proof. For each n, pick $\xi_1, \dots, \xi_{N_n}$ such that

$$\sum_{i=1}^{N-1} |m(\xi_{i+1}) - m(\xi_i)| \geqslant V(m) - 1/n.$$

If we define

$$m_n = m(\xi_1)\mathbf{I}_{(-\infty,\xi_1)} + \sum_{i=1}^{N_n-1} m(\xi_i)\mathbf{I}_{(\xi_i,\xi_{i+1})} + m(\xi_N)\mathbf{I}_{(\xi_N,\infty)}$$

Then $\|m - m_n\|_{L^1(\mathbf{R})} \leqslant 1/n$. But this means that $m_n(D)$ converges to $m(D)$ weakly as operators on $\mathcal{S}(\mathbf{R})$, and so

$$\|m\|_{M^p(\mathbf{R})} \leqslant \limsup_{n\to\infty} \|m_n\|_{M^p(\mathbf{R})}.$$

Now we can rewrite

$$m_n(\xi) = m(\xi_1)\mathbf{I}_{(-\infty,\xi_1)} + \sum_{i=1}^{N-1} [m(\xi_i) - m(\xi_{i+1})]\mathbf{I}_{(\xi_1,\xi_i)}(\xi) + m(\xi_N)\mathbf{I}_{(\xi_N,\infty)}.$$

Since the multiplier operators corresponding to indicator functions of intervals are uniformly bounded in $M^p(\mathbf{R})$ for $1 < p < \infty$, we conclude that

$$\|m_n\|_{M^p(\mathbf{R})} \lesssim_p |m(\xi_1)| + \sum_{i=1}^{N-1} |m(\xi_i) - m(\xi_{i+1})| + |m(\xi_N)| \leqslant \|m\|_{L^\infty(\mathbf{R})} + V(m).$$

Taking $n \to \infty$ gives $\|m\|_{M^p(\mathbf{R})} \lesssim_p \|m\|_{L^\infty(\mathbf{R})} + V(m)$. $\square$

The theorem of Hörmander-Mikhlin is a more sophisticated instance of the smoothness-regularity phenomenon, giving L^p to L^p bounds to Fourier multipliers which decay smoothly away from the origin. There are several formulations of the phenomenon, of greater and greater generality. Without loss of generality, we will assume that $m \in L^\infty(\mathbf{R}^d)$, since otherwise m cannot lie in any of the spaces M^p. Let us state them in increasing order of generality, the former being most easily seen as relating to the smoothness of the multipliers:

- The multiplier m lies in $C^\infty(\mathbf{R}^d - \{0\})$, and satisfies

$$|D_\xi^\beta m(\xi)| \lesssim_\beta |\xi|^{-\beta} \tag{19.1}$$

  for all multi-indices β. In particular, this is true if m is smooth away from the origin, and homogeneous of degree zero.

- For some integer N with $N > d/2$, m has weak derivatives up to order N which are functions, and satisfy (19.1) for all β with $|\beta| \leqslant N$.

- For some integer $N > d/2$, m has weak derivatives in $L^2_{\text{loc}}(\mathbf{R}^d)$ up to order N, and satisfies an L^2 average estimate

$$\sup_{R>0} \left(\fint_{|\xi|\leqslant 2R} |\xi|^{2\beta} |D_\xi^\beta m(\xi)|^2 \, d\xi \right)^{1/2} \lesssim 1.$$

- For some non-zero, smooth radial test function φ, compactly supported away from the origin,

$$\sup_{R>0} \|\varphi \, \text{Dil}_R m\|_{L^2_\alpha} < \infty \tag{19.2}$$

  for some $\alpha > d/2$ (not necessarily an integer).

- For some φ as above, and for some $\varepsilon > 0$,

$$\sup_{R>0} \int |\mathcal{F}(\varphi \text{Dil}_R m)(x)|(1 + |x|)^\varepsilon \, dx < \infty.$$

If the second last point holds, the ε in the last point can be chosen to be any number smaller than $\alpha - d/2$.

Theorem 19.6. *Consider $m \in L^\infty(\mathbf{R}^d)$ and suppose there exists $\varepsilon > 0$ and φ as above with*

$$\sup_{R>0} \int |\mathcal{F}(\varphi Dil_R m)(x)|(1 + |x|)^\varepsilon \, dx < \infty.$$

Then for any $1 < p < \infty$ and $f \in \mathcal{S}(\mathbf{R}^d)$,

$$\|m(D)f\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}.$$

Moreover, we have a bound $\|m(D)f\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}$, which, together with the fact that $m \in M^2(\mathbf{R}^d)$ because it is bounded, implies the bounds above by Riesz-Thorin and duality.

Proof. If the assumptions hold for some ε, it is simple to argue they hold for *any* such φ. In particular, if we take a non-negative φ such that for any $\xi \neq 0$,

$$\sum_{n=-\infty}^{\infty} \varphi(2^n \xi) = 1.$$

Write $m_n(\xi) = \varphi(\xi) m(2^n \xi)$, and $K_n = \widehat{m_n}$. Then our assumption implies that

$$\int K_n(x)(1+|x|)^{\varepsilon}\, dx \lesssim 1,$$

In particular, this means that

$$\int_{|x| \geqslant R} |K_n(x)|\, dx \leqslant (1+R)^{-\varepsilon}.$$

uniformly in n. Similarily, Bernstein's inequality implies that

$$\|\nabla K_n\|_{L^1(\mathbf{R}^d)} \lesssim \|K_n\|_{L^1(\mathbf{R}^d)} \lesssim 1.$$

This implies that for all $y \in \mathbf{R}^d$, uniformly in n,

$$\int |K_n(x+y) - K_n(x)|\, dx \lesssim |y|.$$

For any $f \in \mathcal{S}(\mathbf{R}^d)$, we have

$$m(D)f = \sum_{n=-\infty}^{\infty} (\mathrm{Dil}_{2^n} m_n)(D),$$

where the sum converges absolutely in $L^p(\mathbf{R}^d)$. Since $\widehat{\mathrm{Dil}_{2^n} m_n} = 2^{nd} \mathrm{Dil}_{1/2^n} K_n$, it follows that

$$K * f = \sum_{n=-\infty}^{\infty} 2^{nd} \cdot (\mathrm{Dil}_{1/2^n} K_n) * f.$$

The cancellation bound we obtained for the functions K_n indicate the singular-integral nature of the kernels K_n, i.e. they satisfy a cancellation criterion, and the singularitity is quantitatively concentrated near the origin. Indeed, the cancellation bound allows us to use the standard Calderon-Zygmund singular integral results to obtain a uniform bound

$$\|K_n * f\|_{L^p(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}.$$

in n. However, these bounds cannot be summed in n as $n \to \infty$ to yield an L^p bound on $K * f$, since the L^p operator norm of $\text{Dil}_{1/2^n} K_n$ is the same as the L^p operator norm of K_n. To get a better bound, we must utilize the fact that the functions decay away from the origin more strongly. We will still use the Calderon Zygmund decomposition to understand this. So fix an integrable function b supported on a cube Q of some sidelength R, with $\int b(x)\, dx = 0$. Since convolution commutes with translation, we may assume without loss of generality for the calculation of L^p norms that Q is centred at the origin. Let Q^* denote the cube with the same centre and twice the width. Then

$$\begin{aligned}
\int_{(Q^*)^c} 2^{nd} |(\text{Dil}_{1/2^n} K_n) * b| &\leqslant 2^{nd} \int_{(Q^*)^c} \left| \int_Q K_n(2^n(x-y)) b(y)\, dy \right| dx \\
&\leqslant 2^{nd} \int_{(Q^*)^c} \left| \int_Q K_n(2^n(x-y)) b(y)\, dy \right| dx \\
&= 2^{nd} \int_{(Q^*)^c} \left| \int_Q [K_n(2^n(x-y)) - K_n(2^n x)] b(y)\, dy \right| dx \\
&\leqslant 2^{nd} \int_Q \int_{(Q^*)^c} |b(y)| |K_n(2^n(x-y)) - K_n(2^n x)|\, dx\, dy \\
&= \int_Q \int_{(Q^*)^c} |b(y)| |K_n(x - 2^n y) - K_n(x)|\, dx\, dy \\
&\lesssim 2^n \int_Q |y| |b(y)|\, dx \lesssim 2^n R \cdot \|b\|_{L^1(\mathbf{R}^d)}.
\end{aligned}$$

This is a standard Calderon-Zygmund type calculation, as in the singular integral theory, and gives good bounds for $R \leqslant 1/2^n$. For $R \geqslant 1/2^n$, we apply the decay estimate to get a more optimal bound, writing

$$\begin{aligned}
\int_{(Q^*)^c} 2^{nd} |(\text{Dil}_{1/2^n} K_n) * b| &\leqslant 2^{nd} \int_{(Q^*)^c} \int_Q |K_n(2^n(x-y))| |b(y)|\, dy\, dx \\
&\leqslant 2^{nd} \int_{|z| \geqslant R} \int_Q |K_n(2^n z)| |b(y)|\, dy\, dz \\
&\leqslant (1 + R/2^n)^{-\varepsilon} \|b\|_{L^1(\mathbf{R}^d)}
\end{aligned}$$

This is a good for $R \geqslant 2^n$. In particular, we now have good bounds for any value of R, which we can sum up to conclude that

$$\int_{(Q^*)^c} |K * b| \lesssim \|b\|_{L^1(\mathbf{R}^d)}.$$

Now given a general $f \in \mathcal{S}(\mathbf{R}^d)$, we apply a Calderon-Zygmund decomposition, fixing $\alpha > 0$, and writing

$$f = g + \sum_{k=1}^{\infty} b_k$$

where $\|g\|_{L^1(\mathbf{R}^d)} + \sum_{k=1}^{\infty} \|b_k\|_{L^1(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}$, $\|g\|_{L^\infty(\mathbf{R}^d)} \lesssim \alpha$, and there exists a family of disjoint cubes $\{Q_k\}$ such that b_k is supported on Q_k, $\int b_k = 0$, and $\sum_{k=1}^{\infty} |Q_k| \lesssim \alpha^{-1} \|f\|_{L^1(\mathbf{R}^d)}$. If $\Omega = \bigcup_{k=1}^{\infty} Q_k$, then

$$\int_{\Omega^c} |K * (\sum_{k=1}^{\infty} b_k)| \lesssim \sum_{k=1}^{\infty} \|b_k\|_{L^1(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}.$$

Thus

$$|\{x \in \Omega^c : |K * (\sum_{k=1}^{\infty} b_k)| \geqslant \alpha/2\}| \lesssim \|f\|_{L^1(\mathbf{R}^d)}/\alpha,$$

and since $|\Omega| \leqslant \|f\|_{L^1(\mathbf{R}^d)}/\alpha$,

$$|\{x \in \mathbf{R}^d : |K * (\sum_{k=1}^{\infty} b_k)| \geqslant \alpha/2\}| \lesssim \|f\|_{L^1(\mathbf{R}^d)}/\alpha.$$

Since α was arbitrary, we conclude that $\|K * (\sum b_k)\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}$. Next, we use the fact that K is bounded on L^2 (since m is bounded) to conclude that

$$\|K * g\|_{L^2(\mathbf{R}^d)} \lesssim \|g\|_{L^2(\mathbf{R}^d)} \lesssim \alpha^{1/2} \|g\|_{L^1(\mathbf{R}^d)}^{1/2}$$

and so

$$|\{x \in \mathbf{R}^d : |(K * g)(x)| \geqslant \alpha/2\}| \lesssim \|K * g\|_{L^2(\mathbf{R}^d)}^2 \alpha^{-2} \lesssim \|g\|_{L^1(\mathbf{R}^d)} \alpha^{-1}.$$

But now we know $\|K * g\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim \|g\|_{L^1(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}$. We sum and get $\|K * f\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim \|f\|_{L^1(\mathbf{R}^d)}$. □

Remark. A fastidious reader may complain that the Calderon-Zygmund decomposition of a Schwartz function yields a family of functions that are only integrable, and not necessarily even continuous. To remedy this situation, we replace the kernel operator K we are originally studying with the more regularized kernels $K^{\leqslant N} = \sum_{|n| \leqslant N} 2^{nd} \cdot \mathrm{Dil}_{1/2^n} K_n$. The operators $K^{\leqslant N}$ are then Schwartz convolution kernels, and, independently of this proof, are thus easily seen to lie in $M^p(\mathbf{R}^d)$ for all $1 \leqslant p \leqslant \infty$, and thus extend to be well defined on the components of Calderon-Zygmund decompositions of functions. The important part of this proof is that it shows that the operators $K^{\leqslant N}$ are *uniformly* when they induce convolution operators from $L^1(\mathbf{R}^d)$ to $L^{1,\infty}(\mathbf{R}^d)$, and thus imply the resultant bounds for the limiting convolution kernel K.

Since convolution can be seen as a way of smoothing out some of the irregularities of a function, one might ask whether the convolution of a multiplier in $M^{p,q}(\mathbf{R}^d)$ with a function lies in $M^{p,q}(\mathbf{R}^d)$. One can obtain such a result, provided the function is integrable.

Theorem 19.7. *If $u \in L^1(\mathbf{R}^d)$, then $\|m * u\|_{M^{p,q}(\mathbf{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbf{R}^d)} \|u\|_{L^1(\mathbf{R}^d)}$.*

Proof. Let $K = \widehat{m}$. If $v = \widehat{u}$, then

$$\begin{aligned}
(m * u)(D) f(x) &= \int K(x-y) v(x-y) f(y)\, dy \\
&= \int u(\xi) \int K(x-y) f(y) e^{2\pi i (x-y)\cdot \xi}\, dy\, d\xi \\
&= \int u(\xi) \cdot (\mathrm{Mod}_\xi K * f)(x)\, d\xi \\
&= \int u(\xi) (\mathrm{Trans}_\xi m)(D) f(x)\, d\xi.
\end{aligned}$$

Since $\|\mathrm{Trans}_\xi m\|_{M^{p,q}(\mathbf{R}^d)} = \|m\|_{M^{p,q}(\mathbf{R}^d)}$, the result follows. ☐

One might ask whether multipliers operators need at least *some* regularity to get a bound at all. Provided that $1 \leqslant p \leqslant q \leqslant 2$, one at least needs some kind of integrability condition.

Theorem 19.8. *If $m \in M^{p,q}(\mathbf{R}^d)$, and $1 \leqslant p \leqslant q \leqslant 2$, $M^{p,q}(\mathbf{R}^d) \subset L^q(\mathbf{R}^d)^*$.*

Proof. TODO ☐

On the other hand, if $q > 2$, then $M^{p,q}(\mathbf{R}^d)$ contains distributions of positive order (distributions in $M_{\mathrm{loc}}(\mathbf{R}^d)$). We will show an explicit examples for $d \geqslant 4$, and non constructive examples otherwise.

Example. *Recall that the surface measure* σ *on the sphere in* $\mathbf{R}^d$ *satisfies* $|\widehat{\sigma}(\xi)| \lesssim_d |\xi|^{\frac{1-d}{2}}$. *Consider the multiplier corresponding to the distribution* $\Lambda = \partial\sigma/\partial x$, *which is not a distribution of order zero. On the other hand,* $|\widehat{\Lambda}(\xi)| \lesssim_d |\xi|^{\frac{3-d}{2}}$. *Provided* $d \geqslant 4$, *this implies that* $\|\widehat{\Lambda}\|_{L^q(\mathbf{R}^d)} < \infty$ *for* $(d-3)/2q > d$, *i.e. for* $q > 2d/(d-3) = 2 + 6/(d-3)$, *and thus* $\|\Lambda\|_{M^{1,q}(\mathbf{R}^d)} < \infty$ *for this range. Taking* d *appropriately large, applying duality, and iterpolating gives, for each* $\varepsilon > 0$, *a distribution* Λ *lying in* $M^{p,q}(\mathbf{R}^d)$ *for any* $p \leqslant 2 - \varepsilon$ *and any* $q \geqslant 2 + \varepsilon$.

Example. *Our other, non constructive examples apply Baire category arguments. Fix* $\eta \in C_c^\infty(\mathbf{R}^d)$, *and for any* $\lambda > 0$, *let* $m_\lambda(\xi) = \eta(\xi)e^{2\pi i\lambda|\xi|^2}$. *We claim that if* $K_\lambda = \widehat{m_\lambda}$, *then* $\|K_\lambda\|_{L^\infty(\mathbf{R}^d)} \lesssim \lambda^{-1/2}$. *This follows from a simple oscillatory integral argument. Thus it follows that* $\|m_\lambda\|_{M^{1,2}(\mathbf{R}^d)} \lesssim 1$ *and* $\|m_\lambda\|_{M^{1,\infty}} \lesssim \lambda^{-1/2}$. *Interpolating these bounds gives, for each* $q > 2$, *a bound of the form* $\|m_\lambda\|_{M^{1,q}(\mathbf{R}^d)} \lesssim \lambda^{-\varepsilon}$ $for some \varepsilon > 0$. *If* $M^{1,q}(\mathbf{R}^d)$ *solely contained elements of* $M_{loc}(\mathbf{R}^d)$, *if* $\chi \in C_c^\infty(\mathbf{R}^d)$ *equal to one in a neighborhood of the support of* η, *then the operator* $m \mapsto \chi m$ *would be a bounded operator from* $M^{1,q}(\mathbf{R}^d)$ *to* $M(\mathbf{R}^d)$ *by the closed graph theorem. But this would imply that for* $q > 2$,

$$\|\eta\|_{L^1(\mathbf{R}^d)} = \|\chi m_\lambda\|_{M(\mathbf{R}^d)} \lesssim \|m_\lambda\|_{M^{1,q}(\mathbf{R}^d)} \lesssim \lambda^{-\varepsilon},$$

which certainly cannot be true.

De Leeuw's theorem shows slices of continuous $d+1$ dimensional multipliers are bounded by the original mutiplier.

Theorem 19.9. *Let* $m \in C(\mathbf{R}^{d+1})$. *For each* $\xi_0 \in \mathbf{R}$ *define* $m_0 \in C(\mathbf{R}^d)$ *by setting*

$$m_0(\xi) = m(\xi, \xi_0).$$

Then for any $1 \leqslant p \leqslant \infty$, $\|m_0\|_{M^p(\mathbf{R}^d)} \leqslant \|m\|_{M^p(\mathbf{R}^{d+1})}$.

Proof. Without loss of generality, assume $\xi_0 = 0$. For $\lambda > 0$ set

$$L(\xi_1, \dots, \xi_{d+1}) = (\xi_1, \dots, \xi_d, \xi_{d+1}/\lambda).$$

Then

$$\|m \circ L_\lambda\|_{M^p(\mathbf{R}^{d+1})} = \|m\|_{M^p(\mathbf{R}^{d+1})}.$$

Take $\lambda \to \infty$. Since m is continuous, $m \circ L_\lambda$ converges to $m \circ L_\infty$ pointwise, where $L_\infty(\xi_1, \dots, \xi_{d+1}) = (\xi_1, \dots, \xi_d, 0)$. On the other hand,

$$\|m \circ L_\infty\|_{M^p(\mathbf{R}^d)} = \|m_0\|_{M^p(\mathbf{R}^d)}.$$

Indeed, this follows from the simple fact that $(m \circ L_\infty)(D) = m_0(D) \otimes 1$, and $m_0(D) \otimes 1$ has the same operator norm as $m_0(D)$. Thus it suffices to show that

$$\|m \circ L_\infty\|_{M^p(\mathbf{R}^d)} \leqslant \limsup_{\lambda \to \infty} \|m \circ L_\lambda\|_{M^p(\mathbf{R}^d)}.$$

But this follows from a simple weak convergence argument; for any $f, g \in \mathcal{S}(\mathbf{R}^{d+1})$, the dominated convergence theorem implies that

$$\lim_{\lambda \to \infty} |\langle (m \circ L_\lambda)(D) f, g \rangle| = \langle (m \circ L_\infty)(D) f, g \rangle. \qquad \square$$

If m is no longer continuous, this theorem only holds for *almost all* slices of the function.

Theorem 19.10. *Fix $m \in M^p(\mathbf{R}^{d+1})$, and for each λ, let $m_\lambda(\xi) = m(\xi, \lambda)$ be defined on $\mathbf{R}^d$. Then for almost every $\lambda \in \mathbf{R}$, $\|m_\lambda\|_{M^p(\mathbf{R}^d)} \leqslant \|m\|_{M^p(\mathbf{R}^{d+1})}$.*

Proof. By duality and the multiplication formula for the Fourier transform,

$$\left| \int m(\xi) \widehat{f}(\xi) \widehat{g}(\xi)\, d\xi \right| \leqslant \|m\|_{M^p(\mathbf{R}^{d+1})} \|f\|_{L^p(\mathbf{R}^{d+1})} \|g\|_{L^q(\mathbf{R}^{d+1})},$$

where q is the dual of p. If $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ then

$$\left| \int m_\lambda(\xi) \widehat{f_1}(\xi) \widehat{f_2}(\lambda) \widehat{g_1}(\xi) \widehat{g_2}(\lambda)\, d\xi\, d\lambda \right| \leqslant \|m\|_{M^p(\mathbf{R}^d)} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$

Write the left hand side as

$$\left| \int J(\lambda) \widehat{f_2}(\lambda) \widehat{g_2}(\lambda)\, d\lambda \right|,$$

where

$$J(\lambda) = \int m_\lambda(\xi) \widehat{f_1}(\xi) \widehat{f_2}(\xi)\, d\xi.$$

The inequality above this implies that

$$\|J\|_{L^\infty(\mathbf{R})} \leqslant \|J\|_{M^p(\mathbf{R})} \leqslant \|m\|_{M^p(\mathbf{R}^{d+1})} \|f_1\|_{L^p(\mathbf{R}^d)} \|g_1\|_{L^p(\mathbf{R}^d)}.$$

But this implies precisely that for almost every $\lambda \in \mathbf{R}$,

$$|J(\lambda)| \leqslant \|f_1\|_{L^p(\mathbf{R}^d)} \|g_1\|_{L^p(\mathbf{R}^d)},$$

which completes the proof (modulo a separability argument). $\square$

Chapter 20

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 20.1. *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. $\square$

20.1 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 20.2. *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. $\square$

If we are a little more careful, then we can fully approximate elements of $W^{k,p}(\Omega)$ by smooth functions on U.

Theorem 20.3. *$C_c^\infty(\Omega) \cap W^{k,p}(\Omega)$ 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 $W^{k,p}(\Omega)$. $\square$

Approximation by elements of $C^\infty(\overline{\Omega})$ requires some more care, and additional assumptions on the behaviour of $\partial\Omega$.

Chapter 21

Time Frequency Analysis

In harmonic analysis, it is often useful to study a function $f : \mathbf{R}^d \to \mathbf{C}$ via it's Fourier transform $\widehat{f} : \mathbf{R}^d \to \mathbf{C}$. The goal of time-frequency analysis is to think of such a function as being a function living simultaneously in both spaces, i.e. a function on the domain $\mathbf{R}^d \times \mathbf{R}^d$, which can be either written in temporal coordinates, or frequential coordinates, and in certain special cases, a combination of the two. We call this space the *phase plane*, combining both time and frequency information together. There are several difficulties with rigorously incorporating this approach, for instance, resulting from the uncertainty principle, but the utility makes this . Given a function f, we define a *phase portrait* for f to be a subset of $\mathbf{R}^d \times \mathbf{R}^d$ where the majority of the 'mass' of f and $\widehat{f}$ are concentrated (for an arbitrary locally compact abelian group G, phase space is $G^* \times G$). Let us consider a simple example.

Example. *Consider the Gaussian $f(t) = e^{-\pi t^2}$. Then 70% of the mass of f is concentrated on $[-1,1]$, and the mass decays exponentially away from this interval. Thus the function f is concentrated in $[-1,1]$. We have $\widehat{f_\delta}(\omega) = e^{-\pi\omega^2}$, which similarily, is concentrated in $[-1,1]$. A natural choice of the phase portrait of f is therefore $[-1,1] \times [-1,1]$.*

Example. *The Fourier transform of the Dirac delta function δ at a point x is the plane wave $\xi \mapsto e^{-2\pi i \xi \cdot x}$. Thus a natural phase portrait for the Dirac delta function is $\{x\} \times \mathbf{R}^d$. Similarily, the phase portrait of a pure plane wave $x \mapsto e^{2\pi i \xi \cdot x}$ is $\mathbf{R}^d \times \{\xi\}$, since the Fourier transform is the Dirac delta function at ξ.*

The symmetries of the Fourier transform have natural effects on the phase portrait of a function f, which has phase portrait S.

- The phase portrait of $\text{Trans}_x f$ is obtained by translating S horizontally by x units, since on the Fourier side the translation acts as a modulation, and does not move mass. Similarily, the phase portrait of the modulation $\text{Mod}_\xi f$ is obtained by translating S vertically by ξ units, since modulation does not affect the position of mass on the spatial side of things.

- Scaling in physical space has a 'dual' effect in phase space. More precisely, $\text{Dil}_\delta f$ has phase portrait

$$\{(x,\xi) \in \mathbf{R}^d \times \mathbf{R}^d : (x/\lambda, \lambda\xi) \in S\}$$

  More generally, given a linear transformation T, the phase portrait of $f \circ T^{-1}$ is equal to

$$\{(x,\xi) \in \mathbf{R}^d \times \mathbf{R}^d : (T(x), T^{-t}(\xi)) \in S\}.$$

  The rescaling above is a special case. In particular, if $T \in O(d)$, then $f \circ T^{-1}$ has phase portrait

$$\{(x,\xi) \in \mathbf{R}^d \times \mathbf{R}^d : (T(x), T(\xi)) \in S\}.$$

- The phase portrait of $\widehat{f}$ is equal to

$$\{(x,\xi) \in \mathbf{R}^d \times \mathbf{R}^d : (\xi, -x) \in S\},$$

  i.e. a clockwise rotation by ninety degrees.

Notice that all the transformations above preserve area, which is where symplectic geometry enters the picture.

We note that the phase portrait of a rescaled, translated, and modulated Gaussian has phase portrait consisting of an axis-oriented rectangle with sidelengths δx and $\delta \xi$, where $\delta x \cdot \delta \xi \sim 1$. Such a rectangle in phase space is called a *Heisenberg tile*, and functions whose phase portraits consist of Heisenberg tiles are called *wave packets*. In light of the uncertainty principle, these functions are the best alternatives to a function which is compactly supported in a sidelength one interval, and whose Fourier transform is also supported in a sidelength one interval.

21.1 Localization in Time and Space

Physical space localization is easy, we just multiply by a function, either a rough cutoff, or a smooth cutoff. Frequency localization is only slightly harder, where we can apply a basic Fourier multiplier. To localize in both time and frequency, one approach is to first smoothly localize in space, then time, or vice versa. This works out fine, as long as we do not localize too finely in both time and space, i.e. breaking the uncertainty principle. Here is a characteristic result in this setting

Lemma 21.1. *Fix two cubes I and J, and two smooth functions ψ_I and ψ_J adapted to I and J, and consider the localization operator $\pi_{I\times J} = \psi_I(D) \circ \psi_J(X)$. Then $\pi_{I\times J} f$ has Fourier support in I, and is localized in J in the sense that for $|x| \geqslant$*

$$(\pi_{I\times J} f)(x) \lesssim_n |I|^{1-n} |J|^{1/2} \|f\|_{L^2(\mathbf{R}^d)} d(x,J)^{-n}$$

Proof. If we let K_I be the inverse Fourier transform of ψ_I, then for all $n > 0$,

$$|K_I(x)| \lesssim_n |I|^{1-n} |x|^{-n}.$$

The proof then follows from Cauchy-Schwartz applied to the representation $\pi_{I\times J} f = K_I * (\psi_J \cdot f)$.

$$(\int |K_I(x-y)|^2 |\psi_J(y)|^2)^{1/2} \|f\|_{L^2(\mathbf{R}^d)}$$

$\square$

Chapter 22

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 22.1. *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. $\square$

Theorem 22.2 (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.

22.1 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

$$\lambda_n = \frac{\mathrm{sgn}(n)}{i} = \begin{cases} +1/i & n > 0 \\ 0 & n = 0 \\ -1/i & n < 0 \end{cases}$$

Because

$$[|n| \leqslant N] = \frac{\mathrm{sgn}(n+N) - \mathrm{sgn}(n-N)}{2} + \frac{[n = N] + [n = -N]}{2}$$

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 22.3. *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. □

Corollary 22.4. *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,

$$(Hf)(x) = \lim_{r \to 1} \int \cot(t/2) f(x-t)\, dt$$

22.2 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

$A_r(f)(t) =$

22.3 Conjugate Fourier Series

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

$$\sum_{n=-\infty}^{\infty} \widehat{f}(n)e_n(t) = \operatorname{Re}\left(\widehat{f}(0) + 2\sum_{n=1}^{\infty} \widehat{f}(n)e_n(t)\right)$$

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

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

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

Chapter 23

Oscillatory Integrals

The goal of the theory of oscillatory integrals is to obtain estimates of integrals with highly oscillatory integrands, 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

$$I(\lambda) = \int a(x) e^{2\pi i \lambda \phi(x)}\, dx,$$

where a is the *amplitude function*, and ϕ is the *phase*. The value λ is a parameter measuring the degree of oscillation. As λ increases, the degree of the oscillatory factor increases, which implies more cancellation should occur on average. Thus we should expect $I(\lambda)$ to decay as $\lambda \to \infty$. One of the main problems in the study of oscillatory integrals is to measure the asymptotic decay more precisely.

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 $\lambda \in \mathbf{R}$, we consider the quantity*

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

Thus in the oscillatory integral defining the Fourier transform of f, the function f plays the role of the amplitude, the phase function is $\phi(x) = x$. The basic theory of the Fourier transform hints that the decay of the oscillatory integral is related to the smoothness of the amplitude function f.

There are two main tools to estimate oscillatory integrals. Most classically, the method of steepest descent uses contour integration techniques from complex analysis to shift the integral to a domain where less oscillation occurs, so we can apply standard estimation strategies. However, it is difficult to apply this method to oscillatory integrals over multivariate domains. The second tool, known as the method of stationary phase, determines the behaviour of the decay of an oscillatory integral by isolating the oscillation of a smooth phase ϕ to points where $\nabla\phi$ is smooth. If the zeroes of $\nabla\phi$ are isolated, then the oscillatory integrals can be localized near these values. Roughly speaking, each zero x_0 contributes $\psi(x_0)e^{2\pi i\lambda\phi(x_0)}$, times the volume of the region around x_0 where ϕ deviates by $\approx 1/\lambda$, to the asymptotic decay of $I(\lambda)$ as $\lambda \to \infty$.

23.1 One Dimensional Theory

Let us begin with a simple example of an oscillatory integral, i.e.

$$I(\lambda) = \int_J e^{2\pi i\lambda\phi(x)}\,dx,$$

where J is a closed interval, and $\phi : J \to \mathbf{R}$ is Borel measurable. Taking in absolute values shows that $|I(\lambda)| \leqslant |J|$ for all λ. If ϕ is a constant, then $I(\lambda) = |J|e^{i\lambda\phi}$, so in this case the estimate is sharp. But if ϕ varies, we expect $I(\lambda)$ to decay as $\lambda \to \infty$. For instance, there are various results, such as the Esseén concentration inequality, which show that if we are to expect *average* decay in the integral I over a range of λ, then ϕ must not be concentrated around any point.

Theorem 23.1 (Esseén Concentration Inequality). *Let $\phi : J \to \mathbf{R}$ be Borel measurable, and for each $\lambda \in \mathbf{R}$, set*

$$I(\lambda) = \int_J e^{2\pi i\lambda\phi(x)}\,dx.$$

Then for any $\varepsilon > 0$,

$$\sup_{a\in\mathbf{R}} |\{x \in J : |\phi(x) - a| \leqslant \varepsilon\}| \lesssim \varepsilon \int_0^{1/\varepsilon} |I(\lambda)|\,d\lambda,$$

where the implicit constant is independant of ϕ.

Proof. By rescaling, we may assume that $J = [0,1]$. Moreover, for any choice of a, we may replace ϕ with $\phi - a$, reducing the analysis to the case where $a = 0$. Similarily, replacing ϕ with ϕ/ε reduces us to the situation where $\varepsilon = 1$. Thus we must show

$$|\{x \in [0,1] : |\phi(x)| \leqslant 1\}| \lesssim \int_0^1 |I(\lambda)|\, d\lambda,$$

where the implicit constant is independant of the function ϕ. If ψ is an integrable function supported on $[0,1]$, then Fubini's theorem implies

$$\begin{aligned}\int_0^1 \psi(\lambda) I(\lambda)\, d\lambda &= \int_0^1 \int_0^1 \psi(\lambda) e^{2\pi i \lambda \phi(x)}\, d\lambda\, dx\\ &= \int_0^1 \widehat{\psi}(-\phi(x))\, dx.\end{aligned}$$

In particular, this means that

$$\left| \int_0^1 \widehat{\psi}(-\phi(x))\, dx \right| \leqslant \|\psi\|_{L^\infty[0,1]} \int_0^1 |I(\lambda)|\, d\lambda.$$

If we choose a bounded function ψ such that $\widehat{\psi}$ is non-negative, and bounded below on $[-1,1]$, then

$$\left| \int_0^1 \widehat{\psi}(-\phi(x))\, dx \right| \gtrsim |\{x \in [0,1] : |\phi(x)| \leqslant 1\}|,$$

and so the claim follows easily. $\square$

Remark. Recently, I have seen results that *reverse* the concentration bound, for instance, a paper of Basu, Guo, Zhang, and Zorin-Kranich entitled "A Stationary Set Method For Estimating Oscillatory Integrals", and Wright's paper "A Theory of Complex Oscillatory Integrals: A Case Study". I do not know how general these techniques can be developed, but given an oscillatory integral in which we are able to prove stationary set estimates, but do not know asymptotics, it might be useful to look into these papers more.

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 varies significantly, then I exhibits cancellation as $\lambda \to \infty$. One way to quantify this rate is through the derivative of the function ϕ, i.e. if the derivative has a large magnitude, the phase is varying fast. A bound on $|\phi'|$ from below is not sufficient to guarantee cancellation independant of the function ϕ, as the next example shows, if the integrand oscillates at a wavelength $1/\lambda$.

Example. *Fix a positive integer λ_0, and let $\phi_{\lambda_0}(x) = x + f(\lambda_0 x)/\lambda_0$, where f is smooth and 1-periodic, $\|f'\|_{L^\infty(\mathbf{R})} \leqslant 1/2$, and*

$$\int_0^1 e^{2\pi i (x + f(x))}\, dx \neq 0.$$

Then for each $x \in \mathbf{R}$, $1/2 \leqslant |\phi'_{\lambda_0}(x)| \leqslant 2$, and in particular, is bounded independently of λ_0. Since $\phi_{\lambda_0}(x + 1/\lambda_0) = \phi_{\lambda_0}(x) + 1/\lambda_0$, we find $e^{2\pi i \lambda_0 \phi_{\lambda_0}(x)}$ is $1/\lambda_0$ periodic. In particular, this means

$$\begin{aligned} I(\lambda_0) &= \int_0^1 e^{2\pi i \lambda_0 \phi_{\lambda_0}(x)}\, dx \\ &= \lambda_0 \int_0^{1/\lambda_0} e^{2\pi i (\lambda_0 x + f(\lambda_0 x))}\, dx \\ &= \int_0^1 e^{2\pi i (x + f(x))}\, dx. \end{aligned}$$

Thus $I(\lambda_0; \phi_{\lambda_0}) \sim 1$, independant of ϕ_0, despite a uniform lower bound on the derivatives of the family of functiosn $\{\phi_{\lambda_0}\}$.

One option is a uniform upper bound on ϕ'', in addition to lower bounding ϕ', which eliminates the counterexample above, and yields a positive result.

Theorem 23.2. *Let $\phi : J \to \mathbf{R}$ be twice continuously differentiable, and suppose there exists constants $A, B > 0$ with $|\phi'(x)| \geqslant A$ and $|\phi''(x)| \leqslant B$ for all $x \in J$. Then for all $\lambda > 0$, we find*

$$|I(\lambda)| \lesssim \frac{1}{\lambda} \left(\frac{1}{A} + \frac{B}{A^2} |J| \right).$$

Proof. A dimensional analysis shows that the inequality is invariant under rescalings in x and λ, so we may assume that $J = [0,1]$, and $\lambda = 1$. An integration by parts shows that

$$\begin{aligned}\int_0^1 e^{2\pi i \phi(x)}\,dx &= \int_0^1 \frac{1}{2\pi i \phi'(x)} \frac{d}{dx}\left(e^{2\pi i \phi(x)}\right)\,dx \\ &= \left(\frac{e^{2\pi i \phi(1)}}{2\pi i \phi'(1)} - \frac{e^{2\pi i \phi(0)}}{2\pi i \phi'(0)}\right) - \int_0^1 \frac{d}{dx}\left(\frac{1}{2\pi i \phi'(x)}\right) e^{2\pi i \phi(x)}.\end{aligned}$$

Now

$$\frac{d}{dx}\left(\frac{1}{\phi'(x)}\right) = -\frac{\phi''(x)}{\phi'(x)^2},$$

so taking in absolute values to the tree quantities in the sum above completes the proof. $\square$

One can keep applying absolute values to obtain further bounds in terms of higher order derivatives of ϕ. For instance, another integration by parts shows that if there is $A, B, C > 0$ such that for $x \in J$, if $\phi'(x) \geqslant A$, $\phi''(x) \leqslant B$, and $\phi'''(x) \leqslant C$, then

$$|I(\lambda)| \lesssim \frac{1}{\lambda}\left(\frac{1}{A}\right) + \frac{1}{\lambda^2}\left(\frac{B}{A^3} + \frac{C}{A^3}|J| + \frac{B^2}{A^4}|J|\right).$$

One can keep taking in absolute values, but the $1/\lambda$ decay will still remain. This is to be expected; for instance, if $\phi(x) = x$, and J$= [0,1]$, then as $\lambda \to \infty$,

$$I(\lambda) = \int_0^1 e^{2\pi i \lambda x}\,dx = \frac{e^{2\pi i \lambda} - 1}{2\pi i \lambda}$$

and so

$$\limsup_{\lambda \to \infty} |I(\lambda) \cdot \lambda| = 2,$$

so we cannot obtain any better decay than $1/\lambda$ here.

Another option is to not require control on the second derivative of the phase, but instead to assume that ϕ' is monotone, which prevents the kind of oscillation present in our counterexample.

Lemma 23.3 (Van der Corput). *Let $\phi : \mathbf{R} \to \mathbf{R}$ be a smooth phase such that $|\phi'(x)| \geqslant A$ for all $x \in J$, and ϕ' is monotone. Then for all $\lambda > 0$ we have*

$$|I(\lambda)| \lesssim \frac{1}{A\lambda},$$

where the implicit constant is independent of J.

Proof. We may rescale so that $\lambda = 1$ and $J = [0,1]$. Then the same integration by parts shows that

$$\int_0^1 e^{2\pi i \phi(x)}\, dx = \left(\frac{e^{2\pi i \phi(b)}}{2\pi i \phi'(1)} - \frac{e^{2\pi i \phi(0)}}{2\pi i \phi'(0)} \right) + \frac{1}{2\pi i} \int_0^1 \frac{d}{dx} \left(\frac{1}{2\pi i \phi'(x)} \right) e^{2\pi i \phi(x)}\, dx.$$

The two boundary terms are $O(1/A)$. For the integral, we apply a simple trick. Since ϕ' is monotone, so too is $1/\phi'$, so in particular, it's derivative has a constant sign. Thus by the fundamental theorem of calculus,

$$\begin{aligned}
\left| \int_J \frac{d}{dx} \left(\frac{1}{2\pi i \phi'(x)} \right) e^{2\pi i \phi(x)}\, dx \right| &\leqslant \int_J \left| \frac{d}{dx} \left(\frac{1}{2\pi i \phi'(x)} \right) \right| dx \\
&= \left| \int_J \frac{d}{dx} \left(\frac{1}{2\pi i \phi'(x)} \right) \right| dx \\
&= \frac{1}{2\pi \phi'(b)} - \frac{1}{2\pi \phi'(a)}.
\end{aligned}$$

Again, this term is $O(1/A)$. □

Since the Van der Corput bound does not depend on $|J|$, it can be easily iterated using a decomposition strategy to give a theorem about higher derivatives of a function ϕ.

Lemma 23.4. *Let $\phi : \mathbf{R} \to \mathbf{R}$ be smooth, and suppose there is some $k \geqslant 2$ such that $|\phi^{(k)}(x)| \geqslant A$ for all $x \in J$. Then for all $\lambda > 0$, we find*

$$|I(\lambda)| \lesssim_k \frac{1}{(A\lambda)^{1/k}},$$

where the implicit constant is independant of J.

Proof. We perform an induction on k, the case $k = 1$ already proven. By scale invariance, we may assume $\lambda = 1$. Now $\phi^{(k-1)}$ is monotone, so for each $\alpha > 0$, outside an interval of length at most $O(\alpha/A)$, $|\phi^{(k-1)}(x)| \geqslant \alpha$. Thus applying the trivial bound in the excess region, and the case $k-1$ on the other intervals, we conclude

$$|I(\lambda)| \lesssim_k \frac{\alpha}{A} + \alpha^{-1/(k-1)}$$

Optimizing over α, we find $|I(\lambda)| \lesssim_k A^{-1/k}$. □

Remark. If $\phi^{(k-1)}$ vanishes at some point x_0 in J, then Taylor expansion shows that

$$|\{x \in J : |\phi(x) - \phi(x_0)| \leqslant \varepsilon\}| \gtrsim \varepsilon^{1/k}.$$

The Berry-Esseen theorem thus implies that the estimate $I(\lambda) \lesssim_k \lambda^{-1/k}$ is tight; we cannot have $I(\lambda) \lesssim \lambda^{-\alpha}$ for any $\alpha > 1/k$.

Let us now consider a one dimensional oscillatory integral with a varying amplitude function a, i.e.

$$I(\lambda) = \int_{-\infty}^{\infty} a(x) e^{2\pi i \lambda \phi(x)}\, dx.$$

The Van der Corput lemma also applies here.

Lemma 23.5. *Fix $k \geqslant 1$. Suppose a is supported on J, $|\phi^{(k)}(x)| \geqslant A$ for all $x \in J$, with ϕ' monotone if $k = 1$. Then*

$$|I(\lambda)| \lesssim_k \frac{\|a\|_{L^\infty(\mathbf{R})} + \|a'\|_{L^1(\mathbf{R})}}{(A\lambda)^{1/k}}.$$

Proof. Again, without loss of generality, we may assume $\lambda = 1$. Rescaling x means we can assume $J = [0,1]$. For $x \in [0,1]$, define

$$I_0(x) = \int_0^x e^{2\pi i \lambda \phi(t)}\, dt.$$

The standard Van-der Corput lemma implies that for all x,

$$|I_0(x)| \lesssim_k \frac{1}{(A\lambda)^{1/k}}.$$

Integrating by parts, we find that

$$\begin{aligned}\int_0^1 a(x)e^{2\pi i\lambda\phi(x)}\,dx &= \int_0^1 a(x)I_0'(x)\,dx\\ &= [a(1)I_0(1) - a(0)I_0(0)] - \int_0^1 a'(x)I_0(x)\,dx.\end{aligned}$$

Now

$$|a(1)I_0(b) - a(0)I_0(a)| \lesssim \frac{\|a\|_{L^\infty(\mathbf{R})}}{(A\lambda)^{1/k}}$$

and

$$\left|\int_0^1 a'(x)I_0(x)\,dx\right| \lesssim_k \frac{\|a'\|_{L^1(\mathbf{R})}}{(A\lambda)^{1/k}}.$$

Putting these two estimates together completes the proof. □

Remark. Since the bound doesn't depend on the interval, if a is not compactly supported, but a' is integrable, and the other assumptions of the last result holds, then we still have a bound

$$\sup_{a<b}\left|\int_a^b a(x)e^{2\pi i\lambda\phi(x)}\,dx\right| \lesssim_k \frac{\|a\|_{L^\infty(\mathbf{R})} + \|a'\|_{L^1(\mathbf{R})}}{(A\lambda)^{1/k}}$$

If a is smooth and compactly supported, repeated integration by parts is very successful because there are no boundary terms, and so we get very fast decay as $\lambda \to \infty$.

Theorem 23.6. *If a and ϕ are smooth functions, with a compactly supported, and $|\phi'(x)| \geqslant \delta$ for all x in the support of a, then for all $N > 0$,*

$$|I(\lambda)| \lesssim_{a,\phi,N} \delta^{-2N}\lambda^{-N}.$$

For a fixed N, the implicit constant depends only on the support of a, and L^∞ upper bounds on the derivatives of ϕ up to order $N+1$ (not including ϕ itself, which makes sense since the magnitude of the integral is invariant under translation), and upper bounds on the derivatives of a up to order N.

Proof. A single integration by parts gives

$$\begin{aligned} I(\lambda) &= \frac{1}{\lambda} \int \frac{a(x)}{2\pi i \phi'(x)} \frac{d}{dx} \left(e^{2\pi i \lambda \phi(x)} \right) \\ &= -\frac{1}{2\pi i \lambda} \int \frac{d}{dx} \left(\frac{a(x)}{\phi'(x)} \right) e^{2\pi i \lambda \phi(x)} \\ &= \frac{1}{2\pi i \lambda} \int \frac{\phi'(x) a'(x) - a(x) \phi''(x)}{\phi'(x)^2} e^{2\pi i \lambda \phi(x)}. \end{aligned}$$

Further integration by parts give, for each N, that

$$I(\lambda) = \lambda^{-N} \int \frac{P(x)}{\phi'(x)^{2N}} e^{2\pi i \lambda \phi(x)},$$

where $P(x)$ is a polynomial, with variables in the derivatives of ϕ up to order $N+1$, and derivatives in a up to order N, homogeneous of order N of the former variables, and linear in the variables of the latter. Thus we can take in absolute values and integrate to conclude $|I(\lambda)| \lesssim_{a,N} \lambda^{-N}$. $\square$

Remark. Differentiation under the integration sign shows

$$\left(\frac{d}{d\lambda} \right)^n I(\lambda) = (2\pi i)^n \int \phi(x)^n a(x) e^{2\pi i \lambda \phi(x)}\, dx.$$

Since $\phi^n a$ satisfies the same assumptions that a does, it follows that for any N, we have

$$\left| \left(\frac{d}{d\lambda} \right)^n I(\lambda) \right| \lesssim_{a,\phi,N,n} \lambda^{-N}.$$

Thus I is actually a Schwartz function on the real line.

Let us now move onto a 'stationary phase', i.e. a phase ϕ whose derivative vanishes at a point. The simplest example of such a phase is the integral

$$I(\lambda) = \int_{-\infty}^{\infty} a(x) e^{2\pi i \lambda x^2}\, dx,$$

where $a \in C_c^\infty(\mathbf{R})$. If a is nonzero in a neighborhood of the origin, and if $\phi(x) = x^2$, then

$$|\{x \in \mathbf{R} : |\phi(x)| \leqslant 1/\lambda\}|\} \lesssim \lambda^{-1/2}.$$

Thus our heuristics tell us that we should expect $I(\lambda)$ decays on the order of $\lambda^{-1/2}$, which agrees with the asymptotics we now find. A dyadic decomposition about the origin, combined with a single integration by parts at each scale yields a bound $|I(\lambda)| \lesssim \lambda^{-1/2}$.

Theorem 23.7. *Let $a \in \mathcal{S}(\mathbf{R})$. Then*

$$I(\lambda) \sim e^{i\pi/4}(2\lambda)^{-1/2} \sum_{n=0}^{\infty} (i/8\pi)^n \frac{a^{(2n)}(0)}{n!}$$

That is, for each pair of nonnegative integers N and k

$$\left(\frac{d}{d\lambda}\right)^k I(\lambda) = \left(\frac{d}{d\lambda}\right)^k \left\{ e^{i\pi/4} \cdot (2\lambda)^{-1/2} \sum_{n=0}^{N} (i/8\pi)^n \frac{a^{(2n)}(0)}{n!} \right\} + O_{N,m,a}(1/\lambda^{N+m+3/2}).$$

In particular, if a equals one near the origin, then $I(\lambda) \sim e^{i\pi/4}(2\lambda)^{-1/2}$.

Proof. Applying the multiplication formula for the Fourier transform, noting that the distributional Fourier transform of $e^{2\pi i\lambda x^2}$ is

$$(2\lambda)^{-1/2} e^{i\pi/4} e^{-i\pi\xi^2/2\lambda}$$

Thus

$$I(\lambda) = (2\lambda)^{-1/2} e^{i\pi/4} \int_{-\infty}^{\infty} e^{-i\pi\xi^2/2\lambda} \widehat{a}(\xi)\, d\xi.$$

Now for any N, we can write

$$e^{-i\pi\xi^2/2\lambda} = \sum_{n=0}^{N} \frac{1}{n!} \left(\frac{-i\pi\xi^2}{2\lambda}\right)^n + O_N\left((\xi^2/\lambda)^{N+1}\right)$$

Thus substituting in the Taylor series, and then applying the Fourier inversion formula, we find

$$\begin{aligned}
I(\lambda) &= (2\lambda)^{-1/2} e^{i\pi/4} \sum_{n=0}^{N} \frac{1}{n!} \int_{-\infty}^{\infty} \left(\frac{-i\pi\xi^2}{2\lambda}\right)^n \widehat{a}(\xi)\, d\xi + O_{a,N}\left(1/\lambda^{N+3/2}\right) \\
&= (2\lambda)^{-1/2} e^{i\pi/4} \sum_{n=0}^{N} (i/8\pi)^n \frac{1}{n!} \frac{1}{\lambda^n} \int_{-\infty}^{\infty} (2\pi i\xi)^{2n} \widehat{a}(\xi)\, d\xi + O_{a,N}\left(1/\lambda^{N+3/2}\right) \\
&= (2\lambda)^{-1/2} e^{i\pi/4} \sum_{n=0}^{N} (i/8\pi)^n \frac{a^{(2n)}(0)}{n!} \frac{1}{\lambda^n} + O_{a,N}\left(1/\lambda^{N+3/2}\right).
\end{aligned}$$

□

One obtains the asymptotic formula for the derivative of I by noting that

$$I^{(k)}(\lambda) = \int (-2\pi i x^2)^k a(x) e^{-2\pi i \lambda x^2}\, dx,$$

which reduces to the case where $k = 0$.

Remark. The implicit constant can be made independent of a given uniform upper bounds on

$$\int_{-\infty}^{\infty} |\widehat{a}(\xi)||\xi|^{2(N+m+1)}\, d\xi.$$

In particular, this can be obtained by uniform bounds on the support of a, upper bounds on the magnitude of a, and upper bounds on the magnitude of the $(2N+4)$th derivative of a.

It requires only a simple change of variables to extend this theorem to arbitrary quadratic phases. We say a critical point of a function is *non-degenerate* if the second derivative at that point is nonzero.

Theorem 23.8. *Let ϕ be a smooth phase with a single, non-degenerate critical point x_0, and let a be a smooth compactly supported amplitude function. Then there exists a sequence of constants $\{c_n\}$, depending solely on the derivatives of a and ϕ at x_0, such that*

$$I(\lambda) \sim e^{2\pi i \lambda \phi(x_0)} \lambda^{-1/2} \sum_{n=0}^{\infty} c_n \lambda^{-n}.$$

The implicit constants in the asymptotic formula are uniform over a family of amplitude functions a given uniform bounds on the support of a, and upper bounds on $2N$ derivatives of a. We can explicitly calculate that

$$c_0 = \sqrt{\frac{i}{\phi''(x_0)}} \cdot a(x_0).$$

More generally, in this case there exists a sequence of linear differential operators $\{L_n\}$, with coefficients depending on the derivatives of ϕ at x_0, such that $c_n = (L_n a)(x_0)$.

Proof. Translating our integration variables if necessary, we may assume that the critical point x_0 lies at the origin. Partitioning the support of a, applying the principle of nonstationary phase away from the critical point, we may assume without loss of generality that a is supported on an arbitrarily small neighborhood of the critical point. In particular, we may assume that $\phi(x) \neq 0$ for all nonzero x in the support of a. A coordinate change $x \mapsto -x$ means we may assume that $\phi''(0) > 0$. We can then define a function

$$y(x) = \operatorname{sgn}(x) \cdot \phi(x)^{1/2}.$$

It follows from our assumptions that y is a smooth diffeomorphism on the support of a. By the change of variables formula, there exists a smooth, compactly supported function $a_0(y)$ such that

$$I(\lambda) = \int a(x) e^{2\pi i \lambda \phi(x)}\, dx = \int a_0(y) e^{2\pi i \lambda y^2}\, dy.$$

Thus we can apply the previous theorem to conclude that there exists a sequence of constants $\{c_n\}$ such that for each N,

$$\left(\frac{d}{d\lambda} \right)^k I(\lambda) = \left(\frac{d}{d\lambda} \right)^k \left\{ \lambda^{-1/2} \sum_{n=0}^{N} c_n \lambda^{-n} \right\} + O_{\phi,a,N,k}(1/\lambda^{N+m+3/2}).$$

The existence in this theorem is a *constructive* existence statement. The proof gives an effective algorithm to produce as many constants c_n as required for any particular phase ϕ, provided one can explicitly write out the function $y(x)$. In particular, if the phase has only a single stationary point at the origin,

$$c_0 = 2^{-1/2} e^{i\pi/4} a_0(0) = 2^{-1/2} e^{i\pi/4} a(0)/y'(0).$$

Since

$$\begin{aligned} y'(0) &= \lim_{x \to 0^+} y'(x) = \lim_{x \to 0^+} \frac{\phi'(x)}{2\phi(x)^{1/2}} \\ &= \lim_{x \to 0^+} (1/2) \frac{\phi'(x)}{x} \left(\frac{x^2}{\phi(x)} \right)^{1/2} = (\phi''(0)/2)^{1/2}. \end{aligned}$$

This means that $c_0 = e^{i\pi/4} a(0) \phi''(0)^{-1/2}$. $\square$

If the phase ϕ has a critical point of order greater than two, than the asymptotics of the oscillatory integral get worse. In particular, if ϕ has a zero of order k, then around this region ϕ differs by $1/\lambda$ on an interval of length $1/\lambda^{1/k}$, so we might expect $I(\lambda)$ to be proportional to $\lambda^{1/k}$. This is precisely what happens, but our proof will not rely on the Fourier transform since the computation of the Fourier transform of $e^{\lambda i x^k}$ is quite difficult to calculate when $k > 2$. The next proof also works for the case $k = 2$, but the proof involves some contour shifting. Since the large majority of the examples we consider will have nondegenerate critical points (this is the generic behaviour of critical points), these complicated asymptotics can be safely skipped on a first reading.

Lemma 23.9. *For any non-negative integers l and k, there is a positive constant $A_{kl} > 0$ such that for any $\lambda \in \mathbf{R}$ and $\varepsilon > 0$,*

$$\int_0^\infty e^{2\pi i \lambda x^k} e^{-\varepsilon x^k} x^l \, dx = A_{kl} (\varepsilon - 2\pi i \lambda)^{-(l+1)/k},$$

where the kth root is the principal root for non-negative complex numbers.

Proof. If $z = (\varepsilon - 2\pi i \lambda)^{1/k} x$, and if α_N is the ray between the origin and the point $N(\varepsilon - 2\pi i \lambda)^{1/k}$, then

$$\int_0^N e^{2\pi i \lambda x^k} e^{-\varepsilon x^k} x^l \, dx = (\varepsilon - 2\pi i \lambda)^{-(l+1)/k} \int_{\alpha_N} e^{-z^k} z^l \, dz.$$

Let $\theta \in (-\pi/2, 0]$ be the argument of $(\varepsilon - 2\pi i \lambda)^{1/k}$, and set β_N to be the arc between $N(\varepsilon - 2\pi i \lambda)^{1/k}$ and $N(\varepsilon^2 + \lambda^2)^{1/2}$. Then β_N has length $O(N)$, with implicit constant depending on λ and ε. Moreover, any point z on β_N has modulus $N(\varepsilon^2 + \lambda^2)^{1/2}$ and argument less than or equal to θ/k. But this implies that $\text{Re}(z^k) \geqslant N^k (\varepsilon^2 + \lambda^2)^{k/2} \cos(\theta)$, and so there exists a constant c depending on ε and λ such that $|e^{-z^k}| \leqslant e^{cN^k}$. But this means that $|z^l e^{-z^k}| \leqslant N^l e^{-cN^k}$. Thus taking in absolute values gives that

$$\lim_{N \to \infty} \int_{\beta_N} e^{-z^k} z^l \, dz = 0.$$

In particular, applying Cauchy's theorem, we conclude that

$$\lim_{N \to \infty} \int_{\gamma_N} e^{-z^k} z^l \, dz = \int_0^\infty e^{-x^k} x^l \, dx.$$

If we denote the latter integral by $A_{kl} > 0$, then we have shown that

$$\int_0^\infty e^{2\pi i \lambda x^k} e^{-\varepsilon x^k} x^l \, dx = A_{kl} \cdot (\varepsilon - 2\pi i \lambda)^{-(l+1)/k},$$

as was required to be shown. $\square$

Remark. In particular, this implies that for each ε, there exists constants A_{kln} such that

$$\int_0^\infty e^{2\pi i \lambda x^k} e^{-x^k} x^l \, dx = (2\pi\lambda)^{-(l+1)/k} \sum_{n=0}^\infty A_{kln} (2\pi\lambda)^{-n}.$$

This is obtained by taking the Laurent series of

$$(1 - 2\pi i \lambda)^{-(l+1)/k} = (2\pi\lambda)^{-(l+1)/k} (\lambda^{-1} - 2\pi i)^{-(l+1)/k},$$

In particular, for each N and for each λ, we conclude

$$\int_0^\infty e^{2\pi i \lambda x^k} e^{-x^k} x^l \, dx = \lambda^{-(l+1)/k} \sum_{n=0}^N A_{kln} (2\pi\lambda)^{-n} + O_N\left(1/\lambda^{n+1+1/k}\right).$$

Lemma 23.10. *If η is compactly supported and smooth, then*

$$\left| \int_{-\infty}^\infty e^{2\pi i \lambda x^k} x^l \eta(x) \, dx \right| \lesssim_{l,k,\eta} \lambda^{-(l+1)/k}.$$

Proof. Let α be a bump function supported on $[-2,2]$ with $\alpha(x) = 1$ for $|x| \leqslant 1$. For each $\varepsilon > 0$, write

$$\begin{aligned}\int_{-\infty}^\infty e^{2\pi i \lambda x^k} x^l \eta(x) \, dx = &\int_{-\infty}^\infty e^{2\pi i \lambda x^k} x^l \eta(x) \alpha(x/\varepsilon) \, dx \\ &+ \int_{-\infty}^\infty e^{2\pi \lambda x^k} x^l \eta(x) (1 - \alpha(x/\varepsilon)) \, dx,\end{aligned}$$

where we will bound each term and optimize for a small ε. We trivially have

$$\left| \int_{-\infty}^\infty e^{2\pi i \lambda x^k} x^l \eta(x) \alpha(x/\varepsilon) \, dx \right| \lesssim_\eta \varepsilon^{l+1},$$

We apply an integration by parts to the second integral, noting that $e^{2\pi i\lambda x^k}$ is a fixed point of the differential operator

$$Df = \frac{1}{2\pi i\lambda k x^{k-1}} \frac{df}{dx}.$$

If we consider the differential operator

$$D^* g = \frac{d}{dx}\left(\frac{-f}{2\pi i\lambda k x^{k-1}}\right) = \left(\frac{i}{2\pi\lambda k}\right)\left(\frac{f'(x)}{x^{k-1}} - \frac{(k-1)f(x)}{x^k}\right),$$

then for any smooth f and compactly supported g,

$$\int_{-\infty}^{\infty} (Df)(x)g(x) = \int_{-\infty}^{\infty} f(x)(D^*g)(x).$$

In particular,

$$\begin{aligned}\int_{-\infty}^{\infty} e^{2\pi i\lambda x^k} x^l \eta(x)(1-\alpha(x/\varepsilon))\,dx &= \int_{-\infty}^{\infty} D^N(e^{2\pi i\lambda x^k})\, x^l\eta(x)(1-\alpha(x/\varepsilon))\,dx\\ &= \int_{-\infty}^{\infty} e^{2\pi i\lambda x^k}\, (D^*)^N\{x^l\eta(x)(1-\alpha(x/\varepsilon))\}\,dx.\end{aligned}$$

Write $g_N(x) = (D^*)^N\{x^l\eta(x)(1-\alpha(x/\varepsilon))\}$. Since $x^l\eta(x)(1-\alpha(x/\varepsilon))$ vanishes for $|x| \leqslant \varepsilon$, so too does $g_N(x)$. For $N \geqslant l/(k-1)$, and $|x| \geqslant \varepsilon$, we have

$$|g_N(x)| \lesssim_{N,\eta} \lambda^{-N}\varepsilon^{-N}|x|^{l-N(k-1)},$$

where the implicit constant depends on upper bounds for the derivatives of η of order to N. We can thus take in absolute values after integrating by parts to conclude that if $N > (l+1)/(k-1)$, then

$$\left|\int_{-\infty}^{\infty} e^{2\pi i\lambda x^k} x^l\eta(x)(1-\alpha(x/\varepsilon))\,dx\right| \lesssim_{N,\eta} \lambda^{-N}\varepsilon^{l+1-Nk}$$

Thus we can put the two bounds together to conclude that

$$\left|\int_{-\infty}^{\infty} e^{2\pi i\lambda x^k} x^l\eta(x)\,dx\right| \lesssim_{N,\eta} \varepsilon^{l+1} + \lambda^{-N}\varepsilon^{l+1-Nk}.$$

Picking $\varepsilon = \lambda^{-1/k}$ gives

$$\left|\int_{-\infty}^{\infty} e^{2\pi i\lambda x^k} x^l\eta(x)\,dx\right| \lesssim_{N,\psi} \lambda^{-(l+1)/k}. \qquad \square$$

But N was chosen depending only on k and l, so the implicit constants depend on the correct variables.

We can now prove the asymptotics for the model case $\phi(x) = x^k$.

Theorem 23.11. *Suppose a is a smooth compactly supported amplitude, and ϕ is a smooth phase with $\phi'(x) \neq 0$ on the support of a except at some point x_0, where $\phi'(x_0) = \cdots = \phi^{(k-1)}(x_0) = 0$, and $\phi^{(k)}(x_0) \neq 0$. Then there is a sequence $\{c_n\}$ such that*

$$I(\lambda) \sim \lambda^{-1/k} \sum_{n=0}^{\infty} c_n \lambda^{-n/k}.$$

Proof. Let us begin with the model case $\phi(x) = x^k$. Let $\tilde{a}$ be a bump function with $\tilde{a}(x) = 1$ for all x with $a(x) > 0$. Then

$$I(\lambda) = \int_{-\infty}^{\infty} e^{2\pi i \lambda x^k} e^{-x^k} [e^{x^k} a(x)] \tilde{a}(x)\, dx.$$

For each N, perform a Taylor expansion, writing

$$e^{x^k} a(x) = \sum_{n=0}^{N} c_n x^n + x^{N+1} R_N(x).$$

Thus if $P_N(x) = \sum_{n=0}^{N} c_n x^n$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} e^{2\pi i \lambda x^k} e^{-x^k} [e^{x^k} a(x)] \tilde{a}(x)\, dx \\
&\quad = \int_{-\infty}^{\infty} e^{2\pi i \lambda x^k} e^{-x^k} P_N(x)\, dx \\
&\quad + \int_{-\infty}^{\infty} e^{2\pi i \lambda x^k} e^{-x^k} P_N(x) (\tilde{a}(x) - 1)\, dx \\
&\quad + \int_{-\infty}^{\infty} e^{2\pi i \lambda x^k} e^{-x^k} x^{N+1} R_N(x) \tilde{a}(x)\, dx.
\end{aligned}$$

The first integral can be expanded in the required power series. The second integral, since it is supported away from the origin, is $O_M(\lambda^{-M})$ for any $M > 0$. And in the last lemma we showed the third integral is $O(\lambda^{-(N+2)/k})$, so combining these three terms gives the required result. The general case follows from a change of variables. $\square$

Remark. As we saw in the case $k = 2$, if k is even and n is odd then

$$\int_{-\infty}^{\infty} e^{\lambda i x^k} e^{-x^k} x^n = 0.$$

Thus we can actually improve the asymptotics to the existence of a sequence $\{c_n\}$ such that

$$I(\lambda) = \lambda^{-1/k} \sum_{n=0}^{N} c_n \lambda^{-2n/k} + O_{\phi,a,N}\left(1/\lambda^{(2N+3)/k}\right).$$

Let us now consider some examples of the method of stationary phase in one dimension.

Example. *The Bessel function of order m, denoted $J_m(r)$, is defined to be the oscillatory integral*

$$J_m(r) = \int_0^1 e^{2\pi i r \sin(2\pi\theta)} e^{-2\pi i m\theta} d\theta.$$

We want to use the method of stationary phase to determine the decay of $J_m(r)$ as $r \to \infty$. The amplitude is $a(\theta) = e^{-2\pi i m\theta}$, and the phase is $\phi(\theta) = \sin(2\pi\theta)$. We note that the phase $\phi(\theta) = \sin(2\pi\theta)$ is stationary when $\theta = 1/4$ and $\theta = 3/4$, and that these stationary points are nondegenerate. Thus we might expect $|J_m(r)| = O_m(r^{-1/2})$. More precisely, we write $a = a_1 + a_2 + a_3$, where a_1 is supported in a small neighbourhood of $1/4$, a_2 in a neighbourhood of $3/4$, and a_3 is supported away from $1/4$ and $3/4$. Then $I_{a_1}(\lambda)$ and $I_{a_2}(\lambda)$ are oscillatory integrals with a unique nondegenerate stationary point. In fact, we find that

$$I_{a_1}(r) = (1/2\pi) e^{2\pi i(r - 1/8 - m/4)} + O(r^{-3/2}).$$

$$I_{a_2}(r) = (1/2\pi) e^{-2\pi i(r - 1/8 - m/4)} + O(r^{-3/2}).$$

The integral $I_{a_3}(\lambda)$ can be shifted (using periodicity) into a compactly supported integral with smooth amplitude and no stationary points, and thus decays arbitrarily fast, i.e. $|I_{a_3}(\lambda)| = O(\lambda^{-3/2})$. Thus summing up these estimates gives

$$I_a(r) = (1/\pi)\cos(2\pi r - \pi/4 - \pi m/2) + O(r^{-3/2}).$$

Example. *Consider the Airy function*

$$Ai(x) = \int_{-\infty}^{\infty} e^{2\pi i(x\xi + \xi^3/3)}\, d\xi,$$

which arises as a solution to the differential equation $y'' = xy$. Again, this integral is not defined absolutely. Nonetheless, for large N, an application of the Van der Corput lemma implies that for any finite interval I containing only points x with $|x| \geqslant N$,

$$\int_I e^{2\pi i(x\xi + \xi^3/3)}\, d\xi = O(1/N),$$

where the implicit constant is independant of I. Thus we can interpret the integral as

$$\lim_{n\to\infty} \int_{a_n}^{b_n} e^{2\pi i(x\xi + \xi^3/3)}\, d\xi,$$

where $\{a_n\}$ and $\{b_n\}$ are any sequences with $a_n \to -\infty$, $b_n \to \infty$.

Now consider the phase $\phi(\xi) = x\xi + \xi^3/3$. Then $\phi'(\xi) = x + \xi^2$. When x is negative, there are two stationary points. Thus we can rescale the integral, writing $\nu = x^{-1/2}\xi$, so that for $x > 0$,

$$Ai(-x) = x^{1/2} \int_{-\infty}^{\infty} e^{2\pi i x^{3/2}(\nu^3/3 - \nu)}\, d\nu.$$

If we write $\phi_0(\nu) = \nu^3/3 - \nu$, then ϕ_0 has two stationary points, at $\nu = \pm 1$. These stationary points are non-degenerate, so if we write $1 = a_1 + a_2 + a_3 + a_4$, where a_1 equal to one in a neighbourhood of 1, a_2 equal to one in a neighbourhood of -1, and a_3 is supported in the region between -1 and 1, and a_4 vanishes in all such regions, then we decompose $Ai(-x)$ as $I_1 + I_2 + I_3 + I_4$. Now the principle of stationary phase tells us that

$$I_1 = (1/\sqrt{2}) e^{2\pi i(-2/3x^{3/2} + 1/8)} x^{-1/4} + O(x^{-7/4})$$
$$I_2 = (1/\sqrt{2}) e^{2\pi i(2/3x^{3/2} - 1/8)} x^{-1/4} + O(x^{-7/4})$$

Moreover, $I_3 = O_N(x^{-N})$ for all $N \geqslant 0$. It remains to show $I_4 = O(x^{-1})$. Indeed, an integration by parts shows that

$$\begin{aligned} I_4 &= x^{1/2} \int_{-\infty}^{\infty} e^{2\pi i x^{3/2}\phi_0(\nu)} a_4(\nu)\, d\nu \\ &= \frac{-1}{2\pi i x} \int_{-\infty}^{\infty} e^{2\pi i x^{3/2}\phi_0(\nu)} \frac{d}{d\nu}\left(\frac{a_4(\nu)}{\nu^2 - 1}\right) d\nu. \end{aligned}$$

Taking in absolute values shows $|I_4| \lesssim 1/x$. Thus as $x \to \infty$,

$$Ai(-x) = \sqrt{2}x^{-1/4}\cos((-4\pi/3)x^{3/2} + \pi/4) + O(1/x),$$

which gives the first order asymptotics of the integral.

On the other hand, let us consider large positive x. Then the phase ϕ has no critical points, and we therefore expect very fast decay. To achieve this decay, we employ a contour shift, replacing the oscillatory integral with a different oscillatory integral which has *a stationary point, so we can obtain asymptotics here. If we write $\phi(z) = xz + z^3/3$, then $\phi'(z) = 0$ when $z = \pm ix^{1/2}$. A simple contour shift argument to the line $\mathbf{R} + ix^{1/2}$ gives*

$$Ai(x) = \int_{-\infty}^{\infty} e^{2\pi i\phi(\xi + ix^{1/2})}\, d\xi = e^{-(4\pi/3)x^{3/2}} \int_{-\infty}^{\infty} e^{-2\pi\xi^2 x^{1/2}} e^{2\pi i\xi^3/3}\, d\xi.$$

We have

$$\int_{-\infty}^{\infty} e^{-2\pi\xi^2 x^{1/2}} e^{2\pi i\xi^3/3}\, d\xi \approx x^{-1/4} \int_{-\infty}^{\infty} e^{-2\pi\xi^2} e^{2\pi i x^{-3/4}\xi^3/3}\, d\xi.$$

Now a Taylor series shows

$$e^{2\pi i x^{-3/4}\xi^3/3} = 1 + O(x^{-3/4}\xi^3/3),$$

so, plugging in, we conclude

$$Ai(x) = 2^{-1/2}x^{-1/4}e^{-(4\pi/3)x^{3/2}} + O(x^{-3/4}e^{-(2/3)x^{3/2}}).$$

Thus Airy's function decreases exponentially as $x \to \infty$.

Example. *Let us consider the integral quantities*

$$\int_0^1 e^{2\pi ix\xi} e^{2\pi i/x} x^{-\gamma}\, dx$$

where to avoid technicalities we assume $0 \leqslant \gamma < 2$. These integral quantities are not defined absolutely, so we actually interpret this integral as

$$\lim_{\varepsilon \to 0} \int_0^1 e^{2\pi ix\xi} e^{2\pi i/x} x^{-\gamma}\, dx$$

If we write $\phi(x) = x\xi + 1/x$, then

$$\int_0^1 e^{2\pi i x\xi} e^{2\pi i/x} x^{-\gamma}\, dx = \int_0^1 e^{2\pi i\phi(x)} x^{-\gamma}\, dx.$$

For $0 < \varepsilon_1 < \varepsilon_2 < \varepsilon$, since $\phi'(x) = \xi - 1/x^2$, an easy integration by parts shows that for $\varepsilon \leqslant \xi^{-1/2}/2$,

$$\begin{aligned}\int_{\varepsilon_1}^{\varepsilon_2} e^{2\pi i\phi(x)} x^{-\gamma}\, dx &= \frac{1}{2\pi i\xi}\int_{\varepsilon_1}^{\varepsilon_2} \frac{d}{dx}\left(e^{2\pi i\phi(x)}\right)\frac{x^{2-\gamma}}{x^2 - 1/\xi}\, dx \\ &= \frac{-1}{2\pi i\xi}\int_{\varepsilon_1}^{\varepsilon_2} e^{2\pi i\phi(x)}\frac{d}{dx}\left(\frac{x^{2-\gamma}}{x^2 - 1/\xi}\right) + O(\varepsilon^{2-\gamma}) \qquad (23.1)\\ &= O(\varepsilon^{2-\gamma}),\end{aligned}$$

where the constant is independent of ξ. This implies the limit we study exists. We wish to prove an asymptotic formula for this integral as $\xi \to \infty$. If we write $\phi(x) = x\xi + 1/x$, then

$$\int_0^1 e^{2\pi i x\xi} e^{2\pi i/x} x^{-\gamma}\, dx = \int_0^1 e^{2\pi i\phi(x)} x^{-\gamma}.$$

Since ϕ has a nondegenerate stationary point when $x = \xi^{-1/2}$, our heuristics might suggest that if the phase and amplitude were smooth at the origin, then

$$\int_0^1 e^{2\pi i\phi(x)} x^{-\gamma} \approx (1/2)^{1/2} e^{4\pi i\xi^{1/2}} e^{-i\pi/4}\xi^{\gamma/2-3/4}.$$

We shall show that these heuristics continue to hold, up to an error of $O(\xi^{\gamma/2-1})$.

In an attempt to isolate the critical point, we split the interval $[0,1]$ into three parts, $[0, 0.5\xi^{-1/2}]$, $[0.5\xi^{-1/2}, 1.5\xi^{-1/2}]$, and $[1.5\xi^{-1/2}, 1]$, obtaining three integrals I_1, I_2, and I_3. The calculation (23.1) shows that $|I_1| \lesssim \xi^{\gamma/2-1}$, and thus is neglible to our asymptotic formula. To obtain a bound on I_3, we use the Van der Corput lemma, noting that $\phi'(x) = \xi - 1/x^2$ is monotone, and $|\phi'(x)| \gtrsim \xi$ for $x \geqslant 1.5\xi^{-1/2}$. Thus we find $|I_1| \lesssim \xi^{-1}$, and thus is also neglible to our formula. Thus we are left with the trick part of calculating I_2 accurately. It will be easiest to do this by renormalizing the integral, i.e. writing $y = \xi^{1/2}x$, and calculating

$$I_2 = \int_{0.5\xi^{-1/2}}^{1.5\xi^{-1/2}} e^{2\pi i\phi(x)} x^{-\gamma}\, dx = \xi^{\gamma/2-1/2}\int_{0.5}^{1.5} e^{2\pi i\xi^{1/2}(y+1/y)} y^{-\gamma}\, dy.$$

We consider a smooth amplitude function $\psi(x)$ supported on the interior of $[0.5, 1.5]$. Then since $y + 1/y$ is stationary at $y = 1$, but non-degenerate, we can write

$$\int e^{2\pi i \xi^{1/2}(y+1/y)} y^{-\gamma} \psi(y)\, dy = \xi^{-1/4} \pi^{1/2} e^{2\pi i (2\xi^{1/2}+1/4)} + O(\xi^{-1/2}),$$

from which we obtain our main term. On the other hand, we can apply the Van der Corput lemma to show that

$$\int_{0.5}^{1.5} e^{2\pi i \xi^{1/2}(y+1/y)} y^{-\gamma} (1-\psi(y))\, dy = \int e^{2\pi i \xi^{1/2}(y+1/y)} y^{-\gamma} \psi(y)\, dy = O(\xi^{-1/2}).$$

Combining all these estimates gives the theorem.

On the other hand, consider the integral

$$I(\xi) = \int_0^1 e^{-2\pi i \xi x} e^{2\pi i/x} x^{-\gamma}\, dx = \int_0^1 e^{2\pi i \phi(x)} x^{-\gamma},$$

where $\phi(x) = 1/x - \xi x$ is the phase. Then the phase has no critical points so we can assume that we can large decay for large ξ. We decompose the integral onto the intervals $[0, \xi^{-1/2}]$ and $[\xi^{-1/2}, 1]$, inducing the two quantities I_1 and I_2. Now applying the Van der Corput lemma to I_2 with $|\phi'(x)| = |1/x^2 + \xi| \geqslant \xi$ for $x \geqslant 0$, gives $|I_2| \lesssim \xi^{\gamma/2-1}$. On the other hand, renormalizing with $y = \xi^{1/2} x$, we have

$$I_1 = \xi^{\gamma/2-1/2} \int_0^1 e^{2\pi i \xi^{1/2}(1/y-y)} y^{-\gamma}\, dy.$$

For each n, we note that for the phase $\phi_0(x) = 1/y - y$, for $1/2^{n+1} \leqslant y \leqslant 1/2^n$, we have $|\phi_0'(x)| \gtrsim 4^n$. Thus we can apply the Van der Corput lemma to conclude

$$\left| \int_{1/2^{n+1}}^{1/2^n} e^{2\pi i \phi_0(x)} y^{-\gamma}\, dy \right| \lesssim \frac{2^{\gamma n}}{4^n \xi^{1/2}}.$$

Summing up over all $n \geqslant 0$, we conclude $|I_1| \lesssim \xi^{\gamma/2-1}$. Thus $|I(\xi)| \lesssim \xi^{\gamma/2-1}$.

One way to interpret this asymptotic formula is through a Riemann singularity, *i.e. a tempered distribution Λ supported on the half-life $x \geqslant 0$, that agrees with the oscillatory function $e^{2\pi i/x} x^{-\gamma}$ for small x, but is compactly supported and smooth away from the origin. We consider the case $0 \leqslant \gamma < 2$ for simplicity. Thus for Schwartz $f \in \mathcal{S}(\mathbf{R})$, we have*

$$\Lambda(f) = \lim_{\varepsilon \to 0^+} \int_\varepsilon^\infty f(x) e^{2\pi i/x} x^{-\gamma} \psi(x)\, dx,$$

where ψ is smooth and compactly supported, and equals one in a neighbourhood of the origin. An easy integration by parts shows that for a fixed Schwartz f, and for $0 < \varepsilon_1 < \varepsilon_2 < \varepsilon$,

$$\left| \int_{\varepsilon_1}^{\varepsilon_2} f(x) e^{2\pi i/x} x^{-\gamma} \, dx \right| = O\left(\varepsilon^{2-\gamma}\right),$$

where the implicit constants depend on upper bounds for f and f' in a neighbourhood of the origin. Thus we find $\Lambda(f)$ is well defined, and moreover, Λ is a distribution of order one. Since Λ is compactly supported, the Paley-Weiner theorem implies that $\widehat{\Lambda}$ is a distribution represented by a locally integrable function, and

$$\widehat{\Lambda}(\xi) = \int_0^\infty e^{2\pi i/x} x^{-\gamma} e^{-2\pi \xi i x} \, dx.$$

The calculations above give asymptotic formulas for this Fourier transform. In particular, we see that the Fourier transform of Λ decays much faster to the right than to the left.

23.2 Stationary Phase in Multiple Variables

When we move from a single variable oscillatory integral to a multivariable oscillatory integrals. Thus we consider the oscillatory integral

$$I(\lambda) = \int_{\mathbf{R}^d} a(x) e^{2\pi i \lambda \phi(x)} \, dx.$$

for large λ. The method of stationary phase becomes significantly more complicated in this setting because the stationary points of the phase function are no longer necessarily isolated. In certain basic situations, such as when the stationary points are isolated and satisfy a nondegeneracy condition, we can obtain asymptotic formulae.

Theorem 23.12. *Let ϕ and a be smooth functions on $\mathbf{R}^d$, with a compactly supported. If $\nabla\phi$ is nowhere vanishing on the support of a, then for each $N > 0$, $|I(\lambda)| \lesssim_N \lambda^{-N}$ for all N.*

Proof. Set $v = (\nabla\phi)/|\nabla\phi|^2$. Note that the phase ϕ is an eigenfunction of the differential operator D defined such that

$$Df(x) = \frac{v \cdot \nabla f}{2\pi i \lambda}.$$

The adjoint operator of D is the operator D^* defined by setting

$$D^* f(x) = \frac{\nabla \cdot (vf)}{-2\pi i \lambda},$$

i.e. for any smooth f and g, with one of these functions compactly supported,

$$\int Df(x) g(x)\, dx = \int f(x)(D^* g)(x)\, dx.$$

Thus

$$I(\lambda) = \int D^N (e^{2\pi i \lambda \phi(x)}) a(x)\, dx = \int e^{2\pi i \lambda \phi(x)} ((D^*)^N a)(x)\, dx.$$

Taking absolute values in the last integral gives that

$$|I(\lambda)| \leqslant \int |(D^*)^N a(x)|\, dx \lesssim_{\phi, a, N} \frac{1}{\lambda^N}. \qquad \square$$

Remark. The implicit constants for a fixed N can be uniformly bounded given a uniform lower bound on $|\nabla \phi|$, and upper bounds on the derivatives of ϕ up to order $N+1$, on a up to order N, and on the measure of the support of a, just as in the one dimensional case.

A tensorization argument establishes stationary phase asymptotics for a quadratic phase.

Theorem 23.13. *Let A be an invertible $d \times d$ matrix, fix $x_0 \in \mathbf{R}^d$, and consider the phase $\phi(x) = A(x - x_0) \cdot (x - x_0)$. Then for any compactly supported smooth amplitude a, there exists constants $\{c_n\}$ depending only on the derivatives of a at the origin, such that*

$$I(\lambda) \sim \lambda^{-d/2} \sum_{n=0}^{\infty} c_n \lambda^{-n}.$$

Moreover,

$$c_0 = a(x_0) \prod_{k=1}^{d} (i/\mu_k)^{1/2},$$

where $\mu_1, \ldots, \mu_d$ are the eigenvalues of A.

Proof. Suppose first that a is a tensor product of d compactly supported functions in $\mathbf{R}$. The constant c_0 is invariant under affine changes of coordinates. Thus we may assume that A is a diagonal matrix. But then the oscillatory integral splits into the product of single variable integrals, to which we can apply our one-dimensional asymptotics. Since the asymptotics here depend only on the support of a, and upper bounds on the magnitude of a on derivatives up to order $2N + 4$. A density argument then shows the argument generalizes to any smooth a, with implicit constants depending on upper bounds on the measure of the support of a, and upper bounds on the derivative of a of order up to $2N + (d + 4)$. $\square$

Morse's theorem says that if x_0 is a non-degenerate critical point of a smooth function ϕ, then there exists a coordinate system around x_0 and $a_1, \dots, a_d \in \{\pm 1\}$ such that, in this coordinate system,

$$\phi(x_0 + t) = a_1 t_1^2 + \dots + a_d t_d^2.$$

In one dimension, the same is true if x_0 has a higher order critical point, but this does not generalize to higher dimensions, which reflects the lack of as nice a theory in this case. But in the case of functions with finitely many non-degenerate critical points, we can obtain nice asymptotics. Applying Morse's theorem gives the following theorem.

Theorem 23.14. *Let ϕ and a be smooth functions, with a compactly supported. Suppose ϕ has a single critical point x_0 on the support of a, which is nondegenerate. Then there exists constants $\{c_n\}$ depending only on finitely many derivatives of Φ and ψ at x_0, such that*

$$I(\lambda) \sim e^{2\pi i \lambda \phi(x_0)} \lambda^{-d/2} \sum_{n=0}^{\infty} c_n \lambda^{-n}.$$

Moreover,

$$c_0 = a(x_0) \cdot \prod_{k=1}^{d} (i/\mu_k)^{1/2},$$

where $\mu_1, \dots, \mu_d$ are the Eigenvalues of the Hessian of ϕ at x_0.

23.3 Variable Coefficient Results

We can also obtain results given amplitude functions that depend on λ, and also vary in a third variable, under suitable conditions. Consider a *symbol* $a(x,y,\lambda)$, i.e. a smooth function such that $\text{supp}_y(a)$ is compact, and

$$\left|\nabla_x^n \nabla_y^m \nabla_\lambda^k a(x,y,\lambda)\right| \lesssim_{n,m,k} \lambda^{\alpha+\delta k},$$

for some $\alpha \in \mathbf{R}$ and $\delta > 0$, where the implicit constant is *locally uniform* in x. We can then consider the oscillatory integrals

$$I(x,\lambda) = \int a(x,y,\lambda) e^{2\pi i \lambda \phi(x,y)}\, dy.$$

Provided that $\delta < 1$, and $\nabla_y \phi \neq 0$ on the support of a, integration by parts gives that for any n and N,

$$\nabla_x^n \nabla_\lambda^m I(x,\lambda) \lesssim_{N,n,m} \lambda^{-N},$$

where the implicit constant is locally uniform in x.

On the other hand, suppose that we have a nondegenerate critical point, i.e. there exists (x_0,y_0) with $\nabla_y \phi(x_0,y_0) = 0$, but with $H_y\phi(x_0,y_0)$ invertible. Then by the implicit function theorem, there exists a unique solution $(x,y(x))$ to $\nabla_y \phi(x,y)$ in a small neighborhood of y_0, for x in a small neighborhood of x_0. Intuitively, if a is supported on the small neighborhood of x_0 and y_0, we should have

$$\begin{aligned} I(x,\lambda) &= \int a(x,y,\lambda) e^{2\pi i \lambda \phi(x,y)}\, dy \\ &\approx \lambda^{-d/2} e^{2\pi i \lambda \phi(x,y(x))} a(x,y(x),\lambda) \prod_{k=1}^d (i/\mu_k)^{1/2}, \end{aligned}$$

where $\mu_1,\dots,\mu_d$ are the eigenvalues of $H_y\phi(x,y(x))$. Ignoring the $e^{2\pi i \lambda \phi(x,y(x))}$ factor (which introduces powers of λ), differentiating by x should introduce negligible effects on the value of the oscillatory integral. In fact, it is not difficult to prove that if

$$\left|\nabla_x^n \nabla_y^m \nabla_\lambda^k a(x,y,\lambda)\right| \lesssim_{n,m,k} \lambda^{-k},$$

then, under the support assumptions on a,

$$\nabla_x^n \nabla_\lambda^m \left\{ e^{-2\pi i \lambda \phi(x,y(x))} I(x,\lambda) \right\} \lesssim_{N,n,m} \lambda^{-1/2-m}.$$

The result is a simple calculation, left as an exercise. A discussion can be found in Sogge's book 'Fourier Integrals in Classical Analysis'. On the other hand, if we have bounds of the form

$$\left| \nabla_x^n \nabla_y^m \nabla_\lambda^k a(x,y,\lambda) \right| \lesssim_{n,m,k} \lambda^{\alpha+\delta k},$$

for $\delta < 1/2$, then we can still obtain an asymptotic expansion of the form

$$I(x,\lambda) \sim e^{2\pi i \lambda \phi(x,y(x))} \lambda^{-d/2} \sum_{n=0}^\infty c_n(x,\lambda) \lambda^{-n},$$

locally uniform in x, where the coefficients c_n are symbols of order TODO, equal to a linear differential operator applied to a at $(x,y(x),\lambda)$, with coefficients depending solely on the behaviour of the Hessian matrix of ϕ in y, at $(x,y(x))$. In particular,

$$c_0 = a(x,y(x),\lambda) \prod_{k=1}^d (i/\mu_k(x))^{1/2},$$

where $\mu_1(x),\dots,\mu_d(x)$ are the eigenvalues of the Hessian of ϕ at $(x,y(x))$. Duistermaat's book on Fourier Integral Operators includes an explicit description of these differential operators.

23.4 Surface Carried Measures

Let us consider oscillatory integrals on a 'curved' version of Euclidean space. One most basic example is the Fourier transform of the surface measure of the sphere, i.e.

$$\widehat{\sigma}(\xi) = \int_{S^{d-1}} e^{-2\pi i \xi x} d\sigma(x).$$

Studying the decay of this surface measure is of much interest to many problems in analysis. One can reduce the study of this Fourier transform to the study of Bessel functions, to which we have already developed an asymptotic theory.

Theorem 23.15. *If σ is the surface measure on the sphere S^{d-1}, then*

$$\widehat{\sigma}(\xi) = \frac{2\pi \cdot J_{d/2-1}(2\pi|\xi|)}{|\xi|^{d/2-1}}.$$

In particular,

$$\widehat{\sigma}(\xi) = \frac{2\cos(2\pi|\xi| - (d/2-1)(\pi/2) - \pi/4)}{|\xi|^{(d-1)/2}} + O_d(1/|\xi|^{(d+1)/2}).$$

Proof. Since σ is rotationally symmetric, so too is $\widehat{\sigma}$. In particular, we can apply Fubini's theorem to conclude that if V_{d-2} is the surface area of the unit sphere in $\mathbf{R}^{d-2}$, then

$$\begin{aligned}\widehat{\sigma}(\xi) &= \int_{S^{d-1}} e^{-2\pi|\xi|x_1} d\sigma(x)\\ &= V_{d-2} \int_{-1}^{1} e^{-2\pi|\xi|t}(1-t^2)^{d/2-1} dt.\end{aligned}$$

Setting $r = 2\pi|\xi|$ completes the argument. $\square$

Since the multivariate stationary phase approach is essentially 'coordinate independant', we can also generalize the approach to manifolds. If M is a d dimensional Riemmannian manifold, and ϕ and a are complex-valued functions on the manifold, we can consider the oscillatory integral

$$I(\lambda) = \int_M a(x) e^{2\pi i \lambda \phi(x)} d\sigma(x),$$

where σ is the surface measure induced by the metric on M. If ϕ and a are compactly supported, then this integral is well defined in the Lebesgue sense.

Theorem 23.16. *Suppose that a is a compactly supported smooth amplitude on a Riemannian manifold M, ϕ is a smooth phase, and $\nabla\phi$ vanishes at a single critical point x_0 on the support of ψ, upon each of which the Hessian $H\phi$ is non-degenerate at each point. Then there exists constants $\{c_n\}$ such that*

$$I(\lambda) \sim e^{2\pi i \lambda \phi(x_0)} \lambda^{-d/2} \sum_{n=0}^{\infty} c_n \lambda^{-n} + O(1/\lambda^{N+d/2+1}).$$

Moreover,

$$c_0 = a(x_0) \prod_{k=1}^{d} (i/\mu_k)^{1/2},$$

where $\mu_1, \dots, \mu_d$ *are the eigenvalues of the Hessian* $H\phi$ *at* x_0.

The theorem is proved by a simple partition of unity approach which reduces to the Euclidean case. It has the following important corollary.

Theorem 23.17. *If a d dimensional surface* Σ *is a smooth submanifold of* $\mathbf{R}^{d+1}$, *and has non-vanishing Gauss curvature, and if a is a smooth, compactly supported function on* Σ, *then*

$$|\widehat{a\sigma}(\xi)| \lesssim_{a,\sigma} \frac{1}{|\xi|^{d/2}},$$

where σ *is the surface measure of* Σ.

Proof. For each $\xi \in S^d$,

$$\widehat{a\sigma}(\xi) = \int_M a(x) e^{-2\pi i \xi \cdot x} \, d\sigma(x)$$

If $\lambda = |\xi|$, for $|\xi| = 1$, we can write

$$\widehat{a\sigma}(\lambda\xi) = \int_M a(x) e^{2\pi i \lambda \phi(\xi,x)} \, d\sigma(x),$$

where $\phi(\xi, x) = -\xi \cdot x$. Then the gradient of ϕ on the surface M vanishes at a point (ξ, x) precisely when ξ is normal to Σ. There are precisely two different smooth unit normal vector fields n_1, n_2 defined in a neighborhood of any given point x_0.

We can find two smooth vector fields n_1 and n_2

$$I_\xi(\lambda) = \int_M a(x) e^{2\pi i \lambda \phi(\xi,x)} \, d\sigma,$$

where $\phi(\xi, x) = -2\pi i \xi \cdot x$. The derivatives of ϕ_ξ of order $\leqslant N$ on M are $O_N(1)$, independently of ξ. Similarly, $H_M \phi_\xi$ is uniformly non-degenerate, in the sense that the operator norm of $(H_M \phi_\xi)^{-1}(x)$ is $O(1)$, independently of ξ. Working with Σ as a local graph, and then applying the curvature condition on Σ implies that for each $\xi \in S^d$, ϕ_ξ has $O(1)$ stationary points on the support of ψ. There also exists a constant r such that if x does not lie in any ball of radius r around a stationary point, then $|\nabla_M \phi_\xi| \gtrsim 1$. Moreover, the Hessian $H_M \phi_\xi$ is uniformly non-degenerate in the radius r balls around the critical point, independently of ξ. Thus we can apply the last result to conclude

$$I_\xi(\lambda) \lesssim \lambda^{-d/2},$$

where the implicit constant is independent of ξ, because all the required derivatives are uniformly bounded. □

If Ω is a bounded open subset of $\mathbf{R}^{d+1}$ whose boundary is a smooth manifold with non-zero Gaussian curvature at each point, then it's Fourier transform has decay one order better than the Fourier transform of it's boundary.

Corollary 23.18. *If Ω is a bounded open subset of $\mathbf{R}^d$ whose boundary is a smooth manifold Σ with non-zero Gaussian curvature at each point. If I_Ω is the indicator function on Ω, then*

$$|\widehat{I_\Omega}(\xi)| \lesssim_\Omega |\xi|^{-(d+1)/2}.$$

Proof. We have

$$\widehat{I_\Omega}(\xi) = \int_\Omega e^{-2\pi i \xi \cdot x}\, dx$$

Then we can apply Stoke's theorem for each $1 \leqslant k \leqslant d+1$ to conclude

$$\int_\Omega e^{-2\pi i \xi \cdot x}\, dx = \frac{(-1)^k}{2\pi i \xi_k} \int_\Sigma e^{-2\pi i \xi \cdot x} (dx^1 \wedge \dots \wedge \widehat{dx^k} \wedge \dots \wedge dx^n).$$

For each k, there is a smooth function ψ_k such that

$$dx^1 \wedge \dots \wedge \widehat{dx^k} \wedge \dots \wedge dx^n = \psi_k d\sigma.$$

Thus applying the last case, we find

$$\left| \frac{(-1)^k}{2\pi i \xi_k} \int_\Sigma e^{-2\pi i \xi \cdot x} (dx^1 \wedge \dots \wedge \widehat{dx^k} \wedge \dots \wedge dx^n) \right| \lesssim \xi_k^{-1} |\xi|^{-(d-1)/2}.$$

At each point ξ, if we choose ξ_k with the largest value, then $|\xi_k| \sim |\xi|$, so

$$\left| \int_\Omega e^{-2\pi i \xi \cdot x}\, dx \right| \lesssim |\xi|^{-(d+1)/2}.$$

□

The fact that curved surfaces have Fourier decay is of supreme importance to much of modern harmonic analysis. Let us see some basic consequences, which is a classical result due to Hardy, Littlewood, and Hlawka.

Theorem 23.19. *Let B be the unit ball in $\mathbf{R}^d$, and for each $\lambda > 0$, let $N(\lambda)$ denote the number of integer lattice points in λB. Then*

$$N(\lambda) = |B| \cdot \lambda^d + O(\lambda^{n-2+2/(n+1)}).$$

Proof. We have

$$N(\lambda) = \widehat{f}(0),$$

where $f = \mathbf{I}_{\lambda B} \cdot \sum_{n \in \mathbf{Z}^d} \delta_n$. The Poisson summation formula implies that

$$\widehat{f} = \widehat{\mathbf{I}_{\lambda B}} * (\sum_{n \in \mathbf{Z}^d} \delta_n)$$

Thus

$$\widehat{f}(0) = \sum_{n \in \mathbf{Z}^d} \widehat{\mathbf{I}_{\lambda B}}(n) = \lambda^d \sum_{n \in \mathbf{Z}^d} \widehat{\mathbf{I}}_B(\lambda n) = |B|\lambda^d + \sum_{n \neq 0} \widehat{\mathbf{I}}_B(\lambda n).$$

We want to apply the estimate $\widehat{\mathbf{I}}_B(\lambda n) \lesssim (\lambda n)^{-(d+1)/2}$, but the decay here isn't enough for the sum in the Poisson summation formula to converge. To fix this, we mollify $\mathbf{I}_B$ slightly. Let β be a non-negative bump function supported on $|x| \leqslant 1/2$ with $\int \beta(x)\, dx = 1$. Fix $\varepsilon > 0$, and define

$$\chi_\lambda(x) = \varepsilon^{-d} \mathrm{Dil}_\varepsilon \beta * \mathbf{I}_{\lambda B}.$$

Then set $\widetilde{N}(\lambda) = \sum_j \chi_\lambda(j)$. Then $\mathbf{I}_{\lambda B}$ is equal to χ_λ everywhere except on $\lambda - \varepsilon \leqslant |x| \leqslant \lambda + \varepsilon$, and since both functions are positive,

$$\widetilde{N}(\lambda - \varepsilon) \leqslant N(\lambda) \leqslant \widetilde{N}(\lambda + \varepsilon).$$

We can apply the Poisson summation formula now as above to obtain that

$$\widetilde{N}(\lambda) = \lambda^d |B| + \sum_{n \neq 0} \widehat{\mathbf{I}}_B(\lambda n) \widehat{\beta}(\varepsilon n).$$

The first term is bounded by $(\lambda n)^{-(d+1)/2}$, and the second by $(\varepsilon n)^{-K}$, where $n \geqslant 1/\varepsilon$ and K is large, and bounded by $O(1)$ for $n \leqslant 1/\varepsilon$. Applying these bounds thus gives

$$\widetilde{N}(\lambda) = \lambda^d |B| + O(\lambda^{-(d+1)/2} \varepsilon^{-(d-1)/2}).$$

Thus

$$N(\lambda) = \lambda^d |B| + O(\varepsilon\lambda^{d-1} + \lambda^{(d-1)/2}\varepsilon^{-(d-1)/2})$$

Choosing $\varepsilon = \lambda^{-(d-1)/(d+1)}$ completes the proof. □

Remark. One can never have $N(\lambda) = |B|\lambda^d + O(\lambda^{d-2})$, because TODO

Example. *If M is a hypersurface in $\mathbf{R}^d$, and ψ is a smooth, compactly supported function on M, and f is a smooth, compactly supported function on $\mathbf{R}^d$, we can define a function Af on $\mathbf{R}^d$ by defining*

$$(Af)(y) = \int_M f(y-x)\psi(x)\, d\sigma(x).$$

We note that Af is really the convolution of f with $\psi\sigma$. Thus

$$\widehat{Af}(\xi) = \widehat{f}(\xi)\widehat{\psi d\sigma}(\xi).$$

For each multi-index α, the derivative $(Af)_\alpha$ is equal to

$$\int_M f_\alpha(y-x)\psi(x)\, d\sigma(x) = f_\alpha * (\psi\sigma).$$

In particular, we have

$$\widehat{(Af)_\alpha} = (2\pi i\xi)^\alpha \widehat{f}(\xi)\widehat{\psi\sigma}(\xi).$$

Since we have shown

$$|\widehat{\psi\sigma}(\xi)| \lesssim |\xi|^{-(d-1)/2},$$

we conclude that if $|\alpha| \leqslant k$, where $k = (d-1)/2$,

$$\|(Af)_\alpha\|_{L^2(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^d)}.$$

In particular, this implies that A extends to a unique bounded operator from $L^2(\mathbf{R}^d)$ to $L^2_k(\mathbf{R}^d)$, i.e. to a map such that for each $f \in L^2(\mathbf{R}^d)$, Af is a square integrable function which has square integrable weak derivatives of all orders less than or equal to k, and moreover, $\|(Af)_\alpha\|_{L^2(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^d)}$ for all $|\alpha| \leqslant k$. Thus the operator A is 'smoothening', in a certain sense.

The operator A is obviously bounded from $L^1(\mathbf{R}^d)$ to $L^1(\mathbf{R}^d)$ and $L^\infty(\mathbf{R}^d)$ to $L^\infty(\mathbf{R}^d)$, purely from the fact that $\psi\sigma$ is a finite measure. Using curvature and some analytic interpolation, we will now also show that A is bounded *from $L^p(\mathbf{R}^d)$ to $L^q(\mathbf{R}^d)$, where $p = (d+1)/d$, and $q = d+1$. Interpolation thus yields a number of intermediate estimates. The trick here is to obtain an (L^1, L^∞) bound for an 'improved' version of A, and an (L^2, L^2) bound for a 'worsened' version of A. Interpolating between these two results gives a bound for precisely A. It suffices to prove this bound 'locally' on M, since we can then sum up these bounds, so we may assume that M is given as the graph of some function, i.e. there exists u such that*

$$M = \{(x, u(x)) :\}$$

For each s, we write $A_s f = K_s * f$, where

$$K_s(x) = \gamma_s |x_d - \phi(x')|_+^{s-1} \psi_0(x).$$

Here $\gamma_s = s \dots (s+N)e^{s^2}$, where N is some large parameter to be fixed in a moment. The e^{s^2} parameter is to mitigate the growth of γ_s as $|Im(s)| \to \infty$, which allows us to interpolate. The quantity $|u|_+^{s-1}$ is equal to u^{s-1} where $u > 0$, and is equal to 0 when $u \leqslant 0$. And $\psi_0(x) = \psi(x)(1 + |\nabla_{x'}\phi(x')|^2)^{1/2}$.

23.5 Oscillatory Integral Operators

An oscillatory integral takes as input a particular frequency scale, and gives back a scalar value. The goal of the theory of stationary phase is to determine the decay of the outputs as we increase the value of the frequency scale. Oscillatory integral operators are a generalization of this property. Most often, these oscillatory operators are a family of operators $\{T_\lambda\}$, where λ parameterizes the frequency scale of some oscillatory factor. The goal is to understand how the properties of these operators is affected asymptotically as we increase the frequency factor.

Let us begin with a family of such operators given by the formula

$$T_\lambda f(x) = \int_{\mathbf{R}^n} a(x,y) e^{2\pi i \lambda \phi(x,y)} f(y)\, dy,$$

where $a \in C_c^\infty(\mathbf{R}^m \times \mathbf{R}^n)$ is real-valued and $\phi \in C^\infty(\mathbf{R}^m \times \mathbf{R}^n)$. The regularity of all these parameters means that the operators T_λ are defined even for a general distributional input f. For intuition, we recall the theory of oscillatory integral distributions, and in particular, canonical relations and wave front sets. The kernel $K_\lambda(x,y) = a(x,y)e^{2\pi i\lambda\phi(x,y)}$ is an 'oscillatory integral distribution' of 'dimension zero'. Applying the theory of wave front sets for such distributions blindly, we obtain the formula

$$\mathrm{WF}(K_\lambda) \subset \{(x,y;\lambda\nabla_x\phi(x,y),\lambda\nabla_y\phi(x,y)) : x \in \mathbf{R}^m, y \in \mathbf{R}^n\}.$$

Since K_λ is a smooth function, the wave front set of K_λ is actually empty, so this formula is pretty redundant. Nonetheless, this formula does give us some information about where the operator T_λ may be singular in a quantitative sense. Consider the *canonical relation*

$$C_\phi = \{(x,y;\nabla_x\phi(x,y),-\nabla_y\phi(x,y)) : x \in \mathbf{R}^m, y \in \mathbf{R}^n\}.$$

A rudimentary Taylor expansion calculation shows that if $f(y) = \psi((y-y_0)/\delta_y)e^{2\pi i\eta_0\cdot y}$ for some y_0 and ξ_0, then for any particular x_0, if $|x-x_0| \leqslant \delta_x$,

$$T_\lambda f_\delta(x) \approx \delta_y^n a(x_0,y_0)e^{2\pi i\lambda(\phi(x_0,y_0)+\nabla_x\phi(x_0,y_0)\cdot(x-x_0))}\widehat{\psi}(-\delta_y[\lambda\nabla_y\phi(x_0,y_0)+\eta_0]),$$

where the left and right differ by $\delta_y^n \cdot O(\delta_x + \lambda\delta_x^2 + \lambda\delta_y^2)$. In particular, this approximation is good provided that $\delta_x \lesssim \lambda^{-1/2}$ and for $\delta_y \lesssim \lambda^{-1/2}$. We intuit from this that if f is a wave packet with unit amplitude, localized in space near y_0 with an uncertainty of $\lambda^{-1/2}$, and localized near a frequency η_0, then, when localized in space near x_0, with an uncertainty of $\lambda^{-1/2}$, $T_\lambda f$ looks like a wave packet with amplitude

$$O_N(\lambda^{-n/2}(1+\lambda^{1/2}|\lambda\nabla_y\phi(x_0,y_0)+\eta_0|)^{-N})$$

for all $N > 0$, and with frequency localized near $\lambda\nabla_x\phi(x_0,y_0)$. In particular, if $\eta_0 = -\lambda\nabla_y\phi(x_0,y_0)$, then we obtain a wave packet in the output with an amplitude of $\approx \lambda^{-n/2}$ when localized in a $\Theta(\lambda^{-1/2})$ radius neighbourhood of x_0. And if η_0 is chosen away from this value, the amplitude becomes very small as $\lambda \to \infty$. This indicates the canonical relation determines which wave packets induce a noticable effect on the output, and where this effect occurs. In particular, if we take f with the former properties, then $\|T_\lambda f\|_{L^q(\mathbf{R}^m)} \gtrsim \lambda^{-n/2}\lambda^{-m/2q}$, whereas $\|f\|_{L^p(\mathbf{R}^n)} \sim \lambda^{-n/2p}$, which implies that the operator norm of T_λ from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^m)$ is at least $\Omega(\lambda^{(n/2)(1/p-1)-(m/2)(1/q)})$. In particular, we cannot expect to do any better than the trivial bound $\|T_\lambda f\|_{L^\infty(\mathbf{R}^m)} \lesssim \|f\|_{L^1(\mathbf{R}^n)}$ when $p = 1$ and $q = \infty$. In certain situations, we will be able to conclude that T_λ actually does have an operator norm achieving this decay bound, provided we assume suitable geometric conditions on the canonical relation C_ϕ.

To begin with, let us consider the case where the matrices $D_y \nabla_x \phi = (D_x \nabla_y \phi)^T$ always has full rank on the support of the function a. If $m \geqslant n$, this information can be expressed by saying that $C_\phi \to T^* \mathbf{R}^n$ is a submersion, and $C_\phi \to T^* \mathbf{R}^m$ is an immersion, with these properties swapped if $m \leqslant n$. Thus, locally speaking, we have a smooth map $T^* \mathbf{R}^m \to C_\phi$ which is a left inverse to the corresponding projection map. For intuitions sake, let us assume this smooth map is globally defined, so we can obtain a smooth map $\alpha = (y, \eta) : T^* \mathbf{R}^m \to T^* \mathbf{R}^n$ such that if $(x_0, y_0; \xi_0, \eta_0) \in C_\phi$, then $y_0 = y(x_0, \xi_0)$ and $\eta_0 = \eta(x_0, \xi_0)$. Thus, roughly speaking, the only possible wave packet f localized spatially at a scale $\lambda^{-1/2}$ whose corresponding output $T_\lambda f$ has significant amplitude when localized at a scale $\lambda^{-1/2}$ near x_0 and oscillates at a frequency ξ_0 is a wave packet at position $y(x_0, \xi_0)$ and with frequency $\eta(x_0, \xi_0)$. The fact that $|\alpha(x_0, \xi_0) - \alpha(x_1, \xi_1)| \gtrsim |(x_0, \xi_0) - (x_1, \xi_1)|$ reflects a kind of 'orthogonality property' of the operator T_λ. To understand this orthogonality, we rely on L^2 techniques, namely a T^*T argument.

Theorem 23.20. *Consider a family of oscillatory integral operators*

$$\{T_\lambda : L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^m)\}$$

with associated phase ϕ. If $D_y \nabla_x \phi$ is injective, i.e. the map $C_\phi \to T^* \mathbf{R}^m$ is an immersion, then

$$\|T_\lambda f\|_{L^2(\mathbf{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbf{R}^n)}.$$

Proof. Once sufficiently localized, the assumption above implies that

$$|\nabla_x \phi(x, y_1) - \nabla_x \phi(x, y_2)| \gtrsim |y_1 - y_2|.$$

We calculate that

$$T_\lambda^* g(y) = \int \overline{a(x,y)} e^{-2\pi i \lambda \phi(x,y)}\, dx.$$

Thus the kernel of the operator $T_\lambda^* T_\lambda$ is equal to

$$K(y_1, y_2) = \int \overline{a(x, y_1)} a(x, y_2) e^{2\pi i \lambda [\phi(x, y_1) - \phi(x, y_2)]}\, dx.$$

The principle of nonstationary phase thus tells us that

$$|K(y_1, y_2)| \lesssim (1 + \lambda |y_1 - y_2|)^{-N}.$$

Thus for $|K(y_1, y_2)| \lesssim 1$ for $|y_1 - y_2| \lesssim 1/\lambda$, and $|K(y_1, y_2)| \lesssim_N \lambda^{-N} |y_1 - y_2|^{-N}$ for $|y_1 - y_2| \gtrsim 1/\lambda$. Schur's test thus gives that $\|T_\lambda^* T_\lambda f\|_{L^1(\mathbf{R}^n)} \lesssim \lambda^{-n} \|f\|_{L^1(\mathbf{R}^n)}$, and that $\|T_\lambda^* T_\lambda f\|_{L^\infty(\mathbf{R}^n)} \lesssim \lambda^{-n} \|f\|_{L^\infty(\mathbf{R}^n)}$, and interpolation that $\|T_\lambda^* T_\lambda f\|_{L^2(\mathbf{R}^n)} \lesssim \lambda^{-n} \|f\|_{L^2(\mathbf{R}^n)}$, and so $\|T_\lambda f\|_{L^2(\mathbf{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbf{R}^n)}$. □

Remark. The best possible decay bound for the operator norm of T_λ for general n and m is $\lambda^{-(n+m)/4}$, and since the theorem above can only be applied when $n \leqslant m$, the result is only optimal when $n = m$.

Interpolating between the trivial bound $\|T_\lambda f\|_{L^\infty(\mathbf{R}^n)} \lesssim \|f\|_{L^1(\mathbf{R}^n)}$ yields that for $2 \leqslant q \leqslant \infty$, if p is the conjugate dual of q, then

$$\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^n)}.$$

More generally, using the fact that $T_\lambda f$ only depends on the behaviour of f on $\mathrm{supp}_y(a)$, we conclude that we have a bound $\|T_\lambda f\|_{L^\infty(\mathbf{R}^n)} \lesssim \|f\|_{L^p(\mathbf{R}^n)}$ for any $1 \leqslant p \leqslant \infty$. Interpolating this bound with $\|T_\lambda f\|_{L^2(\mathbf{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbf{R}^n)}$ then gives $\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^n)}$ whenever $1/p + 1/q \leqslant 1$ and $1 \leqslant p \leqslant 2$ (equivalently, $2 \leqslant q \leqslant \infty$).

One might see an analogy with this family of inequalities to the Hausdorff-Young inequality $\|\widehat{f}\|_{L^q(\mathbf{R}^n)} \lesssim \|f\|_{L^p(\mathbf{R}^n)}$, which holds when $1/p + 1/q = 1$. Indeed, the Hausdorff-Young inequality actually follows from these bounds, since one can consider the truncated Fourier transforms

$$F_\lambda f(\xi) = \int a(x/\lambda, \xi/\lambda) e^{2\pi i \xi \cdot x} f(x)\, dx = \lambda^n \int a(x, \xi) e^{2\pi i \lambda \xi \cdot x} f(\lambda x)\, dx$$

for some bump function $a \in C_c^\infty(\mathbf{R})$, which equal the normal Fourier transform in the limit as $\lambda \to \infty$ for suitably regular functions f. The operators $T_\lambda = \lambda^{-n} F_\lambda \circ \mathrm{Dil}_\lambda$ are given by

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

and $D_\xi \nabla_x (\xi \cdot x)$ is the identity matrix, so the theory above implies that $\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^n)}$. Rescaling this inequality shows that $\|F_\lambda f\|_{L^q} \lesssim$

$\|f\|_{L^p(\mathbf{R}^n)}$, and taking limits as $\lambda \to \infty$ then yields Hausdorff Young. Thus we see obtaining results for oscillatory integral operators without compact support can be obtained from this theory if we can obtain oscillatory integral bounds with a decay of the form $\lambda^{-n/q}$.

One phase to which the last result does not imply is that phase $\phi(x,y) = |x-y|$, since, for $x \neq y$,

$$\nabla_x \phi = \frac{x-y}{|x-y|} \quad \text{and} \quad \nabla_y \phi = \frac{y-x}{|y-x|}.$$

This implies that $D_y \nabla_x \phi$ cannot possibly be invertible, since the map $y \mapsto \nabla_x \phi(x,y)$ is not an open map (it's image lies in S^{n-1}). The canonical relation C_ϕ is a cone of points of the form

$$\left\{ \left(x, y; \frac{x-y}{|x-y|}, \frac{x-y}{|x-y|} \right) \right\},$$

and so each pair (x_0, ξ_0) is related to a one dimensional family of points (y, ξ_0), where y lies on the half-line of points through x_0 pointing in the direction $-\xi_0$. Similarily, each pair (y_0, η_0) is related to a one dimensional family of points (x, ξ_0), where x lies on a half-life of points pointing in the direction ξ_0. In particular, this means that obtaining a bound of the form $\|T_\lambda f\|_{L^2(\mathbf{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbf{R}^n)}$ is impossible for such a phase. More precisely, fix $x_0 \in \text{supp}_x(a)$ and a half-line $l = \{x \in \mathbf{R}^n : (x - x_0)/|x - x_0| = \xi_0\}$ starting at x_0 for some $|\xi_0| = 1$ whose intersection with $\text{supp}_y(a)$ has non-empty interior in l. Fix $N \sim \lambda^{1/2}$, and consider a family of N wave packets $f_1, \dots, f_N$, each with pairwise disjoint support, but supported spatially near a family of $O(\lambda^{-1/2})$ separated points lying on $l \cap \text{supp}_y(a)$, with spatial uncertainty $\lambda^{-1/2}$, and with frequency concentrated near ξ_0. If $f = f_1 + \dots + f_N$, then the formula above implies that $T_\lambda f$ will have the majority of it's spatial support in the set

$$\{x_0 \in \mathbf{R}^n : |\lambda (x_0 - y_0)/|x_0 - y_0| - \xi_0| \lesssim \lambda^{-1/2}\},$$

and have amplitude $\lambda^{-(n-1)/2}$ over there. Roughly speaking, this is a tube centered at y_0, pointing in the direction ξ_0, and with thickness $O(\lambda^{-1/2})$, and thus has volume $\Theta(\lambda^{-(n-1)/2})$. But this means that $\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \gtrsim \lambda^{-(n-1)/2} \lambda^{-(n-1)/2q}$, and so combined with the fact that $\|T_\lambda f\|_{L^p(\mathbf{R}^n)} \sim \lambda^{-(n-1)/2p}$, this means the operator norm of T_λ must be $\Omega(\lambda^{(n-1)/2(1/p-1/q-1)})$, which for $p = q = 2$ gives $\Omega(\lambda^{-(n-1)/2})$. On the other hand, this does not preclude a bound of the form $\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^n)}$ from holding when $q \geqslant (n+1)/(n-1) \cdot p^*$, which forms a subset of the family of all of the cases we proved above when, in addition, $1 \leqslant p \leqslant 2$. We will show, using the fact that the projections of C_ϕ onto $T^* \mathbf{R}^n$ are surfaces of constant curvature, that we can obtain the bound under this assumption when $1 \leqslant p \leqslant 2$, and even when $1 \leqslant p \leqslant 4$ when $n = 2$.

To analyze the operator, we consider a cutoff ψ with $\psi(z) = 0$ for $|z| \lesssim 1$, and with $\psi(z) = 1$ for $|z| \gtrsim 1$. Write $T_\lambda = \tilde{T}_\lambda + R_\lambda$, where

$$\tilde{T}_\lambda f(x) = \int a(x,y) \psi(x-y) e^{2\pi i \lambda |x-y|} f(y)\, dy.$$

If we can establish that $\|\tilde{T}_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^n)}$, then we know that

$$R_\lambda f(x) = \int a(x,y)(1 - \psi(x-y)) e^{2\pi i \lambda |x-y|}$$

TODO: COMPLETE ARGUMENT.

To study $\tilde{T}_\lambda$, we can cover the support of a by finitely many open sets $\{U_\alpha \times V_\alpha\}$, upon each of which we may find a diffeomorphism $y_\alpha : W_\alpha \times I_\alpha \to V_\alpha$, where $W_\alpha \subset \mathbf{R}^{n-1}$, I_α is an interval, and for any $x \in U_\alpha$ and $t \in I_\alpha$, the map

$$v \mapsto \frac{y_\alpha(v,t) - x}{|y_\alpha(v,t) - x|}$$

is an immersion from W_α to $\mathbf{R}^n$. Applying a partition of unity $\{\psi_\alpha\}$ over these open sets, we can write $\tilde{T}_\lambda = \sum_\alpha T_\lambda^\alpha$. Moreover, we can write

$$\begin{aligned}
T_\lambda^\alpha &= \int a(x,y) \psi_\alpha(x,y) e^{2\pi i \lambda |x-y|}\, dy \\
&= \int \frac{a(x, y_\alpha(v,t)) \psi_\alpha(x, y_\alpha(v,t))}{|\det(D y_\alpha(v,t))|} e^{2\pi i \lambda |x - y_\alpha(v,t)|} f(y_\alpha(v,t))\, dv\, dt \\
&= \int_{I_\alpha} \int_{W_\alpha} a_\alpha(x, v; t) e^{2\pi i \lambda |x - y_\alpha(v,t)|} f(y_\alpha(v,t))\, dv\, dt \\
&= \int_{I_\alpha} T_\lambda^{\alpha,t} f_{t,\alpha}(x),
\end{aligned}$$

where $f_{t,\alpha}(v) = f(y_\alpha(v,t))$, and if we set $\phi(x,v;t) = |x - y_\alpha(v,t)|$, then

$$T_\lambda^{\alpha,t} f(x) = \int_{W_\alpha} a_\alpha(x,v;t) e^{2\pi i \lambda \phi(x,v;t)} f(v)\, dv.$$

The next Lemma will justify that $\|T^{\alpha,t} f_{t,\alpha}\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f_{t,\alpha}\|_{L^p(W_\alpha)}$, from which, together with Hölder's inequality, it follows that

$$\begin{aligned}
\|T_\lambda^\alpha f\|_{L^q(\mathbf{R}^n)} &\lesssim \int_{I_\alpha} \|T_\lambda^{\alpha,t} f_{t,\alpha}\|_{L^q(\mathbf{R}^n)} \\
&\lesssim \lambda^{-n/q} \int_{I_\alpha} \|f_{t,\alpha}\|_{L^p(W_\alpha)}\, dt \\
&\lesssim \lambda^{-n/q} |I_\alpha|^{1-1/p} \left(\int \|f_{t,\alpha}\|_{L^p(W_\alpha)}^p \right)^{1/p} \\
&\lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^n)}.
\end{aligned}$$

Let us now prove this lemma, we note that the canonical relation associated with the family of operators $\{T_\lambda^{\alpha,t}\}$ is

$$C^{\alpha,t} = \left\{ \left(x, v; \frac{|x - y_\alpha(v,t)|}{|x - y_\alpha(v,t)|}, D_v(y_\alpha)(v,t)^T \cdot \frac{x - y_\alpha(v,t)}{|x - y_\alpha(v,t)|} \right) \right\}.$$

Our assumptions on the coordinate system y_α implies that the projection map $C^{\alpha,t} \to T^* W_\alpha$ is an immersion, i.e. the matrix $D_v \nabla_x \phi_{t,\alpha}$ has full rank $n-1$ on the domain of the integral. And the image of the projection $C^{\alpha,t} \to T^* \mathbf{R}^n$ is an open subset of the submanifold of unit vectors in $T^* \mathbf{R}^n$, which, at each point x, is a $n-1$ dimensional manifold with non-vanishing curvature. These assumptions justify the application of the next Lemma.

Lemma 23.21. *Suppose $\phi : \mathbf{R}^n \times \mathbf{R}^{n-1}$ is a C^∞ phase, that $\nabla_x D_y \phi$ has rank $n-1$ on the support of $a \in C_c^\infty(\mathbf{R}^n \times \mathbf{R}^{n-1})$, and that the image of the projection $C_\phi \to T * \mathbf{R}^n$ is on each fibre a hypersurface of non-vanishing curvature. Then if $1/p + 1/q \leqslant 1$, $1 \leqslant p \leqslant 2$, and $q \geqslant [(n+1)/(n-1)] \cdot p^*$, then*

$$\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^{n-1})}.$$

Remark. Using a similar construction as for the operator associated with the phase $\phi(x,y) = |x - y|$, one can show from the assumption that there exists f such that $\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \geqslant \lambda^{(n-2)/2p-(n-2)/2-n/2q}\|f\|_{L^p(\mathbf{R}^{n-1})}$, which shows that one can only obtain a bound of the form above when $q \geqslant [(n+1)/(n-1)] \cdot p^*$. We also note that the theorem above, once rescale, implies extension estimates of the form

$$\left\| \int_{\mathbf{R}^{n-1}} e^{2\pi i x \cdot \phi(\xi)} f(\xi)\, d\xi \right\|_{L^q(\mathbf{R}^n)} \lesssim \|f\|_{L^p(\mathbf{R}^{n-1})}$$

for the same range of p and q, which are dual to the corresponding restriction estimates for the hypersurface $H = \{(\xi, \phi(\xi))\}$.

Proof. To prove the result, we may assume that $q = [(n+1)/(n-1)] \cdot p^*$. Since the result is proved when $p = 1$, it suffices to show that

$$\|T_\lambda f\|_{L^{(n+1)/2(n-1)}(\mathbf{R}^n)} \lesssim \lambda^{-n(n-1)/2(n+1)}\|f\|_{L^2(\mathbf{R}^{n-1})}.$$

The assumption that the surface $y \mapsto \nabla_x \phi(x,y)$ has non-vanishing curvature can be expressed in the following manner: I can, locally at least, for each $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^{n-1}$, pick a unit normal vector $\nu(x,y) \in S^{n-1}$ to the surface $y \mapsto \nabla_x \phi(x,y)$ at the point $\nabla_x \phi(x,y)$. This means precisely that for each fixed x_0 and y_0,

$$\nabla_y(\nabla_x \phi(x_0,y) \cdot \nu(x_0,y_0)) = 0$$

when $y = y_0$. If $H(x_0,y_0)$ is the Hessian matrix of the operator $y \mapsto \nabla_x \phi(x_0,y) \cdot \nu(x_0,y_0)$, then this means precisely that

$$\nabla_x \phi(x_0,y) \cdot \nu(x_0,y_0) = (y-y_0)^T H(x_0,y_0)(y-y_0) + O(|y-y_0|^3).$$

The statement that the surface does not have vanishing curvature at (x_0,y_0) is then equivalent to the fact that $H(x_0,y_0)$ is invertible.

To prove the theorem, perhaps after localizing and taking a change of x-variables, we may assume that $x = (z,t)$, where $D_z \nabla_y \phi$ is invertible. Introduce the operators T_λ^t, mapping functions from $\mathbf{R}^{n-1}$ to $\mathbf{R}^{n-1}$ by the formula

$$T_\lambda^t f(z) = \int a(z,y;t) e^{2\pi i \lambda \phi(z,y;t)} f(y)\, dy.$$

Then

$$\|T_\lambda f\|_{L^q(\mathbf{R}^d)} \sim \left(\int \|T_\lambda^t f\|_{L^q(\mathbf{R}^{n-1})}^q\, dt \right)^{1/q},$$

and it suffices to bound this integral. The advantage of rephrasing the problem in terms of the oscillatory integral operators T^t_λ is because we already have some tight estimates for such operators, namely

$$\|T^t_\lambda f\|_{L^q(\mathbf{R}^{n-1})} \lesssim \lambda^{-(n-1)/q} \|f\|_{L^p(\mathbf{R}^{n-1})}.$$

However, directly plugging this into the inequality only gives a bound $\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^q(\mathbf{R}^{n-1})}$. To gain a boost on what is essentially a trivial application of the triangle inequality, we need to know more about the interactions between the family of operators $\{T^t_\lambda\}$. This is where the curvature of the canonical relation comes into play, since it implies a kind of orthogonality of the operators.

To utilize this kind of orthogonality, we again apply adjoint techniques. There is $b \in C_c^\infty$ and $\Phi(z_1, z_2, y; t_1, t_2) = \phi(z_1, y, t_1) - \phi(z_2, y, t_2)$ such that the kernel of the operator $T^{t_1}_\lambda (T^{t_2}_\lambda)^*$ is

$$K_{t_1, t_2}(z_1, z_2) = \int a(z_1, z_2, y; t_1, t_2) e^{2\pi i \lambda (\phi(z_1, y, t_1) - \phi(z_2, y, t_2))}\, dy.$$

Let's analyze this using stationary phase. We have $\nabla_y \Phi(z_0, z_0, y_0, t_0, t_0) = 0$ for any choice of z_0, y_0, and t_0, yet we have $D_{z_1} \nabla_y \Phi(z, z, y, t, t) = D_{z_1} \nabla_y \phi(z_1, y, t_1)$, which is an invertible matrix by assumption. Thus by the implicit function theorem, there exists a function $z_1 = z_1(z_2, y, t_1, t_2)$ defined in a neighborhood of (z_0, y_0, t_0, t_0) which is a unique solution to the equation $\nabla_y \Phi(z_1, z_2, y, t_1, t_2) = 0$. By localizing in these variables, we may assume these are the only such stationary points. A Taylor expansion gives

$$\begin{aligned} \nabla_y \Phi(z_1, z_2, y, t_1, t_2) = D_z \nabla_y \phi(z_2, y, t_2) \cdot (z_1 - z_2) \\ + D_t \nabla_y \phi(z_2, y, t_2) \cdot (t_1 - t_2) \\ + O(|z_1 - z_2|^2 + |t_1 - t_2|^2). \end{aligned}$$

Thus

$$|D_z \nabla_y \phi(z_2, y, t_2) \cdot (z_1 - z_2) + D_t \nabla_y \phi(z_2, y, t_2) \cdot (t_1 - t_2)| \lesssim O(|z_1 - z_2|^2 + |t_1 - t_2|^2).$$

It follows from this that $(z_1 - z_2, t_1 - t_2)$ must be in a small, conical neighborhood of $\pm \nu(z_2, y, t_2)$ (For any matrix A, $|Ax| \gtrsim |x|$ unless x is close to the kernel of A). If we write $x_i = (z_i, t_i)$ again, this means that for the apropriate sign $|(x_1 - x_2) \pm |x_1 - x_2|\nu(x_2, y)| \lesssim |x_1 - x_2|$. But then

$$\begin{aligned} D_y \nabla_y \Phi(x_1, x_2, y) &= D_y \nabla_y \{D_x \phi(x_2, y) \cdot (x_1 - x_2)\} + O(|x_1 - x_2|^2) \\ &= |x_1 - x_2| D_y \nabla_y \{D_x \phi(x_2, y) \cdot \nu(x_2, y)\} + O(|x_1 - x_2|), \end{aligned}$$

If we choose all the localization values above carefully enough, our curvature assumption implies that this matrix is invertible with determinant $|x_1 - x_2|^{n-1}$, so we have nondegenerate stationary points. Thus isolating near these critical points, we obtain a decay in the oscillatory integral of the form $\lesssim (1 + \lambda|x_1 - x_2|)^{-(n-1)/2}$. Away from these critical points, we have $|\nabla_y \Phi(x_1, x_2, y)| \gtrsim |x_1 - x_2|$ (since if $|\nabla_y \Phi(x_1, x_2, y)| \ll |x_1 - x_2|$, then $x_1 - x_2$ will be in one of the neighborhoods we considered above), and so when we integrate here, nonstationary phase implies that we have a decay of the form $O_N((1 + \lambda|x_1 - x_2|)^{-N})$ for all N. Putting together these estimates gives that

$$|K^{t_1, t_2}(z_1, z_2)| \lesssim (1 + \lambda|t_1 - t_2| + \lambda|z_1 - z_2|)^{-(n-1)/2}.$$

This implies the estimates

$$\|T_\lambda^{t_1} (T_\lambda^{t_2})^* f\|_{L^\infty(\mathbf{R}^{n-1})} \lesssim \lambda^{-(n-1)/2} |t_1 - t_2|^{-(n-1)/2} \|f\|_{L^1(\mathbf{R}^{n-1})}.$$

This looks like something we could apply the theory of fractional integrals to understand. And indeed we can, as the next result shows: TODO. $\square$

Chapter 24

Restriction Theorems

If a function f lies in $L^p(\mathbf{R}^d)$, there does not exist a numerically meaningful way to restrict f to a set of measure zero, since f is only defined up to measure zero. More precisely, for any $0 < p < \infty$, $0 < q \leqslant \infty$, and any Radon measure σ supported on a set S with Lebesgue measure zero, there does not exist a bound $\|f\|_{L^q(S,\mu)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$ for all $f \in C(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$. We can consider a very similar problem for the Lebesgue measure; given $f \in L^1(\mathbf{R}^d)$, $\widehat{f} \in C(\mathbf{R}^d)$, and so $\widehat{f}|_S$ is a well defined function on S. If there exists a bound of the form

$$\|\widehat{f}\|_{L^q(S,\mu)} \lesssim \|f\|_{L^p(\mathbf{R}^d)} \tag{24.1}$$

for all $f \in L^1(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$, then (assuming $p < \infty$, which will always be the case in nontrivial situations) there exists a unique bounded operator $R : L^p(\mathbf{R}^d) \to L^q(S,\mu)$ such that $Rf = \widehat{f}|_S$ for $f \in L^1(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$; thus we have a meaningful way of restricting the Fourier transforms of functions in $L^p(\mathbf{R}^d)$ to S. In light of the failure to restrict elements of $L^p(\mathbf{R}^d)$ to $L^q(S,\mu)$ without taking the Fourier transform, this indicates that the Fourier transform of $L^p(\mathbf{R}^d)$ is a fairly special element of $L^q(\mathbf{R}^d)$. We refer to an estimate of the form above as a *restriction estimate*.

The approximate translation invariance of restriction bounds implies (from Littlewood's principle) that $q \geqslant p$ for any restriction estimate. For $p = 1$, a restriction bound from $L^1(\mathbf{R}^d)$ to $L^\infty(S,\mu)$ is trivial in light of the bound $\|\widehat{f}\|_{L^\infty(\mathbf{R}^d)} \leqslant \|f\|_{L^1(\mathbf{R}^d)}$. On the other hand, for $p = 2$ and $0 < q \leqslant \infty$, *no such bound is possible* unless μ is absolutely continuous with respect to the Lebesgue measure. This follows from a simple application of Parseval's theorem. Interpolation shows we also should not expect any bounds when $p > 2$, and indeed such restriction estimates even fail when μ is absolutely continuous, for the Hausdorff-Young inequality fails in this setting. Thus the interesting bounds occur when $1 < p < 2$. For a particular pair (S,μ) and exponent q, the goal is to push the value of p as close to 2 as possible. Often, we study a smooth surface S, and consider a measure μ which is absolutely continuous with respect to the surface measure on S. We shall find that in this setting there is a rich theory relating the existence of restriction maps to the curvature of the surface S.

There is a dual problem associated with the pair (S,μ). For $g \in L^1(S,\mu)$, we consider the *extension operator*

$$E_S g(x) = \int_S g(\xi) e^{2\pi i \xi \cdot x}\, d\mu(\xi).$$

In other words, $E_S g$ is the inverse Fourier transform of the finite Borel measure $g \cdot \mu$. It is simple to see using Fubini's theorem that if $f \in L^1(\mathbf{R}^d)$ and $g \in L^1(S,\mu)$, then

$$\int_S R_S f(\xi) \overline{g(\xi)}\, d\mu(\xi) = \int_{\mathbf{R}^d} f(x) \overline{(E_S g)(x)}\, dx.$$

For $1 \leqslant p < \infty$, and $1 < q \leqslant \infty$, $\mathcal{S}(\mathbf{R}^d)$ is dense in $L^p(\mathbf{R}^d)$ and $C_c^\infty(\mathbf{R}^d)$ is dense in $L^{q'}(S,\mu)$. Moreover, for this range of parameters, $R_S f$ lies in $L^q(\mathbf{R}^d)$ for all $f \in \mathcal{S}(\mathbf{R}^d)$, and $E_S g$ lies in $L^{p'}(\mathbf{R}^d)$ for all $g \in C_c^\infty(\mathbf{R}^d)$. We therefore conclude by duality that for $1 \leqslant p < \infty$ and $1 < q \leqslant \infty$, a bound of the form

$$\|R_S f\|_{L^q(S,\mu)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

holding for all $f \in \mathcal{S}(\mathbf{R}^d)$ is equivalent to a bound

$$\|E_S g\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|g\|_{L^{q'}(S,\mu)},$$

for all $g \in C_c^\infty(\mathbf{R}^d)$. Since bounds only hold for $1 \leqslant p < 2$ then this duality result essentially covers all cases of interest.

If μ is a singular measure, using the fact that $E_S g = g \cdot \mu$, for any nonzero g we see that we cannot have $E_S g \in L^{p'}(\mathbf{R}^d)$ for any $2 \leqslant p \leqslant \infty$, because then Hausdorff-Young and the Fourier inversion theorem would imply that $g \cdot \mu \in L^p(\mathbf{R}^d)$, which is not the case for any such value of p. Thus, just like restriction estimates, extension estimates in nontrivial situations only hold when $1 \leqslant p < 2$.

It is simple to prove that restriction bounds for product sets reduce to bounds on their projections. Thus the interesting sets S we consider will never be product sets.

Theorem 24.1. *Consider two pairs (S_1, μ_1) and (S_2, μ_2) in $\mathbf{R}^n$ and $\mathbf{R}^m$, and let $S = S_1 \times S_2$, $\mu = \mu_1 \times \mu_2$. Then for $0 < p \leqslant q \leqslant \infty$, a restriction bound of the form*

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

for all $f \in \mathcal{S}(\mathbf{R}^{n+m})$ is equivalent to a pair of bounds of the form

$$\|\widehat{f_1}\|_{L^q(S_1,\mu_1)} \lesssim \|f_1\|_{L^p(\mathbf{R}^n)} \quad \textit{and} \quad \|\widehat{f_2}\|_{L^q(S_2,\mu_2)} \lesssim \|f_2\|_{L^p(\mathbf{R}^m)}.$$

for all $f_1 \in \mathcal{S}(\mathbf{R}^n)$ and $f_2 \in \mathcal{S}(\mathbf{R}^m)$.

Proof. We calculate that for any pair f_1, f_2,

$$\|\widehat{f_1 \otimes f_2}\|_{L^q(S,\mu)} = \|\widehat{f_1}\|_{L^q(S_1,\mu_1)} \|\widehat{f_2}\|_{L^q(S_2,\mu_2)}.$$

and

$$\|f_1 \otimes f_2\|_{L^p(\mathbf{R}^{n+m})} = \|f_1\|_{L^p(\mathbf{R}^n)} \|f_2\|_{L^p(\mathbf{R}^m)}.$$

If a restriction estimate held on (S, μ), it would then follow that

$$\|\widehat{f_1}\|_{L^q(S_1,\mu_1)} \lesssim \frac{\|f_2\|_{L^p(\mathbf{R}^d)}}{\|\widehat{f_2}\|_{L^q(S_2,\mu_2)}} \|f_1\|_{L^p(\mathbf{R}^n)}.$$

Choosing any nontrivial choice of f_2 gives a restriction estimate on (S_1, μ_1). By symmetry, we also get a restriction estimate on (S_2, μ_2).

Conversely, suppose we have a restriction estimate on (S_1, μ_1) and (S_2, μ_2), and $f \in L^1(\mathbf{R}^{n+m}) \cap L^p(\mathbf{R}^{n+m})$. Without loss of generality, we may assume by Littlewood's principle that $q \geqslant p$. If we temporarily let $\mathcal{F}_1$ and $\mathcal{F}_2$ denote the Fourier transform in the first and second variables respectively, then by applying Minkowski's inequality and Fubini's theorem we find that

$$\begin{aligned}
\|\widehat{f}\|_{L^q(S,\mu)} &= \left(\int_{\mathbf{R}^m} \int_{\mathbf{R}^n} |\mathcal{F}_1(\mathcal{F}_2 f)(\xi_1,\xi_2)|^q \, d\mu_1(\xi_1) d\mu_2(\xi_2) \right)^{1/q} \\
&\lesssim \left(\int_{\mathbf{R}^m} \left(\int_{\mathbf{R}^n} |(\mathcal{F}_2 f)(x_1,\xi_2)|^p dx_1 \right)^{q/p} d\mu_2(\xi_2) \right)^{1/q} \\
&\leqslant \left(\int_{\mathbf{R}^n} \left(\int_{\mathbf{R}^m} |(\mathcal{F}_2 f)(x_1,\xi_2)|^q \, d\mu_2(\xi_2) \right)^{p/q} dx_1 \right)^{1/p} \\
&\lesssim \left(\int_{\mathbf{R}^n} \int_{\mathbf{R}^m} |f(x_1,x_2)|^p dx_2 dx_1 \right)^{1/p} = \|f\|_{L^p(\mathbf{R}^d)}. \qquad \square
\end{aligned}$$

The most classical example of a restriction conjecture is when $\mu = \sigma$ is the surface measure on $S^{d-1} \subset \mathbf{R}^d$, or when $\mu = \psi\sigma$ for some $\psi \in C_c^\infty(S)$. If S is the graph of some smooth function $\phi : U \to \mathbf{R}$, where U is an open subset of $\mathbf{R}^{d-1}$, then (24.1) is equivalent to an estimate of the form

$$\left(\int_U \psi(\eta) |\widehat{f}(\eta,\phi(\eta))|^q \, d\eta \right)^{1/q} \lesssim \left(\int_{\mathbf{R}^d} |f(x)|^p \, dx \right)^{1/p},$$

for some $\psi \in C_c^\infty(U)$. Dual to this are the extension estimates, which are equivalent to bounds of the form

$$\begin{aligned}
&\left(\int_{\mathbf{R}^{d-1} \times \mathbf{R}} \left| \int_U \psi(\eta) f(\eta) e^{2\pi i (x\cdot\eta + y\phi(\eta))} \, d\eta \right|^{p'} dx\, dy \right)^{1/p'} \\
&\qquad \lesssim \left(\int_U \psi(\eta) |f(\eta)|^{q'} \, d\eta \right)^{1/q'}.
\end{aligned}$$

These types of estimates often appear in the theory of dispersive partial differential equations, which makes them useful outside of harmonic analysis. For instance, we can use restriction estimates to understand the free Schrödinger equation $u_t = 2\pi i \Delta_x u$. In particular, if $u : \mathbf{R}^{d-1} \times \mathbf{R} \to \mathbf{R}$ solves this equation, and $u(x,0) = f(x)$ for all $x \in \mathbf{R}^{d-1}$, then

$$u(x,t) = (e^{2\pi i \Delta t} f)(x) = \int_{\mathbf{R}^{d-1}} \widehat{f}(\xi) e^{2\pi i (\xi \cdot x - 4\pi^2 |\xi|^2 t)} \, d\xi.$$

The map $f \mapsto u$ is thus essentially the extension operator associated to the *elliptic paraboloid* $S = \{(\xi, |\xi|^2) : \xi \in \mathbf{R}^d\}$, together with the measure μ induced by the projection of S onto $\mathbf{R}^{d-1}$.

Example. *Consider extension estimates to a hyperplane, i.e. to the graph of the map* $\phi : \mathbf{R}^{d-1} \to \mathbf{R}$ *with* $\phi(x) = 0$ *for all* $x \in \mathbf{R}^{d-1}$. *We are thus attempting to obtain a bound of the form*

$$\left(\int_{\mathbf{R}^{d-1} \times \mathbf{R}} \left| \widehat{f}(x) \right|^{p'} \, dx\, dy \right)^{1/p'} \lesssim \left(\int_{\mathbf{R}^{d-1}} |f(\eta)|^{q'} \, d\eta \right)^{1/q'}.$$

The left hand side doesn't depend on y, *and thus is infinite unless* $f = 0$. *Thus no extension estimates can hold unless* $p' = \infty$ *and* $q' = 1$, *i.e. for the trivial estimates. The general heuristic here is that 'flatness' is preventing us from getting better estimates. For curved surfaces, one expects to get a much wider range of exponents.*

The general heuristic suggests that restriction estimates for surfaces with non-vanishing curvature behave the best. The classical 'toy' examples of such results are the restriction estimates for the paraboloid and the sphere, where we have additional symmetries to use not present in all surfaces. The conjectured estimates for such surfaces still remain open problems today. The conjectured exponents come from two examples which are seen to be 'worst case examples'.

The standard theory of stationary phase shows that if $\psi \in C_c^\infty(\mathbf{R}^{d-1})$ is supported in a small enough neighborhood of the origin, then there exists a continuous, non-vanishing function $a : S^{d-1} \to \mathbf{C}$ such that for $r > 0$ and $x \in S^{d-1}$,

$$E_S \psi(rx) = a(x) r^{-(d-1)/2} + O(r^{-(d+1)/2}).$$

Thus $E_S \psi \in L^{p'}(\mathbf{R}^d)$ if and only if $2d/(d-1) < p'$, i.e. $p < 2d/(d+1)$. Since $\psi \in L^{q'}(S, \mu)$ for any q', the restriction theorem is thus only interesting when $p < 2d/(d+1)$.

The second example is due to Knapp, and called the 'Knapp example' in the literature, which is really just studying the extension operator applied to a wave packet. To construct the Knapp example, by translating and rotating (which by symmetry does not change any estimates in the Fourier transform) we may assume our surface contains the origin, and near the origin is the graph of a map $\phi : U \to \mathbf{R}$, where $U \subset \mathbf{R}^{d-1}$ contains the origin, $\phi(0) = 0$, and $\nabla \phi(0) = 0$. Furthermore, rescaling μ if necessary, we may assume that $\mu(B_r(0)) \geqslant r^{d-1}$ for all $r \leqslant 1$. Now if we fix $r \leqslant 1$, and consider a bump function $\psi \in C_c^\infty(\mathbf{R}^d)$ supported on a ball of radius r with $|\psi(x)| = 1$ for all $|x| \leqslant r/2$ and with $\|\psi\|_{L^\infty(S)} \leqslant 1$, then

$\|\psi\|_{L^{q'}(S,\mu)} \sim r^{(d-1)(1-1/q)}$. On the other hand, $\psi \cdot \mu$ is actually supported on a 'cap'

$$\theta_r = \{(\xi, \xi_d) : |\xi| \leqslant r, |\xi_d| \leqslant Cr^2\}$$

for some constant C depending on ϕ, independant of r. Now $E_S \psi$ is equal to $\widehat{\psi \cdot \mu} = \widehat{\psi} * \widehat{\mu}$ once reflected about the origin. Since $\psi \in C_c^\infty(\mathbf{R}^d)$, $\widehat{\psi}$ is Schwartz, and since μ has finite measure, $\widehat{\mu}$ is a continuous, bounded function, which implies that $E_S(\psi)$ is Schwartz. Since $\widehat{E_S(\psi)}$ is supported in θ_r, the uncertainty principle heuristically implies that $E_S(\psi)$ is locally constant on translations of the 'tube'

$$\theta_r^* = \{(x, x_d) : |x| \leqslant 1/r, |x_d| \leqslant C/r^2\}.$$

Since

$$E_S(\psi)(0) = \int \psi(x) d\mu(x) \gtrsim r^{d-1}$$

it follows from (TODO: record Prop 5.5 of Wolff's notes in these notes) that

$$\int |E_S(\psi)(x, x_d)|(1 + rx + r^2 x_d)^{-N}\, dx\, dx_d \gtrsim_N |\theta_r^*| r^{d-1} \gtrsim r^{-2}$$

Hölder's inequality implies that if N is chosen larger than d, then

$$\int |E_S(\psi)(x, x_d)|(1 + rx + r^2 x_d)^{-N}\, dx\, dx_d \lesssim r^{-(d+1)(1/p)} \|E_S \phi\|_{L^{p'}(\mathbf{R}^d)},$$

and so we conclude that $\|E_S \phi\|_{L^{p'}(\mathbf{R}^d)} \geqslant r^{(d+1)(1/p)-2}$. Letting $r \to 0$ shows that an extension estimate for S from $L^{q'}(S, \mu)$ to $L^{p'}(\mathbf{R}^d)$ cannot hold if $(d+1)(1/p) - 2 < (d-1)(1-1/q)$. Thus an extension estimate can only hold for $(d-1)/q' < (d+1)/p'$.

Remark. TODO: If S is a surface passing through the origin whose derivatives vanish to order k, consider a variant of the Knapp example which shows how restriction fails in this domain.

The general belief in the field is that these examples are essentially the only barriers to restriction. Thus for any such pair $1 \leqslant p, q < \infty$ such that $p < 2d/(d+1)$ and $(d+1)(1/p) - 2 \geqslant (d-1)(1-1/q)$, we have restriction (and thus extension) bounds from $L^{q'}(S, \mu)$ to $L^{p'}(\mathbf{R}^d)$, and thus a restriction bound from $L^p(\mathbf{R}^d)$ to $L^q(S, \mu)$. This is the *restriction conjecture*, which, to a large extent, still remains an open problem.

24.1 Stein-Tomas Theorem

The classical result for restriction estimates on surfaces with non-vanishing curvature is the *Stein-Tomas theorem*. If S is a hypersurface in $\mathbf{R}^d$ with nonvanishing curvature, and $\mu = \psi\sigma$ for some $\psi \in C_c^\infty(\mathbf{R}^d)$, where σ is the surface measure on S, then for $p \leqslant 2(d+1)/(d+3)$,

$$\|\widehat{f}\|_{L^2(S)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}.$$

This result is tight, as shown by the Knapp example. There are many proofs of the Stein-Tomas theorem. The simplest uses the theory of stationary phase, but has the disadvantage that it does not include the end-point case where $p = 2(d+1)/(d+3)$.

Theorem 24.2. *If $p < 2(d+1)/(d+3)$, then for $f \in L^1(\mathbf{R}^d)$,*

$$\|R_S f\|_{L^2(S,\mu)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}.$$

Proof. We apply a TT^* argument, which can be used to study the boundedness of any operator whose codomain is a Hilbert space. For $f \in L^1(\mathbf{R}^d)$,

$$(E_S R_S f)(x) = \int_S \widehat{f}(\xi) e^{2\pi i \xi \cdot x} d\mu(\xi) = f * \widehat{\nu}.$$

where ν is the reflection of μ about the origin. The TT^* method implies that the restriction estimate we wish to prove holds if and only if we have a bound of the form

$$\|E_S R_S f\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}.$$

The operator $E_S R_S$ is just a convolution operator by $\widehat{\nu}$, which makes it much simpler to understand. In particular, the fact that S has nonvanishing curvature means that

$$|\widehat{\nu}(\xi)| \lesssim \langle \xi \rangle^{-(d-1)/2}.$$

This implies that $\widehat{\nu} \in L^r(\mathbf{R}^d)$ for $r > 2d/(d-1)$, which implies via Young's inequality that

$$\|E_S R_S f\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

for $p < 4d/(d-1)$. To improve this bound to hold for $p < 2(d+1)/(d-1)$, we will also exploit the fact that $\nu(B_r(\xi)) \lesssim r^{d-1}$ uniformly over $\xi \in \mathbf{R}^d$ and $r > 0$.

Consider a dyadic phase-space decomposition of ν. In other words, fix a bump function $\phi_0 \in C_c^\infty(\mathbf{R}^d)$ with $\phi_0(x) = 1$ for $|x| \leqslant 1$, and supported on $|x| \leqslant 2$. Define $\phi_n(x) = \phi_0(x/2^n) - \phi(x/2^{n-1})$. Then ϕ_n is supported on the annulus $2^{n-1} \leqslant |x| \leqslant 2^{n+1}$ and $\sum_{n \geqslant 0} \phi_n = 1$. Define $K_n = \phi_n \widehat{\nu}$ for $n \geqslant 0$. Then $\|K_n\|_{L^\infty(\mathbf{R}^d)} \leqslant 2^{-n(d-1)/2}$, so Young's inequality implies that

$$\|f * K_n\|_{L^\infty(\mathbf{R}^d)} \lesssim 2^{-n(d-1)/2} \|f\|_{L^1(\mathbf{R}^d)}$$

On the other hand, $\widehat{K_n} = \widehat{\phi_n} * \mu$, where for all $N > 0$,

$$|\widehat{\phi_n}(\xi)| \lesssim_N \frac{2^{dn}}{(1 + 2^n|\xi|)^N}$$

is a function which rapidly decays for $|\xi| \geqslant 1/2^n$. Thus a dyadic decomposition with $N > d-1$ yields

$$\begin{aligned}
|\widehat{K_n}(\xi)| &= \left| \int \widehat{\phi_n}(\eta) d\mu(\xi - \eta) \right| \\
&\lesssim_N 2^{dn} \mu(B_{1/2^n}(\xi)) + \sum_{k=1}^\infty (2^{dn}/2^{Nk}) \mu(B_{2^k/2^n}) \\
&\lesssim 2^n + \sum_{k=1}^\infty 2^{k(d-1) - Nk} 2^n \\
&\lesssim 2^n.
\end{aligned}$$

But this means we have a bound $\|f * K_n\|_{L^2(\mathbf{R}^d)} \lesssim 2^n \|f\|_{L^2(\mathbf{R}^d)}$. Interpolating between these two bounds gives that for $p \leqslant 2$,

$$\begin{aligned}
\|f * K_n\|_{L^{p'}(\mathbf{R}^d)} &\lesssim 2^{-(2/p-1)n(d-1)/2} 2^{2n(1-1/p)} \|f\|_{L^p(\mathbf{R}^d)} \\
&= 2^{n((d+3)/2 - (1/p)(d+1))} \|f\|_{L^p(\mathbf{R}^d)},
\end{aligned}$$

The operator norm bound here decays exponentially for

$$p < 2(d+1)/(d+3),$$

in which case we can sum over n and use the triangle inequality to obtain the bound we desire. □

Remark. The only structure we used about S and μ here is that μ has a uniform, Frostman type bound, and that $\widehat{\mu}$ has a geometric decay as frequencies become large. For such a pair one has a bound

$$\|R_S f\|_{L^2(S,\mu)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

provided that $1 \leqslant p \leqslant 2(d+\beta-\alpha)/(2d+\beta-2\alpha)$, where $0 < \alpha < n$, $\beta > 0$, where $|\widehat{\mu}(\xi)| \lesssim \langle \xi \rangle^{-\beta}$ and $|\mu(B_r(x)) \lesssim r^\alpha$.

The proof at the endpoint $p = 2(d+1)/(d+3)$ is quite different, requiring some analytic interpolation bounds. TODO. In the special case of the sphere, relying on an alternate property of the sphere.

Lemma 24.3. *Let σ be the surface measure of the sphere $S = S^{n-1}$. Then $\sigma * \sigma$ is absolutely continuous with respect to the Lebesgue measure, supported on $|\xi| \leqslant 2$, and on this set,*

$$(\sigma * \sigma)(\xi) \lesssim \begin{cases} 1/|\xi| & : 0 < |\xi| \leqslant 1, \\ (2-|\xi|)^{(n-3)/2} & : 1 \leqslant |\xi| \leqslant 2. \end{cases}$$

Proof. For each $\varepsilon > 0$, let $\sigma_\varepsilon = (1/\varepsilon)\mathbf{I}_{S_\varepsilon}$. Then σ_ε converges weakly to σ as $\varepsilon \to 0$, which implies that $(\sigma_\varepsilon * \sigma_\varepsilon)$ converges to $\sigma * \sigma$ weakly. Now $\sigma_\varepsilon * \sigma_\varepsilon$ is radial. Thus it suffices to bound $(\sigma_\varepsilon * \sigma_\varepsilon)(\xi)$ for $\xi = (r, 0)$, where $0 < r \leqslant 2$. Now

$$(\sigma_\varepsilon * \sigma_\varepsilon)(\xi) = |S_\varepsilon \cap (\xi + S_\varepsilon)|.$$

If $(\alpha, \beta) \in S_\varepsilon \cap (\xi + S_\varepsilon)$, where $\alpha \in \mathbf{R}$, $\beta \in \mathbf{R}^{n-1}$, then

$$1 - 2\varepsilon \leqslant \alpha^2 + \beta^2 \leqslant 1 + 3\varepsilon \quad \text{and} \quad 1 - 2\varepsilon \leqslant (\alpha - r)^2 + \beta^2 \leqslant 1 + 3\varepsilon.$$

Together, these inequalities imply that $\alpha = r/2 + O(\varepsilon/r)$ and thus $\beta^2 = 1 - r^2/4 + O(\varepsilon + \varepsilon^2/r)$, so $\beta = \sqrt{1 - r^2/4} + O()$. TODO Thus α ranges over an interval of length $O(\varepsilon/r)$, and for each α, β can vary over a region of $n-1$ dimensional volume $O(\varepsilon)$, so we conclude that □

24.2 Restriction on the Paraboloid

The main way we can extend local estimates to global estimates is to apply some kind of symmetry property, normally scaling. Let us use this technique to show that we can extend the result of the Tomas-Stein theorem to give the same bounds for restriction to the non-compact elliptic paraboloid

$$S = \{(\xi, \omega) \in \mathbf{R}^{d-1} \times \mathbf{R} : |\xi|^2 = \omega\}.$$

We work with extension estimates, setting, for a function $f : \mathbf{R}^{d-1} \to \mathbf{C}$,

$$Ef(x,t) = \int_{\mathbf{R}^{d-1}} e^{2\pi i(|\xi|^2 t + \xi \cdot x)} f(\xi)\, d\xi.$$

Our goal is to show bounds of the form $\|Ef\|_{L^p(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^{d-1})}$. This is possible, provided that$p = 2(d+1)/(d+3)$.

Theorem 24.4. *If $p = 2(d+1)/(d+3)$, then for any $f \in \mathcal{S}(\mathbf{R}^{d-1})$,*

$$\|Ef\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^{d-1})}.$$

Proof. The Tomas-Stein theorem gives the required bound for any $f \in C_c^\infty(\mathbf{R}^d)$, where the implicit constant depends on the support of f. For such f, define

$$Ef(x,t) = \int_{\mathbf{R}^{d-1}} e^{2\pi i(|\xi|^2 t + \xi \cdot x)} f(\xi)\, d\xi.$$

We note that $E(\text{Dil}_{\xi,a} f) = a^{d-1} \cdot \text{Dil}_{a^{-2},t} \text{Dil}_{a^{-1},x}(Ef)$, in light of the 'parabolic symmetry' of the extension operator. Given a function $f \in C_c^\infty(\mathbf{R}^d)$ supported on a ball of radius R at the origin, $\text{Dil}_{\xi,1/R} f$ is supported on a ball of radius 1 at the origin, and so Tomas-Stein says that

$$\|E(\text{Dil}_{\xi,1/R} f)\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|\text{Dil}_{\xi,1/R} f\|_{L^2(\mathbf{R}^{d-1})} = R^{-(d-1)/2} \cdot \|f\|_{L^2(\mathbf{R}^{d-1})},$$

where the implicit constant is independent of f. But we also know that

$$\begin{aligned} \|E(\text{Dil}_{\xi,1/R} f)\|_{L^{p'}(\mathbf{R}^d)} &= R^{1-d} \|\text{Dil}_{R^2,t} \text{Dil}_{R,x}(Ef)\|_{L^{p'}(\mathbf{R}^d)} \\ &= R^{1-d} R^{2/p'} R^{(d-1)/p'} \|Ef\|_{L^{p'}(\mathbf{R}^d)}, \end{aligned}$$

and so we conclude that

$$\|Ef\|_{L^{p'}(\mathbf{R}^d)} \lesssim R^{(d-1)/2 - (d+1)/p'} \|f\|_{L^2(\mathbf{R}^{d-1})} = \|f\|_{L^2(\mathbf{R}^{d-1})}.$$

In other words, we have found a bound independant of the support of f. A simple approximation argument extends the bound to general $f \in \mathcal{S}(\mathbf{R}^d)$. $\square$

Remark. The same scaling symmetries show that we can only have an extension bound

$$\|Ef\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|f\|_{L^{q'}(\mathbf{R}^{d-1})}$$

if $(d+1)/p' = (d-1)/q$. For $q = 2$, this shows the above result is tight.

It is a general heuristic that the extension theory for the paraboloid is essentially the same as the extension theory for a cap of a curve point. In particular, if we have a bound of the form

$$\|Ef\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|f\|_{L^{q'}(S,\mu)},$$

where $(d+1)/p' = (d-1)/q$, and where μ is a smooth measure supported on a small cap around the neighbourhood of a point with non-vanishing curvature. One can then approximate the paraboloid by a rescaling of the cap, and thus obtain a bound of the form

$$\|Ef\|_{L^{p'}(\mathbf{R}^d)} \lesssim \|f\|_{L^{q'}(\mathbf{R}^{d-1})}$$

for any $f \in \mathcal{S}(\mathbf{R}^{d-1})$, where E is the extension function on the paraboloid TODO: Prove This. Thus, despite one theory being local and the other being global, the theories of these two extension operators are roughly equivalent. Every result obtained in one setting can essentially be translated to the other setting.

24.3 Restriction to the Cone

TODO: Conjectured Estimates

TODO: Lorentz Invariant

TODO: Application to the wave equation.

24.4 TODO: Hardy Littlewood Majorant Conjecture

Chapter 25

Almost Orthogonality

It is a standard result of Hilbert space theory that if $\{e_n\}$ are a family of pairwise orthogonal vectors in a Hilbert space H, then Parseval's inequality

$$\|\sum_n a_n e_n\|_H = \left(\sum |a_n|^2\right)^{1/2}$$

holds. In many cases in analysis, we do not have *perfect* orthogonality, but we know that $\langle e_n, e_m \rangle$ is small for most pairs n and m. For general non-orthogonal vectors we do have the identity

$$\|\sum_n a_n e_n\|_H^2 = \sum a_n \overline{a_m} \langle e_n, e_m \rangle.$$

Here is a simple result which exploits this identity.

Lemma 25.1. *Suppose $\{e_n : n \in \mathbf{Z}^d\}$ is a family of vectors such that there is $\delta > 0$ with*

$$\langle e_n, e_m \rangle \leqslant \frac{C}{\langle n - m \rangle^{d+\delta}}.$$

Then

$$\|\sum_n a_n e_n\| \lesssim_\delta C^{1/2} \left(\sum |a_n|^2\right)^{1/2}$$

Proof. Applying Cauchy-Schwartz and Young's convolution inequality, we conclude that

$$
\begin{aligned}
\|\sum_n a_n e_n\|_H^2 &\leqslant C \sum_n |a_n| \sum_m \frac{|a_m|}{\langle n-m\rangle^{d+\delta}} \\
&\leqslant C \left(\sum_n |a_n|^2 \right)^{1/2} \left(\sum_n \left(\sum_m \frac{|a_m|}{\langle n-m\rangle^{d+\delta}} \right)^2 \right)^{1/2} \\
&\lesssim_\delta C \left(\sum_n |a_n|^2 \right).
\end{aligned}
$$

We then just take square roots on both sides of the formula. □

Let us consider some examples of almost orthogonal systems.

Example. *Consider $f \in L^2(\mathbf{R}^d)$ with compact support, and set $e_n = Trans_n f$. The $\langle e_n, e_m \rangle \neq 0$ only if $|n-m| \lesssim 1$, and then we have the trivial bound $\langle e_n, e_m \rangle \leqslant \|f\|_{L^2(\mathbf{R}^d)}^2$. Thus we conclude that*

$$\|\sum a_n e_n\|_{L^2(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^d)} \left(\sum |a_n|^2 \right)^{1/2}.$$

In this situation, another way to prove this inequality is to break up the sum into arithmetic subsequences which have a high enough gap so that the sums are *over functions with disjoint support, in which case we have true orthogonality. Summing up the subsums then gives the result.*

Example. *Given $\phi \in C_c^\infty(\mathbf{R}^d)$, and consider $e_n = Mod_n f$. Then*

$$\langle e_n, e_m \rangle = \int |\phi(x)|^2 e^{2\pi i (n-m)\cdot x}\, dx.$$

We can then apply the theorem of nonstationary phase, which gives

$$|\langle e_n, e_m \rangle \lesssim_k \frac{1}{\langle n-m \rangle^k}$$

for any k. Thus this is an almost orthogonal system.

For almost orthogonal systems, we do *not* have a lower bound

$$\|\sum a_n e_n\|_{L^2(\mathbf{R}^d)} \gtrsim \left(\sum |a_n|^2 \right)^{1/2}.$$

Indeed, the decay does not preclude us from choosing the same two vectors for different indices, which give complete cancellation for appropriately chosen constants.

Example. *Consider a family of vectors $\{e_n\}$ in $L^2(\mathbf{R}^d)$ such that for each n, the support of e_n intersects the support of $O(1)$ other vectors e_m. Do we have almost orthogonality here?*

The most well used almost orthogonality result is the Cotler-Stein lemma, which enables us to bound sums of operators which are almost orthogonal (the bounded operators from a Hilbert space to itself form a Hilbert space).

Theorem 25.2 (Cotler Stein)**.** *Let H_1 and H_2 be Hilbert spaces, and consider a family of bounded operators $\{T_\alpha : H_1 \to H_2\}$. Let $a_{\alpha\beta} = \|T_\alpha T_\beta^*\|$ and let $b_{\alpha\beta} = \|T_\alpha^* T_\beta\|$. Then if*

$$A = \sup_\alpha \sum_\beta \sqrt{a_{\alpha\beta}} \quad \textit{and} \quad B = \sup_\alpha \sum_\beta \sqrt{b_{\alpha\beta}}$$

are both finite, then $\sum T_\alpha$ converges unconditionally in the strong operator topology, and

$$\left\| \sum T_\alpha \right\| \leqslant \sqrt{AB}.$$

One can often extend almost orthogonality to obtain bounds in other L^p spaces via applying interpolation. Here is such a version I encountered in Heo, Nazarov, and Seeger's paper *Radial Fourier Multipliers in High Dimensions*.

Theorem 25.3. *Consider a family of functions $\{f_n : n \in \mathbf{Z}^d\}$ in $L^2(\mathbf{R}^d)$, together with a family of sidelength one cubes $\{Q_n\}$ in $\mathbf{R}^d$ such that $supp(f_n) \subset Q_n$. Suppose*

$$|\langle f_n, f_m \rangle| \leqslant \frac{1}{\langle n - m \rangle^\beta}.$$

for some $\beta \in (0, d)$. Then for $p < 2d/(2d - \beta)$,

$$\left\| \sum a_n f_n \right\|_{L^p(\mathbf{R}^d)} \lesssim_{d,\beta,p} \left(\sum |a_n|^p \right)^{1/p}$$

Remark. If $\beta > d$, then Young's convolution inequality implies that

$$\begin{aligned}
\left\| \sum a_n f_n \right\|_{L^2(\mathbf{R}^d)} &= \left(\sum_{n,m} a_n \overline{a_m} \langle f_n, f_m \rangle \right)^{1/2} \\
&\leqslant \left(\sum_{n,m} \frac{a_n \overline{a_m}}{\langle n - m \rangle^\beta} \right)^{1/2} \\
&\leqslant \left(\sum_n |a_n|^2 \right)^{1/4} \left(\sum_n \left| \sum_m \frac{\overline{a_m}}{\langle n - m \rangle^\beta} \right|^2 \right)^{1/2} \\
&\leqslant \left(\sum_n |a_n|^2 \right)^{1/2} \left(\sum_n \frac{1}{\langle n \rangle^\beta} \right) \\
&\lesssim_{\beta,d} \left(\sum_n |a_n|^2 \right)^{1/2}
\end{aligned}$$

Thus one can view $\beta < d$ as a case where we have some, but not enough orthogonality to prove an L^2 orthogonality bound.

Proof. We view this result as proving the boundedness of the operator

$$a \mapsto \sum_n a_n f_n$$

from $l^p(\mathbf{Z}^d)$ to $L^p(\mathbf{R}^d)$. We shall prove that for $1 \leqslant p \leqslant 2d/(2d - \beta)$, a *restricted strong type inequality holds*, from which the general bound holds by interpolation. It suffices to show that for any finite set of indices $I \subset \mathbf{Z}^d$,

$$\| \sum_{n \in I} f_n \|_{L^p(\mathbf{R}^d)} \lesssim \#(I)^{1/p}.$$

Partition $\mathbf{R}^d$ into an almost disjoint family of sidelength one cubes $\{R_\alpha\}$, define $I_\alpha = \{n \in I : Q_n \cap R_\alpha \neq \varnothing\}$, and set $F_\alpha = \sum_{n \in I_\alpha} f_n$. Now for each $x \in \mathbf{R}^d$, there are at most 3^d indices α such that $F_\alpha(x) \neq 0$. Thus

$$\| \sum_{n \in I} f_n \|_{L^p(\mathbf{R}^d)} = \| \sum_\alpha F_\alpha \|_{L^p(\mathbf{R}^d)} \leqslant 3^d \left(\sum_\alpha \| F_\alpha \|_{L^p(\mathbf{R}^d)}^p \right)^{1/p}.$$

Applying the almost-orthogonality of the functions $\{f_n\}$,

$$\begin{aligned}
\|F_\alpha\|_{L^2(\mathbf{R}^d)}^2 &\leqslant \sum_{n,m \in I_\alpha} \frac{1}{\langle n-m \rangle^\beta} \\
&\leqslant \sum_{n \in I_\alpha} \sum_{|m| \lesssim \#(I_\alpha)^{1/d}} \frac{1}{\langle n-m \rangle^\beta} \\
&\lesssim \#(I_\alpha) \cdot \#(I_\alpha)^{1-\beta/d}
\end{aligned}$$

Thus $\|F_\alpha\|_{L^2(\mathbf{R}^d)} \lesssim \#(I_\alpha)^{1-\beta/2d}$. Combined with the fact that F_α is supported on a sidelength $O(1)$ cube, we conclude that for $0 < p \leqslant 2$, $\|F_\alpha\|_{L^p(\mathbf{R}^d)} \lesssim_p \|F_\alpha\|_{L^2(\mathbf{R}^d)}$. But putting this together means that

$$\left(\sum_\alpha \|F_\alpha\|_{L^p(\mathbf{R}^d)}^p \right)^{1/p} = \left(\sum_\alpha \#(I_\alpha)^{p(1-\beta/2d)} \right)^{1/p}$$

Provided that $p(1-\beta/2d) \leqslant 1$, i.e. $p \leqslant 2d/(2d-\beta)$, we have

$$\sum_\alpha \#(I_\alpha)^{p(1-\beta/2d)} \leqslant \sum \#(I_\alpha) = \#(I)$$

and so we conclude that

$$\left(\sum_\alpha \|F_\alpha\|_{L^p(\mathbf{R}^d)}^p \right)^{1/p} \lesssim \#(I)^{1/p},$$

which completes the proof of the restricted strong type bound. □

Chapter 26

Weighted Estimates

26.1 Hardy-Littlewood Maximal Function

Let us consider a basic weighted estimate for the Hardy-Littlewood maximal function.

Theorem 26.1. *Suppose $w > 0$ is a measurable function. Then for any $f \in L^1_{loc}(\mathbf{R}^d)$, we have the weak type bound*

$$\int_{\mathbf{R}^d} \mathbf{I}\left(|Mf(x)| > \lambda\right) w(x)\, dx \lesssim_d \frac{1}{\lambda} \int_{\mathbf{R}^d} |f(x)| Mw(x)\, dx,$$

and for all $1 \leqslant p \leqslant \infty$, we have the strong type bound

$$\left(\int_{\mathbf{R}^d} |Mf(x)|^p w(x)\, dx \right)^{1/p} \lesssim_{d,p} \left(\int_{\mathbf{R}^d} |f(x)|^p Mw(x)\, dx \right)^{1/p}.$$

Proof. The result automatically follows for $p = \infty$, so by the Stein-Weiss interpolation theorem it suffices to obtain the weak-type bound. We work similarily to the standard Vitali-type approach. Renormalizing, to complete the proof it suffices to show that for any compact set K such that $|Mf(x)| > 1$ for all $x \in K$,

$$\int_K w(x)\, dx \lesssim_d \int_{\mathbf{R}^d} |f(x)| Mw(x)\, dx.$$

Find a disjoint family of balls $B_1, \dots, B_N$ such that $3B_1, \dots, 3B_N$ covers K, and $\int_{B_i} |f(x)|\, dx \gtrsim_d 1$ for each i. Then

$$\int_K w(x)\, dx \leqslant \sum_{i=1}^N \int_{3B_i} w(x)\, dx$$

and so it suffices to show that

$$\int_{3B_i} w(x)\, dx \lesssim_d \int_{B_i} |f(x)| Mw(x)\, dx.$$

But for $x \in B_i$, we have

$$Mw(x) \gtrsim_d \frac{1}{|B_i|} \int_{3B_i} w(y)\, dy$$

from which the claim follows. □

A simple corollary is a vector-valued generalization of the Hardy-Littlewood inequality.

Theorem 26.2. *If $1 < p, q < \infty$ and $\{f_n\}$ are any sequence of functions in $L^1_{loc}(\mathbf{R}^d)$, then*

$$\left\| \left(\sum_n |Mf_n|^p \right)^{1/p} \right\|_{L^q(\mathbf{R}^d)} \lesssim_{d,p,q} \left\| \left(\sum_n |f_n|^p \right)^{1/p} \right\|_{L^q(\mathbf{R}^d)}$$

and

$$\left| \left\{ x : \left(\sum_n |Mf_n(x)|^p \right)^{1/p} \geqslant \lambda \right\} \right| \lesssim_{d,p} \frac{1}{\lambda} \cdot \left\| \left(\sum_n |f_n(x)|^p \right)^{1/p} \right\|_{L^1(\mathbf{R}^d)}.$$

Proof. For $p = q$, the theorem follows from the standard Hardy-Littlewood maximal inequality. For $p < q$ we apply the equivalence between vector-valued bounds and weight bounds. To prove the remaining case, it suffices to prove the weak-type estimate for $p > 1$. By linearization, we may find radii $r_n(y)$ such that

$$|Mf_n(y)| \leqslant \frac{1}{|B(y, r_n(y))|} \int_{\mathbf{R}^d} \psi_n(x,y) f_n(y)\, dy,$$

where $\psi_n(x,y)$ is a smooth bump function which equals one for $x \in B(y, r_n(y))$ and vanishes for $x \notin 2B(y, r_n(y))$. Thus it suffices to obtain a weak-type bound for the vector-valued operator

$$T(\{f_n\})(y) = \left\{ \int_{\mathbf{R}^d} \frac{1}{|B(y, r_n(y))|} \psi_n(x,y) f_n(x)\, dx \right\}.$$

This is a vector-valued kernel operator with kernel $K(x,y)$ the diagonal matrix with entry. TODO: SEE TAO. $\square$

Chapter 27

Bellman Function Methods

It is interesting to ask whether we can obtain bounds of the form

$$\|Mf\|_{L^p(\mathbf{R}^d)} \lesssim_{d,p} \|f\|_{L^p(\mathbf{R}^d)}$$

without employing any interpolation techniques. This is possible, though nontrivial. We begin with a Bellman function approach, which works best in the dyadic scheme, i.e. proving bounds on M_Δ.

The idea here is to perform an *induction on scales*, i.e. to induct on the complexity of the function f. For a fixed $f \in L^p(\mathbf{R}^d)$, our goal is to obtain bounds of the form

$$\left(\int |M_\Delta f(x)|^p \; dx \right)^{1/p} \lesssim \left(\int |f(x)|^p \right)^{1/p}$$

where the implicit constant is independent of p.

We begin by applying some monotone convergence arguments to simplify our analysis. For each $x \in \mathbf{R}^d$, $|M_\Delta f(x)| = \lim_{m \to -\infty} |M_{\geqslant m} f(x)|$, where $M_{\geqslant m}$ is the operator giving a maximal average over all dyadic cubes containing a point with sidelength exceeding 2^m, and the limit is monotone increasing. It follows that for any $f \in L^p(\mathbf{R}^d)$,

$$\|M_\Delta f\|_{L^p(\mathbf{R}^d)} = \lim_{m \to -\infty} \|M_{\geqslant m} f\|_{L^p(\mathbf{R}^d)}.$$

Thus if we can obtain a bound

$$\|M_{\geqslant m} f\|_{L^p(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

with a bound independant of m, we would obtain the required bound on M_Δ. But if we could obtain a bound

$$\|M_{\geqslant 0} f\|_{L^p(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

for all $f \in L^p(\mathbf{R}^d)$, then a rescaling argument, using the fact that

$$M_{\geqslant m} f = \mathrm{Dil}_{1/2^d} M_{\geqslant 0} \mathrm{Dil}_{2^d} f$$

shows that we in fact have

$$\begin{aligned}
\|M_{\geqslant m} f\|_{L^p(\mathbf{R}^d)} &= 2^{d/p} \|M_{\geqslant 0} \mathrm{Dil}_{2^d} f\|_{L^p(\mathbf{R}^d)} \\
&\lesssim 2^{d/p} \|\mathrm{Dil}_{2^d} f\|_{L^p(\mathbf{R}^d)} = \|f\|_{L^p(\mathbf{R}^d)}.
\end{aligned}$$

Thus we need only concentrate on the operator $M_{\geqslant 0}$. Finally, we note we can *localize* our estimates. Given a function f supported on a dyadic cube Q with sidelength 2^n, and given $x \notin Q$, then there exists a smallest value $m_x > n$ such that x is contained in a dyadic cube with sidelength 2^{m_x} which also contains Q. It then follows that

$$(M_{\geqslant 0} f)(x) = \frac{\int_Q |f(y)|\, dy}{2^{dm_x}} = \frac{\|f\|_{L^1(Q)}}{2^{dm_x}}$$

For each $m > n$, if we set $E_m = \{x \in \mathbf{R}^d : m_x = m\}$, then E_m is contained in a dyadic cube of sidelength 2^m, so $|E_m| \leqslant 2^{dm}$. Thus we have

$$\begin{aligned}
\|M_{\geqslant 0} f\|_{L^p(Q^c)} &= \left(\sum_{m=n+1}^\infty \|M_{\geqslant 0} f\|_{L^p(E_m)}^p \right)^{1/p} \\
&\leqslant \left(\sum_{m=n+1}^\infty \left(\|f\|_{L^1(Q)}^p / 2^{dpm} \right) 2^{dm} \right)^{1/p} \\
&\lesssim_{d,p} \|f\|_{L^1(Q)} 2^{dn(1/p-1)} = \|f\|_{L^1(Q)} |Q|^{1/p-1} \leqslant \|f\|_{L^p(Q)}.
\end{aligned}$$

Thus, if we obtained the bound $\|M_{\geqslant 0} f\|_{L^p(Q)} \lesssim \|f\|_{L^p(Q)}$, then we would find

$$\|M_{\geqslant 0} f\|_{L^p(\mathbf{R}^d)} \leqslant \|M_{\geqslant 0} f\|_{L^p(Q)} + \|M_{\geqslant 0} f\|_{L^p(Q^c)} \lesssim \|f\|_{L^p(Q)}.$$

Thus if f is supported on a dyadic cube Q, it suffices to estimate $M_{\geqslant 0} f$ on the support of f. But by a final monotone convergence argument, it suffices to bound such functions, since given any n we can write $[-2^n, 2^n]$ as the almost disjoint union of 2^d sidelength 2^d dyadic cubes $Q_{n,1}, \dots, Q_{n,2^d}$. For any $f \in L^p(\mathbf{R}^d)$, we consider a pointwise limit $f = \lim_{n \to \infty} f_{n,1} + \dots + f_{n,2^d}$, where $f_{n,i}$ is equal to f restricted to $Q_{n,i}$, and the limit is monotone. We also have

$$M_{\geqslant 0} f = \lim_{n \to \infty} M_{\geqslant 0} f_{n,1} + \dots + M_{\geqslant 0} f_{n,2^d}.$$

where the limit is pointwise and monotone, so

$$\begin{aligned} \|M_{\geqslant 0} f\|_{L^p(\mathbf{R}^d)} &= \lim_{n \to \infty} \|M_{\geqslant 0} f_{n,1} + \dots + M_{\geqslant 0} f_{n,2^d}\|_{L^p(\mathbf{R}^d)} \\ &\lesssim \lim_{n \to \infty} \|f_{n,1}\|_{L^p(\mathbf{R}^d)} + \dots + \|f_{n,2^d}\|_{L^p(\mathbf{R}^d)} \lesssim 2^d \|f\|_{L^p(\mathbf{R}^d)}. \end{aligned}$$

Thus, after a technical reduction argument, we now show that we only have to establish a bound

$$\|M_{\geqslant 0} f\|_{L^p(Q)} \lesssim \|f\|_{L^p(Q)},$$

where $f \in L^p(Q)$, Q is a dyadic cube with sidelength $\geqslant 1$, and the implicit constant is independant of Q.

To carry out the inequality, we perform an *induction on scales*. For each $n \geqslant 0$, we let $C(n)$ denote the optimal constant such that for any function $f \in L^p(\mathbf{R}^d)$ supported on a dyadic cube Q of sidelength 2^n,

$$\|M_{\geqslant 0} f\|_{L^p(Q)} \leqslant C(n) \cdot \|f\|_{L^p(Q)}.$$

If $n = 0$, the problem is trivial, since if Q is dyadic with sidelength 1 and $x \in Q$, then

$$M_{\geqslant 0} f = \fint_Q |f(y)|\, dy$$

so $\|M_{\geqslant 0} f\|_{L^p(Q)} = \|f\|_{L^1(Q)}$, and $C(0) = 1$. Our goal is to show that $\sup_{n \geqslant 0} C(n) < \infty$. Given f supported on a cube Q with sidelength 2^n, the cube has 2^d children $Q_1, \dots, Q_{2^d}$ with sidelength 2^{n-1}. If we decompose $f = f_1 + \dots + f_{2^d}$ onto these cubes, then by induction we know that

$$\|M_{\geqslant 0} f_i\|_{L^p(Q_i)} \leqslant C(n-1) \|f_i\|_{L^p(Q_i)}.$$

Now for $x \in Q_i$,

$$(M_{\geqslant 0} f)(x) = \max\left(M_{\geqslant 0} f_i(x), \fint_Q |f(y)|\, dy \right).$$

Thus if $A = \fint_Q |f(y)|\, dy$, then

$$\begin{aligned}\|M_{\geqslant 0} f\|_{L^p(Q)} &= \left(\|M_{\geqslant 0} f\|_{L^p(Q_1)}^p + \dots + \|M_{\geqslant 0} f\|_{L^p(Q_{2^d})}^p \right)^{1/p} \\ &= \left(\|\max(M_{\geqslant 0} f_1, A)\|_{L^p(Q_1)}^p + \dots + \|\max(M_{\geqslant 0} f_{2^d}, A)\|_{L^p(Q_{2^d})}^p \right)^{1/p}\end{aligned}$$

The bound $\max(M_{\geqslant 0} f_i, A) \leqslant M_{\geqslant 0} f_i + A$ gives

$$\|M_{\geqslant 0} f\|_{L^p(Q)} \leqslant C(n-1)\|f\|_{L^p(Q)} + 2^{d/p}|Q|^{1/p} A = (C(n-1) + 2^{d/p})\|f\|_{L^p(Q)}.$$

This gives $C(n) \leqslant C(n-1) + 2^{d/p}$, which is not enough to obtain a uniform bound. The idea here is to include more information in our induction hypothesis which gives control on $\max(M_{\geqslant 0} f_i, A)$. Since Q contains points not in Q_i, we need to treat A as an arbitrary quantity in our hypothesis.

To do this, we introduce *cost functions*. For each $A, B, D > 0$ and any integer $n \geqslant 0$, we let $V_n(A,B,D)$ be the optimal constant such that

$$\|\max(M_{\geqslant 0} f, A)^p\|_{L^p(Q)} \leqslant V_n(A,B,D)$$

For any function f supported on a dyadic cube Q with sidelength 2^n, with

$$\|f\|_{L^1(Q)} = B \quad \text{and} \quad \|f\|_{L^p(Q)} = D.$$

Our goal will be to show $V_n(A,B,D) \lesssim_p 2^{-dn/p} A + D$ which completes the proof. The role of B is subtle, but will soon become apparan. Of course, we have $\|f\|_{L^1(Q)} \leqslant 2^{dn(1-1/p)}\|f\|_{L^p(Q)}$, so we have $V_n(A,B,D) = -\infty$ unless $B \leqslant 2^{dn(1-1/p)} D$.

The recursive inequality gives an inequality for the values $V_n(A,B,D)$. TODO: COMPLETE THIS PROOF.

Chapter 28

TT^* Arguments

The method of TT^* arguments enables us to obtain bounds on an operator T by exploiting cancellation between an operator and it's adjoint. However, this approach only works when establishing L^2 estimates (or at least where one side of an inequality has a norm induced by an inner product). By monotocity, it suffices to consider maximal operators of the form $\max(A_{r_1} f, \dots, A_{r_N} f)$ (provided the implicit constants are independant of N), and by linearization, it suffices to show that for any measurable function $r : \mathbf{R}^d \to \{r_1, \dots, r_N\}$,

$$\left(\int |A_{r(x)} f(x)|^p \, dx \right)^{1/p} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

where the implicit constant is independant of the function r. Thus we consider the linearized operator $M_r : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$ obtained by setting

$$M_r f(y) = (A_{r(y)} f)(y).$$

We see easily that M_r is a kernel operator with kernel

$$K(x,y) = \frac{1}{|B_{r(y)}(y)|} \mathbf{I}(|x-y| \leqslant r(y)).$$

Thus

$$M_r^* g(x) = \int_{\mathbf{R}^d} \frac{\mathbf{I}(|x-y| \leqslant r(y))}{|B_{r(y)}(y)|} g(y) \, dy,$$

and so one can verify that

$$\begin{aligned}|(M_r M_r^* f)(y)| &= \left|\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} \frac{\mathbf{I}(|z-x| \leqslant r(z))\mathbf{I}(|y-x| \leqslant r(y))}{|B(z,r(z))||B(y,r(y))|} f(z)\, dz\, dx\right| \\ &= \left|\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} \frac{\mathbf{I}(|z-x| \leqslant r(z))\mathbf{I}(|y-x| \leqslant r(y))}{|B(z,r(z))||B(y,r(y))|} f(z)\, dx\, dz\right|.\end{aligned}$$

For a fixed z, the integrand in x vanishes unless $|z-y| \leqslant r(y)+r(z)$, and in this case we find

$$\left|\int_{\mathbf{R}^d} \frac{\mathbf{I}(|z-x| \leqslant r(z))\mathbf{I}(|y-x| \leqslant r(y))}{|B(z,r(z))||B(y,r(y))|}\right| \lesssim_d \frac{1}{\max(r(y)^d, r(z)^d)}.$$

Thus we can write

$$\begin{aligned}|(M_r M_r^* f)(y)| &\lesssim_d \int_{\mathbf{R}^d} \left(\int_{\substack{|z-y|\leqslant r(y)+r(z) \\ r(y)\leqslant r(z)}} \frac{|f(z)|}{r(z)^d} + \int_{\substack{|z-y|\leqslant r(y)+r(z) \\ r(y)\geqslant r(z)}} \frac{|f(z)|}{r(y)^d}\, dx \right) \\ &\lesssim_d M_{2r}|f|(y) + M_{2r}^*|f|(y).\end{aligned}$$

But we verify by rescaling that $\|M_{2r}\| = \|M_{2r}^*\| = \|M_r\|$, so

$$\|M_r\|^2 = \|M_r M_r^*\| \lesssim_d \|M_r\|.$$

But this means that $\|M_r\| \lesssim_d 1$, which gives the bound that we required. Thus we find that the Hardy-Littlewood maximal function is bounded from $L^2(\mathbf{R}^d)$ to $L^2(\mathbf{R}^d)$.

Chapter 29

Maximal Averages Over Curves

29.1 Averages over a Parabola

Given any measurable function $f : \mathbf{R}^2 \to \mathbf{C}$ we can consider the maximal average

$$(Mf)(x,y) = \sup_{\varepsilon > 0} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} |f(x+t, y+t)|\, dt.$$

Thus Mf gives a maximal average over parabolas. Our goal is to show $\|Mf\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}$ for $1 < p < \infty$.

It will be convenient to look at the operator

$$\tilde{M}f(x,y) = \sup_{\varepsilon > 0} \frac{1}{2\varepsilon} \int_{\varepsilon/2}^{\varepsilon} |f(x+t, y+t^2)|\, dt.$$

A dyadic decomposition shows that L^p bounds for $\tilde{M}$ imply L^p bounds for M.

For each $k \in \mathbf{Z}$, let $\tilde{M}_k f(x,y) = 2^{-k} \int_{2^k}^{2^{k+1}} f(x+t, y+t^2)\, dt$. Rescaling shows that

$$\|\tilde{M}_k\|_{L^p(\mathbf{R}^2) \to L^p(\mathbf{R}^2)} = \|\tilde{M}_0\|_{L^p(\mathbf{R}^2) \to L^p(\mathbf{R}^2)}$$

so it suffices to focus on $\tilde{M}_0$. The operator is translation invariant and therefore has a Fourier multiplier

$$\tilde{m}(\xi, \eta) = \int_1^2 e^{2\pi i (\xi t + \eta t^2)}\, dt.$$

Note that $\tilde{m}$ is defined by an oscillatory integral with phase $\phi(t) = \xi t + \eta t^2$. We note that $\phi'(t) = \xi + 2\eta t$, so Van der Corput's lemma implies that for $|\xi| \geqslant 10|\eta|$,

$$|\tilde{m}(\xi,\eta)| \lesssim \frac{1}{|\xi|}.$$

Similarily, $\phi''(t) = 2\eta$, so we find

$$|\tilde{m}(\xi,\eta)| \lesssim \frac{1}{|\eta|^{1/2}}.$$

If $f \in L^2(\mathbf{R}^2)$ and $\widehat{f}$ is supported on the region

$$E_0 = \{(\xi,\eta) : |\eta| \geqslant 1 \text{ or } |\xi| \leqslant 1 \text{ and} |\eta| \geqslant 10\}$$

then $\|\tilde{m}\|_{L^\infty(E_0)} \lesssim 1$ and so

$$\|\tilde{M}_0 f\|_{L^2(\mathbf{R}^2)} = \|\tilde{m}\widehat{f}\|_{L^2(\mathbf{R}^2)} \lesssim \|\widehat{f}\|_{L^2(\mathbf{R}^2)} = \|f\|_{L^2(\mathbf{R}^2)}.$$

On the other hand, we can decompose $\mathbf{R}^2 - E_0$ into
suppose $\widehat{f}$ is supported on the region

$$E_1 = \{(\xi,\eta) : |\xi| \leqslant 1 \text{ and } |\eta| \leqslant 10\}.$$

Then the uncertainty principle implies that f is roughly constant on scales $|\Delta x| \leqslant 1$ and $|\Delta y| \leqslant 1/10$, which should imply good bounds for the maximal average. More precisely, $\widehat{f}$ is supported on the ellipsoid

$$\left\{(\xi,\eta) \in \mathbf{R}^2 : \xi^2/2 + \eta^2/20 \leqslant 1\right\}.$$

Thus the uncertainty principle implies that f is roughly constant on scales $|\Delta x|^2 \leqslant 1/2$ and $|\Delta y|^2 \leqslant 1/20$,

s

$$\phi(x) = \frac{1}{(1 + 2x^2 + 20y^2)^N}$$

Part IV

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 30

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* $|p^{-m}q|_p = p^m$*.*

30.1 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 30.1. *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 30.2. *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 30.3. *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 30.4. *Every Kolmogorov topological group is Hausdorff.*

Theorem 30.5. *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 30.6. *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.

30.2 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 30.7. *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 30.8. *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.

30.3 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

$$L_U = \{(x,y) : yx^{-1} \in U\} \qquad R_U = \{(x,y) : x^{-1}y \in U\}$$

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 30.9. *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 30.10. *All maps with compact support are uniformly continuous.*

Corollary 30.11. *Uniform continuity on compact groups is invariant of the uniform structure chosen.*

30.4 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.

30.5 Topological Groups arising from Normal subgroups

Let G be a group, and $\mathcal{N}$ a family of normal subgroups closed under intersection. If we interpret $\mathcal{N}$ as a neighbourhood base at the origin, the resulting topology gives G the structure of a totally disconnected topological group, which is Hausdorff if and only if $\bigcap \mathcal{N} = \{e\}$. First note that $g_i \to g$ if g_i is eventually in gN, for every $N \in \mathcal{N}$, which implies

$g_i^{-1} \in Ng^{-1} = g^{-1}N$, hence inversion is continuous. Furthermore, if h_i is eventually in hN, then $g_i h_i \in gNhN = ghN$, so multiplication is continuous. Finally note that $N^c = \bigcup_{g \neq e} gN$ is open, so that every open set is closed.

Example. *Consider $\mathcal{N} = \{\mathbf{Z}, 2\mathbf{Z}, 3\mathbf{Z}, \dots\}$. Then $\mathcal{N}$ induces a Hausdorff topology on $\mathbf{Z}$, such that $g_i \to g$, if and only if g_i is eventually in $g + n\mathbf{Z}$ for all n. In this topology, the series $1, 2, 3, \dots$ converges to zero!*

This example gives us a novel proof, due to Furstenburg, that there are infinitely many primes. Suppose that there were only finitely many, $\{p_1, p_2, \dots, p_n\}$. By the fundamental theorem of arithmetic,

$$\{-1, 1\} = (\mathbf{Z}p_1)^c \cap \dots \cap (\mathbf{Z}p_n)^c$$

and is therefore an open set. But this is clearly not the case as open sets must contain infinite sequences.

Chapter 31

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 31.1. *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

$$h = f_1 + f_2 + \varepsilon g$$

$$h_1 = f_1/h \qquad h_2 = f_2/h$$

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

$$|h_1(xy) - h_1(y)| \leqslant \varepsilon \qquad |h_2(xy) - h_2(y)| < \varepsilon$$

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 $\mathrm{supp}(\phi) \subset V$.

Now we have estimates on how well I_ϕ approximates a linear function, so we can apply a limiting process. Consider the product

$$X = \prod_{f \in C_0^+(G)} [(k:f)^{-1}, (k:f_0)]$$

a compact space, by Tychonoff's theorem, consisting of $F : C_c^+(G) \to \mathbf{R}$ such that $(k:f)^{-1} \leqslant F(f) \leqslant (f:k)$. For each neighbourhood V of the identity, let $K(V)$ be the closure of the set of I_ϕ such that $\mathrm{supp}(\phi) \subset V$. Then the set of all $K(V)$ has the finite intersection property, so we conclude there is some $I : C_c^+(G) \to \mathbf{R}$ contained in $\bigcap K(V)$. This means that every neighbourhood of I contains I_ϕ with $\mathrm{supp}(\phi) \subset V$, for all ϕ. This means that if $f_1, f_2 \in C_c^+(G)$, $\varepsilon > 0$, and V is arbitrary, there is ϕ with $\mathrm{supp}(\phi) \subset V$, and

$$|I(f_1) - I_\phi(f_1)| < \varepsilon \quad |I(f_2) - I_\phi(f_2)| < \varepsilon$$

$$|I(f_1+f_2) - I_\phi(f_1+f_2)| < \varepsilon$$

this implies that if V is chosen small enough, then

$$|I(f_1+f_2) - (I(f_1) - I(f_2))| \leqslant 2\varepsilon + |I_\phi(f_1+f_2) - (I_\phi(f_1) + I_\phi(f_2))| < 3\varepsilon$$

Taking $\varepsilon \to 0$, we conclude I is linear. Similar limiting arguments show that I is homogenous of degree 1, and commutes with all left translations. We conclude the extension of I to a linear functional on $C_0(G)$ is well defined, and the Radon measure obtained by the Riesz representation theorem is a Haar measure. □

We shall prove that the Haar measure is unique, but first we show an incredibly useful regularity property.

Proposition 31.2. *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. □

Theorem 31.3. *Haar measures are unique up to a multiplicative constant.*

Proof. Let μ and ν be Haar measures. Fix a compact neighbourhood V of the identity. If $f, g \in C_c^+(G)$, consider the compact sets

$$A = \mathrm{supp}(f)V \cup V\mathrm{supp}(f) \qquad B = \mathrm{supp}(g)V \cup V\mathrm{supp}(g)$$

Then the functions $F_y(x) = f(xy) - f(yx)$ and $G_y(x) = g(xy) - g(yx)$ are supported on A and B. There is a neighbourhood $W \subset V$ of the identity such that $\|F_y\|_\infty, \|G_y\|_\infty < \varepsilon$ if $y \in W$. Now find $h \in C_c^+(G)$ with $h(x) = h(x^{-1})$ and $\mathrm{supp}(h) \subset W$ (take $h(x) = k(x)k(x^{-1})$ for some function $k \in C_c^+(G)$ with $\mathrm{supp}(k) \subset W$, and $k = 1$ on a symmetric neighbourhood of the origin). Then

$$\begin{aligned}\left(\int h d\mu\right)\left(\int f d\lambda\right) &= \int h(y) f(x) d\mu(y) d\lambda(x)\\ &= \int h(y) f(yx) d\mu(y) d\lambda(x)\end{aligned}$$

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 $\mu(A) = \nu(A)$.

31.1 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.

31.2 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 31.4. *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 31.5. *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.

31.3 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 $f d\mu * g d\mu = (f * g)d\mu$. What's more,

$$\|f\|_1 = \|f d\mu\|$$

If $\nu \in M(G)$, then we can still define $\nu * f \in L^1(G)$

$$(\nu * f)(x) = \int f(y^{-1}x)d\mu(y)$$

which holds since

$$\int \phi d(\nu * f\mu) = \int \phi(yx)f(x)d\nu(y)d\mu(x) = \int \phi(x)f(y^{-1}x)d\nu(y)d\mu(x)$$

If G is unimodular, then we also find

$$\int \phi d(f\mu * \nu) = \int \phi(yx)f(y)d\mu(y)d\nu(x) = \int \phi(x)f(y)d\mu(y)d\nu(y^{-1}x)$$

So we let $f * \mu(x) = \int f(y)d\mu(y^{-1}x)$.

Theorem 31.6. *$L^1(G)$ 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. □

If we wish to integrate by right multiplication instead of left multiplication, we find by the substitution $y \mapsto xy$ that

$$\begin{aligned}(f * g)(x) &= \int f(y) g(y^{-1}x) dy \\ &= \int\int f(xy) g(y^{-1}) dy \\ &= \int\int \frac{f(xy^{-1}) g(y)}{\Delta(y)} dy\end{aligned}$$

Observe that

$$f * g = \int f(y) L_{y^{-1}} g\, dy$$

which can be interpreted as a vector valued integral, since for $\phi \in L^\infty(\mu)$,

$$\int (f * g)(x) \phi(x) dx = \int f(y) g(y^{-1}x) \phi(x) dx dy$$

so we can see convolution as a generalized 'averaging' of translate of g with respect to the values of f. If G is commutative, this is the same as the averaging of translates of f, but not in the noncommutative case. It then easily follows from operator computations $L_z(f * g) = (L_z f) * g$, and $R_z(f * g) = f * (R_z g)$, or from the fact that

$$(f * g)(zx) = \int f(y)g(y^{-1}zx)dy = \int f(zy)g(y^{-1}x)dy = [(L_z f) * g](x)$$

$$(f * g)(xz) = \int f(y)g(y^{-1}xz)dy = [f * (R_z g)](x)$$

Convolution can also be applied to the other L^p spaces, but we have to be a bit more careful with our integration.

Theorem 31.7. *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 31.8. *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 31.9. *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 31.10. *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 31.11. *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. □

We have already obtained enough information to characterize the closed ideals of $L^1(G)$.

Theorem 31.12. *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. □

31.4 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 31.13. *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.

31.5 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 31.14. *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 $\phi(yVy_0^{-1}) \subset \phi(VVy_0^{-1})$, so if we fix a compact U, and find V with $V^3 \subset U$, we have shown $\phi(U)$ is open in X. □

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 31.15. *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 31.16. *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

$$Pf = \frac{\phi}{Pg} Pg = \phi$$

$\square$

Lemma 31.17. *If $\phi \in C_c(G/H)$, there is $f \in C_c(G)$ with $Pf = \phi$, and $\pi(\text{supp} f) = \text{supp}(\phi)$, and also $f \geqslant 0$ if $\phi \geqslant 0$.*

Proof. There exists $g \geqslant 0$ in $C_c(G/H)$ with $Pg = 1$ on $\text{supp}(\phi)$, and then $f = (\phi \circ \pi)g$ satisfies the properties of the theorem. $\square$

We can now provide conditions on the existence of a measure on G/H.

Theorem 31.18. *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.

31.6 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 32

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 32.1. *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

$$\begin{aligned}
\varphi(f * g) &= \int \int \frac{f(y)g(y^{-1}x)}{\phi(x)} dy dx \\
&= \int \int \frac{f(y)g(x)}{\phi(y)\phi(x)} dy dx \\
&= \int \frac{f(y)}{\phi(y)} dy \int \frac{g(x)}{\phi(x)} dx
\end{aligned}$$

Since ϕ is continuous, there is a compact subset K of G where $\phi > \varepsilon$ for some $\varepsilon > 0$, and we may then choose a positive f supported on K in such a way that $\varphi(f)$ is non-zero.

The converse results from applying the duality theory of the L^p spaces. Any character on $L^1(G)$ is a linear functional, hence is of the form

$$f \mapsto \int f(x)\phi(x)dx$$

for some $\phi \in L^\infty(G)$. Now

$$\begin{aligned}\int\int f(y)g(x)\phi(yx)dydx &= \int\int f(y)g(y^{-1}x)\phi(x)dydx \\ &= \int f(x)\phi(x)dx \int g(y)\phi(y)dy \\ &= \int f(x)g(y)\phi(x)\phi(y)dxdy\end{aligned}$$

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 32.2. *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 32.3. *$L^1(G)$ 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 ∎

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 33

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 34

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 35

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 35.1. *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 35.2. *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 35.3. *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 35.4 (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 35.5 (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 $1 < p < \infty$.

35.1 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

$$\|f\|_{pq}^s = \|\psi * f\|_p + \left(\sum_{n=1}^{\infty} (2^{sn}\|\varphi_k * f\|_p)^q \right)^{1/q}$$

The convolution $\varphi_n * f$ essentially captures the portion of f whose frequencies lie in the annulus $2^{n-1} \leqslant |\xi| \leqslant 2^{n+1}$

35.2 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 36

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 36.1 (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 36.2 (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 36.3 (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 36.4 (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 V

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 37

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 37.1. *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. □

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 37.2. *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. $\square$

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 37.3. *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 37.4. *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.

37.1 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

37.2 Local Orthogonality