Harmonic Analysis

Jacob Denson

February 27, 2024

Disclaimer

This is a set of notes which I have written over the past decade of learning Harmonic analysis, since I was an undergraduate, as a way to summarize the information I've learnt over the years for myself. The notes were mainly made for my benefit. Covering so much material, there is not as much of a central theme to guide the notes, which might make them hard to learn from. I'd strongly recommend using another source to learn a particular topic, especially those that most helped me learn the content of these notes:

- Tao's Blog Notes.

- Tao's 247A Notes.

- C. Demeter, *Fourier Restriction, Decoupling, and Applications*.

- J.J. Duistermaat, *Fourier Integral Operators*.

- J. Duoandikoetxea, *Fourier Analysis*.

- G.B. Folland, *A Course in Abstract Harmonic Analysis*.

- L. Grafakos, *Classical Fourier Analysis* and *Modern Fourier Analysis*.

- L. Hörmander, *The Analysis of Linear Partial Differential Operators* (4 Volumes).

- L. Hörmander, *Riemmanian Geometry*.

- Y. Katznelson, *An Introduction to Harmonic Analysis*.

- T.W. Körner, *Fourier Analysis*.

- T.W. Körner, *Topics in Fourier and Complex Analysis (Autumn 2009) Notes*.

- T.W. Körner, *Topics in Analysis (Autumn 2010) Notes*.

- S. Krantz, *A Panorama of Harmonic Analysis*.

- R. Lyons, *Seventy Years of Rajchman Measures*.

- R. O'Donnell, *Analysis of Boolean Functions*.

- C. Muscalu and W. Schlag, *Classical and Multilinear Harmonic Analysis*.

- A. Seeger, *Notes on Radial and Quasiradial Fourier Multipliers*.

- C. Sogge, *Fourier Integrals in Classical Analysis*.

- C. Sogge, *Hangzhou Lectures on Eigenfunctions of the Laplacian*.

- E. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*.

- E. Stein, *Singular Integrals and Differentiability Properties of Functions*.

- E. Stein and R. Shakarchi, *Princeton Lectures in Analysis* (4 Volumes).

- E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*.

- B. Street, *Multi-Parameter Singular Integrals* and *Maximal Subellipticity*.

- T. Tao, *Nonlinear Dispersive Equations*.

- L. Tartar, *An Introduction to Sobolev Spaces and Interpolation Spaces*.

- T. Wolff, *Lectures in Harmonic Analysis*.

- P. Zorin-Kranich, *Lecture Notes on Fourier Decoupling*.

- S. Zelditch, *Eigenfunctions of the Laplacian on a Riemannian Manifold*.

- J.F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators* (Two Volumes).

- M. Zworski, *Semiclassical Analysis*.

- A. Zygmund's *Trigonometric Series* (Two Volumes).

Table Of Contents

Part I

Classical Fourier Analysis

Deep mathematical knowledge often arises hand in hand with the recognition of symmetry. The foundations of harmonic analysis prove no exception to this rule, where we attempt to break down mathematical objects into superpositions of more symmetric objects. In the mid 18th century, problems in mathematical physics led D. Bernoulli, D'Alembert, Lagrange, and Euler to consider periodic functions representable uniquely 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 work, *Mémoire sur la propagation de la Chaleur dans les corps solides*, submitted to the Institute de France in 1807, 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 the analogous statement on the real line, that a function f on the real line can be uniquely 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 a general class of functions. Their work is responsible for most of the modern subject of analysis we know today. In particular, their work clarified the subtle differences between convergence properties of functions. If we assume the partial sums of the Fourier series as converging in a pointwise sense, then it is not true that every function has a *unique* expansion as a trigonometric series. If we assume that the Fourier series converges *uniformly*, then a trigonometric expansion, if it exists, must be unique. But we lose some generality; our functions must be continuous if we can ever hope to have some uniform convergence, and one does not even obtain such an expansion for continuous functions. We must therefore assume some smoothness on the functions f we work with. Thus we must introduce more subtle methods of convergence, such as convergence 'on average', which leads to the study of L^p spaces. For $1 < p < \infty$, we will see that all functions in $L^p(\mathbb{R}^d)$ can be expanded uniquely in a trigonometric series that converges in a pointwise sense. Other results can be replacing by interpreting the convergence of the Fourier series as a limit of 'generalized' partial sums, which allow us to recover some uniform convergence results, and convergence in $L^1(\mathbb{R}^d)$.

Chapter 1

Introduction

The trigonometric functions

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

are fundamental to the analysis of oscillatory behaviour. 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 a limit of linear combinations of these trigonometric functions. In the periodic setting, we fix a function $f : \mathbb{R} \to \mathbb{C}$ such that $f(x + 1) = f(x)$ for all $x \in \mathbb{R}$, 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 $f : \mathbb{R} \to \mathbb{C}$, and try 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

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

It was natural, in several mathematical and physical questions in the 19th century, to find *trigonometric expansions* for a given function f, i.e. to find a family of coefficients $\{A_m\}$, $\{B_m\}$, and C such that for each $t \in \mathbb{R}$,

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

where the latter sum is assumed to be convergent pointwise for each $t \in \mathbb{R}$. 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 independent of whether the series actually does converge pointwise 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

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

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

$$\int_0^1 \sin(2\pi mt) \cos(2\pi nt) = 0.$$

This means that for a finite trigonometric sum

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

we have

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

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

The important thing to notice is that these quantities are defined even if f is given a priori, without recourse to a trigonometric polynomial. Thus given *any* periodic integrable function f, we might hope to find a trigonometric expansion for f by considering the trigonometric series given by the values $\{A_m\}$, $\{B_m\}$, and C defined as above. Unlike when f is a trigonometric polynomial, for a general periodic integrable function f, infinitely many of the values $\{A_m\}$, $\{B_m\}$ and C can be nonzero.

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 each $\xi \in \mathbb{R}$, and each $t \in \mathbb{R}$, we consider the function

$$e_\xi(t) = e^{2\pi\xi it}.$$

For each integer $n \in \mathbb{Z}$, 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

$$f \sim \sum_{n \in \mathbb{Z}} D_n e^{2\pi nit}$$

for an integrable, periodic function f, i.e. by setting

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

Euler's formula $e^{nit} = \cos(nt) + i\sin(nt)$ gives an equivalence between expansions in complex exponentials, and expansions in sines and cosines. Thus the values $\{A_m, B_m, C : m \geqslant 0\}$ can be recovered from the values of $\{D_m : m \in \mathbb{Z}\}$, and vice versa. Moreover, we have

$$C + \sum_{m=1}^N A_m \cos(2\pi mt) + B_m \sin(2\pi mt) = \sum_{|n| \leqslant N} D_n e_\xi(t).$$

Thus the convergence properties of these series is identical, provided we take the *symmetric* partial sums of complex exponentials. Because of it's elegance, unifying the three families of coefficients, the expansion by complex exponentials over cosines and sines is the most standard way we discuss Fourier series today.

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

$$\sum_{m \in \mathbb{Z}} C_m e_m.$$

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 $m \in \mathbb{Z}$, 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.

1.3 The Fourier Transform

For a non-periodic function $f : \mathbb{R} \to \mathbb{C}$, a natural analogue to an expansion in a trigonometric series

$$\sum C_m e_m(t)$$

would be an expansion as a *trigonometric integral*, i.e. a formal expression

$$\int a(\xi)e_\xi(t).$$

Let us say a function f *admits a trigonometric expansion* if there exists a locally integrable function a such that

$$f(t) = \lim_{N \to \infty} \int_{|\xi| \leqslant N} a(\xi)e_\xi(t)\, d\xi.$$

Again, a characterization of the type of functions which admit a trigonometric expansion is very difficult, but we can still get away with trying to find a method to obtain an expansion using some kind of orthogonality. The functions $\{e_\xi\}$ are

not even integrable on $\mathbb{R}$, let alone orthogonal, which prevents a direct analogy with the previous approach. But if f is an integrable function on the real line, then for each $R > 0$, we can consider the functions $g_R : [0, 1] \to \mathbb{C}$ by setting $g_R(s) = f(R(s - 1/2))$. We note that formally, we calculate that for $R \geqslant |t|$,

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

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 $\xi \in \mathbb{R}$ 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 of the real line 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 $f : \mathbb{R}^d \to \mathbb{C}$ is periodic, in the sense that

$$
f(x + n) = f(x) \quad \text{for each } x \in \mathbb{R}^d \text{ and } n \in \mathbb{Z}^d,
$$

then we can consider the natural higher dimensional Fourier series

$$f(x) \sim \sum_{n \in \mathbb{Z}^d} \hat{f}(n) e_n(x)$$

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

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

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

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

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

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

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 basic theory of convergence in higher dimensions requires much greater regularity in higher dimensions than the one dimensional theory, and many fundamental questions about the convergence of Fourier series here have much more nuance than in the lower dimensional theory, with many questions about such questions still open today.

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, the general theory of which we shall study later.

Example. *Consider the function $f : [0, \pi] \to \mathbb{R}$ 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

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

10

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 all $x \in [0, \pi]$. If we extend the domain of f to $[-\pi, \pi]$ by making f odd, then

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

In this case, we still have

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

This sum converges absolutely and uniformly for all $x \in [-\pi, \pi]$.

Example. *The tent function*

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

is even, and therefore has a purely real Fourier expansion

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

Thus we obtain an expansion

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

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

Example. *Consider the characteristic function*

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

11

for $(a, b) \subset [-\pi, \pi]$. 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

$$\chi_{(a,b)}(x) = \frac{b - a}{2\pi} + \sum_{n \neq 0} \frac{e_n(-a) - e_n(-b)}{2\pi i n} e_n(x)$$

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

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

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

so that it suffices to show $\sum |\sin(nx)| n^{-1} = \infty$ for every $x \notin \pi \mathbb{Z}$. 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 $m \in \mathbb{Z}$, then

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

Thus if $nx \in (-x/2, x/2) + \pi \mathbb{Z}$, $(n + 1)x \notin (-x/2, x/2) + \pi \mathbb{Z}$. For y outside of $(-x/2, x/2) + \pi \mathbb{Z}$, 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

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

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

12

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 $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, so that a function $f : \mathbb{T} \to \mathbb{C}$ can be identified with a complex-valued periodic function on the real line. We then have a metric on $\mathbb{T}$ given by setting $d(t, s) = |t - s|$, where $|t| = \min_{n \in \mathbb{Z}} |t + n|$ for $t \in \mathbb{T}$. The Lebesgue measure on $\mathbb{R}$ induces a natural Borel measure on $\mathbb{T}$, such that for any periodic function $f : \mathbb{T} \to \mathbb{C}$,

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

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

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

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

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

If the sequence $\{\hat{f}(n) : n \in \mathbb{Z}^d\}$ is absolutely summable, then the formal trigonometric series

$$\sum_n \hat{f}(n) e^{2\pi i n t}$$

can be interpreted non-formally by taking the convergence pointwise for each t, and we would then hope that

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

It turns out that, under the assumption that the coefficients $\{\hat{f}(n)\}$ are absolutely summable, this equation does hold provided f is a continuous function on $\mathbb{T}^d$. Absolute summability will follow if we assume that f is differentiable to a suitably high order. 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 fundamental, if often underappreciated property of the Fourier transform is *linearity*: For any two functions f and g, if $h = f + g$, then $\hat{h} = \hat{f} + \hat{g}$. Linearity is *essential* to most methods in this book; many problems about nonlinear transforms are intractable to current methods. 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 $f \in L^1(\mathbb{T}^d)$, define $\text{Conj}\{f\}$ and $\text{Ref}\{f\} \in L^1(\mathbb{T}^d)$ by setting

$$\text{Conj}\{f\}(x) = \overline{f(x)} \quad \text{and} \quad \text{Ref}\{f\}(x) = f(-x).$$

  Then

$$\widehat{\text{Conj}\{f\}} = (\text{Conj} \circ \text{Ref})\{\hat{f}\} \quad \text{and} \quad \widehat{\text{Ref}\{f\}} = \text{Ref}\{\hat{f}\}.$$

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

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

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

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

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

15

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

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

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

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

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

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

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

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

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

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

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

Moreover,

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

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

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

Thus we get a sine expansion, and moreover,

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

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.

16

2.2 Unique Representation

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 $f \in L^1(\mathbb{T}^d)$, 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 $L^1(\mathbb{T}^d)$, with functions identified if they are equal almost everywhere. Nonetheless, if $f \in C(\mathbb{T}^d)$ 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 $f \in L^1(\mathbb{T}^d)$. If $\hat{f}(n) = 0$ for all $n \in \mathbb{Z}^d$, then f vanishes at all it's continuity points.*

Proof. It suffices to prove that if $f \in L^1(\mathbb{T}^d)$ 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

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

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

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

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 $f, g \in C(\mathbb{T}^d)$ and $\hat{f} = \hat{g}$, then $f = g$.*

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

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

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

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

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

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

In the next section we will show that if $f \in C^m(\mathbb{T}^d)$, 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 $f \in C^\infty(\mathbb{T})$, then using the Fourier series equation for the derivative of a function, for each multi-index α, we conclude that for all $x \in \mathbb{R}^d$

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

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

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

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 $f \in L^1(\mathbb{T})$, it is most natural to consider the partial sums

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

19

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*

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

We will see that these two types of summations can have subtle differences in the higher dimensional theory. Right now, we consider a family of operators that contain both of these operators; we consider an increasing family of sets $\{E_N\}$ in $\mathbb{Z}^d$ with $\lim_{N \to \infty} E_N = \mathbb{Z}^d$, and we then define

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

for $f \in L^1(\mathbb{T}^d)$. 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 $f \in C(\mathbb{T}^d)$ and $\widehat{f} \in L^1(\mathbb{Z}^d)$. Then*

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

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

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

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

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 $\|f - S_N f\|_{L^\infty(\mathbb{T}^d)} \leqslant \varepsilon$. $\qquad\qquad\square$

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 $C(\mathbb{T})$ with integrable Fourier coefficients, the convergence of $\|f - S_N f\|_{L^\infty(\mathbb{T}^d)}$ as $N \to \infty$ can be *as slow as any convergent sequence*.

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

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 $f \in C(\mathbb{T}^d)$, and that $\hat{f}(n_k) = a_{n_k}$ for each $1 \leqslant k \leqslant \infty$. $\qquad\square$

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 $\|f\|_{L^\infty(\mathbb{T}^d)}$ is small (i.e. we have uniform control on the magnitude of f) rather than just if $\|f\|_{L^1(\mathbb{T}^d)}$ 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*

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

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 $f \in C^{m+1}(\mathbb{T}^d)$, then we will see

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

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 $\| \cdot \|_{L^1(\mathbb{T}^d)}$ to analyze the space $C(\mathbb{T}^d)$, most questions are equivalent to questions about the completion of $C(\mathbb{T}^d)$, i.e. the space $L^1(\mathbb{T}^d)$ 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 $\|\widehat{f}\|_{L^q(\mathbb{Z}^d)} \lesssim \|f\|_{L^p(\mathbb{T}^d)}$ 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 $L^p(\mathbb{T}^d)$. The first inequality we give is trivial, but tight for general L^1 functions, i.e. for the functions $f(t) = e_n(t)$.

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

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

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

which was the required bound. □

22

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

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

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* $f \in L^2(\mathbb{T}^d)$, $\|\widehat{f}\|_{L^2(\mathbb{Z}^d)} = \|f\|_{L^2(\mathbb{T}^d)}$.

Proof. With respect to the normalized inner product on the space $L^2(\mathbb{T}^d)$, the calculations of the last chapter tell us that the exponentials $\{e_n : n \in \mathbb{Z}^d\}$ are an orthonormal family of functions, in the sense that for distinct pair $n, m \in \mathbb{Z}^d$, $(e_n, e_m) = 0$ and $(e_n, e_n) = 1$. Since $\widehat{f}(n) = (f, e_n)$, we apply Bessel's inequality to conclude

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

The exponentials $\{e_n\}$ are actually an orthonormal basis for $L^2(\mathbb{T}^d)$; 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 $f \in C^\infty(\mathbb{T}^d)$, then we have shown that if $(f, e_n) = 0$ for all $n \in \mathbb{Z}^d$, then $f = 0$. But $S_N f$ converges to f in $L^2(\mathbb{T}^d)$ (it actually converges uniformly), and so

$$\|f\|_{L^2(\mathbb{T}^d)} = \lim_{N\to\infty} \|S_N f\|_{L^2(\mathbb{T}^d)} = \lim_{N\to\infty} \left(\sum_{n\in E_N} |\widehat{f}(n)|^2 \right)^{1/2}$$

$$= \left(\sum_{n\in\mathbb{Z}^d} |\widehat{f}(n)|^2 \right)^{1/2} = \|\widehat{f}\|_{L^2(\mathbb{Z}^d)}.$$

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

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

Taking $\varepsilon \to 0$ completes the proof. $\qquad\qquad\square$

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

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

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

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 $f \in L^1(\mathbb{T}^d)$, 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

$$|\Lambda_n f| \leqslant \|\widehat{f}\|_{L^\infty(\mathbb{T})} \leqslant \|f\|_{L^1(\mathbb{T})},$$

which shows that the sequence of functionals $\{\Lambda_n\}$ are uniformly bounded as linear functions on $L^1(\mathbb{T}^d)$. Since $\lim_{|n| \to \infty} \Lambda_n(f) = 0$ for any step function f, and the step functions are dense in $L^1(\mathbb{T}^d)$, we conclude that

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

for all $f \in L^1(\mathbb{T}^d)$. $\qquad\square$

Even though the Fourier series of any step function decays at a rate $O(1/n)$, it is *not* true that a general Fourier series decays at a rate of $O(1/n)$. 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 $f \in C^m(\mathbb{T}^d)$, then for each $n \in \mathbb{Z}^d$,*

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

Proof. We have

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

Thus

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

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

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

On the other hand, if $|\hat{f}(n)| \lesssim 1/|n|^{d+m}$, it is easy to see from the pointwise convergence of the Fourier series that $f \in C^m(\mathbb{R}^d)$. 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 $x, h \in \mathbb{T}^d$. We define

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

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

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

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

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

26

so

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

Thus taking in absolute values and applying Hölder continuity gives

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

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 $f \in L^1(\mathbb{T}^d)$. Then*

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

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

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

Then for $|h| < 1$,

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

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

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

$$\lesssim_d 2A|h|^\alpha.$$

Combining these two calculations shows that

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

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

27

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

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

If μ is absolutely continuous with respect to the normalized Lebesgue measure on $\mathbb{T}$, 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

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

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 $\mathbb{T}^d$ is absolutely continuous with respect to the Lebesgue measure if and only if

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

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

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

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

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

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

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

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

Theorem 2.14. *Convolution has the following properties:*

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

- *If $f \in L^p(\mathbb{T}^d)$ and $g \in L^q(\mathbb{T}^d)$, 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*

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

  This is known as Young's inequality for convolutions.

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

- *If $f, g \in L^1(\mathbb{T})$, then $\widehat{f * g} = \hat{f}\hat{g}$.*

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

- *If f is supported on $E \subset \mathbb{T}^d$, and g on $F \subset \mathbb{T}^d$, then $f * g$ is supported on $E + F$.*

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

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

The right hand side is a bound independent of t and converges to zero as $h \to 0$, so $f * g$ is uniformly continuous. Applying Hölder's inequality again gives

29

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

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

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

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

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

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

and this is exactly the identity required. To calculate the weak derivative of $f * g$,

we fix $\phi \in C^\infty(\mathbb{T}^d)$, and calculate using two applications of Fubini's theorem that

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

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

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

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

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

We know that suitably smooth functions have convergent Fourier series. The advantage of convolution is if we want to study the properties of any integrable function f, then for any function $u \in C^\infty(\mathbb{T})$, $f * u$ will be smooth, and if we choose u appropriately, we can let $f * u$ approximate f. In terms of Fourier analysis, provided that $\widehat{u}$ is close to one, $(f * u)^\wedge$ will be close to $\widehat{f}$, and so we might expect $f * u$ to be close to f. If we can establish the convergence properties on the convolution $f * g$, then we can probably obtain results about f. From the frequency side, $\sum \widehat{f}(n)e_n$ might not converge, but $\sum a_n \widehat{f}(n)e_n$ might converge for a suitably fast decaying sequence a_n. But if a_n is close to one, this sequence might still reflect properties of the original sequence.

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

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

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 $\|K_n\|_{L^\infty(\mathbb{T}^d)} \lesssim n^d$, we say it is an *approximation to the identity*.

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

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

and to define

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

Then $\|K_n\|_{L^1(\mathbb{T})} = 1$ for all $n > 0$, and K_n is eventually supported on every small ball around the origin, which implies (2.1). If $K \in L^\infty(\mathbb{T}^d)$, 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 $f \in C(\mathbb{T}^d)$, and $K_n * f$ converges to f in $L^p(\mathbb{T}^d)$ if $f \in L^p(\mathbb{T}^d)$, 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 $L^p(\mathbb{T})$. 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 $C(\mathbb{T}^d)$. But a density argument allows us

32

to conclude that $K_n * f \to f$ in $L^p(\mathbb{T})$ for each $f \in L^p(\mathbb{T}^d)$, for $1 \leqslant p < \infty$. To obtain pointwise convergence for t in the Lebesgue set of f, we calculate

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

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. $\qquad\square$

As mentioned above, if $u \in C^\infty(\mathbb{T})$, then for any $f \in L^1(\mathbb{T})$, $f * u \in C^\infty(\mathbb{T})$, and we have $D^\alpha(f * u) = f * D^\alpha u$. Thus we can use Young's inequality to bound $D^\alpha(f * u)$ in terms of f and $D^\alpha u$. Thus in some contexts, it is useful to find out how smooth we can make u, while still being able to approximate a non-smooth function f with $f * u$. If we fix $u \in C^\infty(\mathbb{T})$ supported on an interval of length $1/4$ centered at the origin, then for $0 < \varepsilon < 1$, we can define $u_\varepsilon(t) = u(t/\varepsilon)$ for $|t| \leqslant \varepsilon$, and $u_\varepsilon(t) = 0$ for $|t| > \varepsilon$. Such a function will lie in $C^\infty(\mathbb{T})$, be supported on an interval of length 2ε interval centered at the origin. Such a function satisfies estimates of the form $\|D^k u_\varepsilon\|_{L^\infty(\mathbb{T})} \lesssim_k \varepsilon^{-k}$. Roughly speaking, this is pretty much the best we can do for a function with such small support. $u \in C^\infty(\mathbb{T})$ is supported on $|t| \leqslant \varepsilon$, and $\int_{\mathbb{T}} u(t) \, dt = 1$, then the mean value theorem can be used to show that $\|D^k u\|_{L^\infty(\mathbb{T})} \gtrsim_k \varepsilon^{-k}$ for all $k > 0$.

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 $f \in L^1(\mathbb{T})$, we calculate that

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

The bracketed part of the final term in the equation is independent 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 (from a more sophisticated viewpoint, analyzing this convolution is key because *all* operators with translational symmetry can be represented via convolutions).

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 for another time.

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

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

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

$$D_N(t) = 1 + \sum_{n=1}^{N} e_n(t) + e_n(-t) = 1 + e(t) \frac{e_N(t) - 1}{e(t) - 1} + e(-t) \frac{e_N(-t) - 1}{e(-t) - 1}$$

$$= 1 + e(t) \frac{e_N(t) - 1}{e(t) - 1} + \frac{e_N(-t) - 1}{1 - e(t)} = \frac{e_{N+1}(t) - e_N(-t)}{e(t) - 1}$$

$$= \frac{e_{N+1/2}(t) - e_{N+1/2}(-t)}{e_{1/2}(t) - e_{1/2}(-t)} = \frac{\sin(2\pi(N + 1/2)t)}{\sin(\pi t)}. \qquad \square$$

34

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

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

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 $f \in C(\mathbb{T})$ 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 $C(\mathbb{T})$. If we let f_N be a continuous function approximating $\mathrm{sgn}(D_N)$ for each N, then $|\Lambda_N f_N| \gtrsim \log N \cdot \|f_N\|_{L^\infty(\mathbb{T})}$. This implies that $\|\Lambda_N\| \to \infty$ as $N \to \infty$. The uniform boundedness principle thus implies that there exists a *single* function $f \in C(\mathbb{T})$ such that $\sup |\Lambda_N f| = \infty$, so $(S_N f)(0)$ diverges as $N \to \infty$. $\qquad\square$

The situation is even worse than this for general integrable functions. In 1927, Andrey Kolmogorov constructed an integrable function whose Fourier series diverges everywhere. But there is some hope. In 1928, Marcel Riesz showed, using the *boundedness of the Hilbert transform*, discussed later in these notes, that if $1 < p < \infty$, and $f \in L^p(\mathbb{T})$, that $S_N f$ converges in the L^p norm to f. After another half century of the development of harmonic analysis, in 1966, Carleson proved that for each $f \in L^p(\mathbb{T})$, 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 if S_N is *square summation*, then $S_N f$ converges to f in $L^p(\mathbb{T}^d)$, 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 $L^2(\mathbb{T}^d)$. The problem of almost everywhere convergence remains open.

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 $a_0, a_1, \ldots$, 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. We encountered such a sum in the study of trigonometric series via the trigonometric expansion associated with the characteristic function i.e.

$$\chi_{(-a,+a)}(x) \sim \frac{a}{\pi} + \sum_{n=1}^{\infty} \frac{2\sin(na)}{n\pi} \cos(nx),$$

which we saw did not converge pointwise for any value of x. This tempts us to try and make sense of the right hand side as a 'divergent series'. There are several methods of assigning values to divergent series. For example, since the time of Euler, a new method of summation developed by Cesaro had been introduced which 'regularized' certain divergent series so that a value could be assigned to them. Rather than considering limits of partial sums, we consider limits of averages of sums, known as Cesaro means. Letting $s_n = \sum_{k=0}^{n} a_k$, we define the Cesaro means

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

A sequence $\{a_k\}$ is *Cesaro summable* to some value σ if σ_n converge to σ as $n \to \infty$. If the series is summable in the usual manner, then the Cesaro sum exists.

However, the Cesaro summation is stronger than normal convergence.

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

There is actually a *system* of Cesaro summation, which allows us to sum more and more divergent series. For each $l \geqslant 0$, we consider the (C, l) *means* σ_k^l of a sequence $\{a_k\}$ by setting $\sigma_n^0 = a_1 + \cdots + a_n$, and for $l \geqslant 1$, defining

$$\sigma_n^l = \frac{\sigma_1^{l-1} + \cdots + \sigma_n^{l-1}}{n}.$$

The $(C, 1)$ means are just the usual Cesaro means defined above. A series is then (C, l) summable if the associated (C, l) means converge. We can actually extend this method to a theory of (C, α) means by noticing that

$$\sigma_n^l = \sum_{k=1}^{n} \binom{n}{k} \binom{n+l}{k}^{-1} a_k,$$

and so we can define the (C, α) means for any $\alpha \in \mathbb{R} - \{-1, -2, \dots\}$ by setting

$$\sigma_n^\alpha = \sum_{k=1}^{n} \binom{n}{k} \binom{n+\alpha}{k}^{-1} a_k.$$

We can then define (C, α) convergence, which get progressively weaker as $\alpha \to \infty$.

Another notion of regularization sums emerged from the theory of power series, a method called *Abel summation*. Given a sequence $\{a_k\}$, we can consider the power series $\sum a_k r^k$. Under suitably weak conditions, i.e. by the root or ratio test, this series is well defined for $|r| < 1$, and we can consider the Abel means $A_r = \sum a_k r^k$, and ask if $\lim_{r \to 1} A_r$ exists, which should be 'almost like' $\sum a_k$. If this limit exists, we call it the Abel sum of the sequence.

Example. *In the Abel sense, we have $1 - 2 + 3 - 4 + 5 - \cdots = 1/4$, because*

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

37

The partial sums are given by

$$s_n = (-1)^{n+1}\lceil n/2 \rceil$$

and clearly do not converge. The Cesaro sums

$$\sigma_n = \begin{cases} \frac{m}{2m-1} & : n = 2m-1 \\ 0 & : n = 2m \end{cases} - n^{-1}\sum_{k=1}^{n}(-1)^k\lceil k/2 \rceil$$

also do not converge, i.e. because the sequence oscillates between a sequence converging to $1/2$, and a sequence which is equal to zero.

Abel summation is always more general than Cesaro summation, a result provided in 1880, due to Frobenius.

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

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

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

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

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

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

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

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

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

Remark. For any α, (C, α) summability implies Abel summability, though the proof is somewhat more technical.

38

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 $f \in L^1(\mathbb{T})$,*

- $(\sigma_N f)(x) \to f(x)$ *for all x in the Lebesgue set of f.*

- $\sigma_N f \to f$ *uniformly if $f \in C(\mathbb{T})$.*

- $\sigma_N f \to f$ *in the L^p norm for $1 \leqslant p < \infty$, if $f \in L^p(\mathbb{T})$.*

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

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

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

Corollary 2.20. *If $f \in L^1(\mathbb{T})$ and $\hat{f} = 0$, then $f = 0$ almost everywhere.*

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

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 $f \in L^1(\mathbb{T}^d)$ 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 $\hat{f}$, then*

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

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

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 $f \in L^1(\mathbb{T})$ we have

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

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 $t \in \mathbb{T}$, 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

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

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

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,
$$\|P_r\|_{L^\infty(\mathbb{T})} \leqslant \frac{1 - r^2}{1 - 2r + r^2} \leqslant \frac{2}{1 - r}.$$
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 $f \in L^1(\mathbb{T})$,*

- $(A_r f)(t) \to f(t)$ *for all x in the Lebesgue set of f.*

- $A_r f \to f$ *uniformly if $f \in C(\mathbb{T})$.*

- $A_r f \to f$ *in the L^p norm for $1 \leqslant p < \infty$, if $f \in L^p(\mathbb{T})$.*

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 Poisson sum does not differ that much from the ordinary sum, which gives useful results.

Theorem 2.22. *If $\hat{f}(n) = O(|n|^{-1})$, then for any integers N and k, if*

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

then

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

Where the implicit constant is independent 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 $S_M f$,

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

Corollary 2.23. *If $f \in L^1(\mathbb{T})$ with $\hat{f}(n) = O(|n|^{-1})$,*

- $S_N f$ *converges to f in the L^p norm for $1 \leqslant p < \infty$.*

- $S_N f$ *converges uniformly to f if $f \in C(\mathbb{T})$.*

- $(S_N f)(x) \to f(x)$ *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 $kN \leqslant M < k(N+1)$,

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

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

Let us briefly describe a more general way to formulate this argument. Fix a function $f : \mathbb{T} \to \mathbb{C}$, and let us suppose that for each N, we can find a trigonometric polynomial P_N of degree N such that $\|f - P_N\|_{L^\infty(\mathbb{T})} \leqslant \varepsilon_N$, then, since S_N has operator norm $O(\log N)$ in the L^∞ norm, and $S_N P_N = P_N$, we conclude that

$$\|S_N f - P_N\|_{L^\infty(\mathbb{T})} \lesssim \log N \cdot \varepsilon_N.$$

But the triangle inequality thus implies that $\|S_N f - f\|_{L^\infty(\mathbb{T})} \lesssim \log N \cdot \varepsilon_N$. The De la Vallée Poisson sum gives precisely such a polynomial if $|\hat{f}(n)| \lesssim |n|^{-1}$, which is what allowed us to prove the required result. This gives convergence provided we can choose $\{P_N\}$ such that $\varepsilon_N = o(\log N)$. Similar results hold in $L^1(\mathbb{T})$.

TODO: RESEARCH MORE INTO NONLINEAR APPROXIMATION BY POLYNOMIALS.

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

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

Then X is a closed subset of $L^1(\mathbb{T})$. 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 the point t_0. $\square$

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

Theorem 2.25 (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

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

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

$$\int_{|t|<\delta_0} f(t) D_N(t) \, dt = \int_{|t|<\delta_0} \left(\sin(2\pi(N+1/2)t)\frac{f(t)}{t} \right) \left(\frac{t}{\sin(\pi t)} \right) \, dt$$

$$\lesssim \|f(t)/t\|_{L^1[-\delta_0,\delta_0]} \leqslant \varepsilon.$$

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

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

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

for any fixed $\delta > 0$.

Lemma 2.26 (Jordan's Criterion). *If $f \in L^1(\mathbb{T})$ 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

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

44

it suffices without loss of generality to show that

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

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

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

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

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

Now an integration by parts gives

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

Thus

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

Remark. The calculations in this proof also show that if $f \in L^1(\mathbb{T})$ 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 $f \in C(\mathbb{T})$, and for any $1 \leqslant p < \infty$, if $f \in L^p(\mathbb{T})$, $S_N f$ converges to f in $L^p(\mathbb{T})$. 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*

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

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

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

Theorem 2.27. *If* $\hat{f}(n) = O(|n|^{-1})$, *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 $\mathbb{T}$ so it is continuous at t_0. Now we find $|\hat{g}| \lesssim |\hat{f}| + |\hat{s}| = O(|n|^{-1})$, 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 J. Gibbs, who discovered the phenomena in 1899 while taking physical measurements, first thought to be a defect in the equipment used to take measurements (though in 1848 H. Wilbraham had already discovered the phenomenon). Gibb's phenomenon is one instance where a series of functions $\{f_n\}$ converges pointwise to some function f, whereas qualitatively with respect to the L^∞ norm, the sequence $\{f_n\}$ does not converge to f, precisely because of this overshooting.

Theorem 2.28. *Given a function f, continuous except at finitely many discontinuity points, one of which being some $t_0 \in \mathbb{R}$, and such that*

$$|\widehat{f}(n)| \lesssim |n|^{-1},$$

we have

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

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

47

$S_N h_j \to 0$ 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 $d(t, 2\pi\mathbb{Z})$ is bounded below, the quantity above tends to zero uniformly. This gives the required result. $\qquad \square$

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 $a_1, \ldots, a_N$ such that

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

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

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

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

Corollary 3.1 (Weirstrass). *The polynomials are uniformly dense in $C[0, 1]$.*

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

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

But if $x = \cos t$, 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. □

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

$$\int_{-1}^1 (1 - x^2)^n \, dx \geqslant = \sum_{k=0}^\infty \int_{1/2^{k+1} \leqslant |x| \leqslant 1/2^k} (1 - x^2)^n$$

$$\gtrsim \sum_{k=0}^\infty \frac{(1 - 1/4^{k+1})^n}{2^k}$$

$$= \sum_{k=0}^\infty \exp\left(n \log(1 - 1/4^{k+1}) - k \log(2) \right)$$

$$\geqslant \sum_{k=0}^\infty \exp\left(-n/4^k \right) / 2^k$$

$$\gtrsim \sum_{k=\log_4 n}^\infty 1/2^k \gtrsim 1/2^{\log_4 n} = 1/n^{1/2}$$

Thus $\|L_n\|_{L^\infty(\mathbb{R})} \leqslant n^{1/2}$ 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 $a_1, a_2, \ldots$. In particular, it is most useful in determining when this distribution is distributed when we consider $2\pi a_1, 2\pi a_2, \ldots$ as elements of $\mathbb{T}$, 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 $I \subset \mathbb{T}$, $\#\{2\pi a_n \in I : n \leqslant N\} \sim N|I|$ as $N \to \infty$. By approximating continuous functions by step functions, this implies that if $f : \mathbb{T} \to \mathbb{C}$ is continuous, then

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

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 $a_1, a_2, \cdots \in \mathbb{T}$ is uniformly distributed if and only if for every n, as $N \to \infty$, $e_n(2\pi a_1) + \cdots + e_n(2\pi a_N) = o(N)$.*

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 $L^\infty(\mathbb{T})$, and T is a bounded functional on this space as well. But this means that $\lim S_N f = Tf$ for all f in $C(\mathbb{T})$, 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 $\gamma, 2\gamma, \ldots$, 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 $\gamma, 2\gamma, \ldots$ is equidistributed. Conversely, if $e_n(2\pi\gamma) = 1$ for some n, we have

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

51

which is not $o(N)$, so the sequence $\gamma, 2\gamma, \ldots$ is not equidistributed. If γ is rational, there certainly is n such that $n\gamma \in \mathbb{Z}$, 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 $\gamma, \gamma^2, \gamma^3, \ldots$ 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 $d(b_n, \mathbb{Z}) = d(\gamma^n, \mathbb{Z}) \to 0$. 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

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

We consider a boundary condition, that $u(e^{2\pi it}) = f(t)$ for some function $f(t)$ on $\mathbb{T}$. 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 $n \in \mathbb{Z}$,

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

53

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

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

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 $L^1(\mathbb{T})$ 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 $\mathbb{T}^d$. We wish to study the propogation of this temperature over time. If we let $u(x,t)$ denote the temperature density at $x \in \mathbb{T}^d$ 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

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

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

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

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

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

54

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*

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

The rapid convergence of this sum implies that $H_t \in C^\infty(\mathbb{T}^d)$ and that $\hat{H}_t(n) = e^{-4\pi^2 |n|^2}$ for each $n \in \mathbb{Z}^d$. 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

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

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

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

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

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

$$\lesssim t^{-d/2} e^{-c'/2t} \sum_{n \neq 0} e^{-c'|n|^2/2}$$

$$\lesssim_d t^{-d/2} e^{-c'/2t} \lesssim_d e^{-c/t}.$$

55

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

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

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 $f \in L^1(\mathbb{T}^d)$, for $1 \leq p < \infty$. Then the function*

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

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

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

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 $u \in C^2(\mathbb{T}^d \times (0, \infty))$, we can take a Fourier series, letting

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

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 $a_n(t) \to \hat{f}(n)$ as $t \to 0$ uniformly in n by the convergence assumption, so $c_n = \hat{f}(n)$. But this implies that $u(x,t) = (H_t * f)(x)$ for each $x \in \mathbb{T}^d$, 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 $\mathbb{R}$ is non-compact and has infinite measure adds some difficulty to the study of the Fourier transform over the Fourier series.

For instance, since $L^p(\mathbb{R}^d)$ is not included in $L^q(\mathbb{R}^d)$ 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 $\mathbb{R}^d$, which, given $f : \mathbb{R}^d \to \mathbb{C}$, considers the quantities

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

for $\xi \in \mathbb{R}^d$. 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 $L^1(\mathbb{R}^d)$. 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 $f \in L^1(\mathbb{R}^d)$, $\|\hat{f}\|_{L^\infty(\mathbb{R}^d)} \leqslant \|f\|_{L^1(\mathbb{R}^d)}$, and $\hat{f} \in C_0(\mathbb{R}^d)$.*

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

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

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

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

58

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 $\xi_k \in \mathbb{R}$, which implies that

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

Thus $\hat{\chi}_I(\xi) \to 0$ 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 $L^1(\mathbb{R}^d)$, a density argument then gives the result for all integrable functions. $\qquad\square$

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

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

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

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

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

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

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

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

and

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

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

- 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 $f, g \in L^1(\mathbb{R}^d)$. This convolution possesses precisely the same properties as convolution on $\mathbb{T}$. 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 $f \in L^1(\mathbb{R}^d)$ has a weak derivative $D^\alpha f \in L^1(\mathbb{R}^d)$, then

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

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

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

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

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

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

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

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

$$= \frac{1}{|\det(T)|} \int f(y)e^{-2\pi i \xi \cdot T^{-1}y)} \, dy$$

$$= \frac{1}{|\det(T)|} \int f(y)e^{-2\pi i T^{-T}\xi \cdot y} \, dy$$

$$= \frac{1}{|\det(T)|}(\widehat{f} \circ T^{-T})(\xi).$$

Thus we conclude that if $T^* : L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$ 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 $a \in \mathbb{R}$ and $\mathrm{Dil}_a : L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$ is the operator defined by setting

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

then

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

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, $\hat{f}$ is dilated so it's support lies in a larger and larger neighborhood.

- Another special case is that if $R \in O_n(\mathbb{R})$, then $\widehat{f \circ R}(\xi) = \hat{f}(R\xi)$, 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 $\hat{f}(R\xi) = \hat{f}(\xi)$ for any $R \in O_n(\mathbb{R})$, so $\hat{f}$ is also a radial function. If f is even, so $f(x) = f(-x)$ for all x, then $\hat{f}(\xi) = \hat{f}(-\xi)$ for all ξ, so $\hat{f}$ is even. Similarily, if f is odd, then $\hat{f}$ 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

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

is called the *Fourier algebra*. The last theorem shows $\mathbf{A}(\mathbb{R}^d) \subset C_0(\mathbb{R}^d)$, but it is *not* the case that $\mathbf{A}(\mathbb{R}^d) = C_0(\mathbb{R}^d)$. Current research cannot give a satisfactory description of the elements of $\mathbf{A}(\mathbb{R}^d)$, and a simple characterization is unlikely. The next lemma will be used to show $\mathbf{A}(\mathbb{R}^d) \neq C_0(\mathbb{R}^d)$.

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

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

Proof. Since $\|\sin(x)/x\|_{L^\infty(\mathbb{R})} \leqslant 1$, 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. $\mathbf{A}(\mathbb{R}) \neq C_0(\mathbb{R})$. *In particular,* $\mathbf{A}(\mathbb{R})$ *does not contain any odd functions g in* $C_0(\mathbb{R})$ *such that*

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

Proof. Suppose $f \in L^1(\mathbb{R})$, and $\widehat{f} \in C_0(\mathbb{R})$ 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

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

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

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

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

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

63

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

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

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

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

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 $f \in C(\mathbb{R}) \cap L^1(\mathbb{R}^d)$ and $\widehat{f} \in L^1(\mathbb{R}^d)$, then the formula

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

holds for all $x \in \mathbb{R}^d$. 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 $f, g \in L^1(\mathbb{R}^d)$,*

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

Proof. If $f, g \in L^1(\mathbb{R}^d)$, then $\widehat{f}$ and $\widehat{g}$ are bounded, continuous functions on $\mathbb{R}^d$. 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 $f \in L^1(\mathbb{R}^d)$ and $\widehat{f} = 0$, then for any $g \in L^1(\mathbb{R}^d)$,

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

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 $g \in L^1(\mathbb{R})$. Then $g'(x) = -2\pi x g(x)$, and since $xg \in L^1(\mathbb{R})$, we 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 $L^1(\mathbb{R}^d)$, then

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

In particular, if for $x_0 \in \mathbb{R}^d$ 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,

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

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

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

65

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

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

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,

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

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

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

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

In particular, if we also assume $f \in C(\mathbb{R}^d)$, 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*

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

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

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

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

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

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

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 $f \in L^1(\mathbb{R}^d)$, we let

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

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*

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

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

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

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

67

Lemma 4.7. *Suppose $f \in L^1_{loc}(\mathbb{R})$, 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 $s > 0$,

$$\int_{-s}^{s} f(x)e^{-\delta x^2} \, dx = \int_{0}^{s} [f(x) + f(-x)]e^{-\delta x^2} \, dx$$

$$= F(s)e^{-\delta s^2} + 2\delta \int_{0}^{s} xF(x)e^{-\delta x^2} \, dx.$$

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

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

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

$$\left| \int f(x)e^{-\delta x^2} \, dx - A \right| \leqslant 2\delta \left| \int_{0}^{t} xF(x)e^{-\delta x^2} \, dx \right|$$

$$+ 2\delta\varepsilon \left| \int_{t}^{\infty} xe^{-\delta x^2} \right|$$

$$+ \left| 2\delta A \int_{t}^{\infty} xe^{-\delta x^2} \, dx - A \right|.$$

68

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 $\varepsilon \to 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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

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

69

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. □

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 $\Phi \in C_0(\mathbb{R}^d)$ 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 $f \in L^1(\mathbb{R}^d)$, 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 $f, g \in L^1(\mathbb{R}^d)$,*

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

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

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

70

If Φ is integrable, then the multiplication formula shows

$$S_\delta(f, \Phi) = \int \hat{f}(\xi) e(\xi \cdot x) \Phi(\delta\xi) d\xi$$
$$= \int f(x)(\mathrm{Mod}_x(\delta_\delta\Phi))^\wedge(x)\, dx = \delta^{-n} \int f(x) \cdot \hat{\Phi}\left(\frac{x-y}{\delta}\right)\, dx.$$

Thus if we define $K_\delta^\Phi(x) = \delta^{-d}\hat{\Phi}(-x/\delta)$, 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 $\{\|K_\delta\|_{L^1(\mathbb{R}^d)}\}$ 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 $f \in L^p(\mathbb{R}^d)$, and converges to f uniformly if f is continuous and bounded. If we have the stronger conditions that

- For any $\delta > 0$,
$$\int K_\delta(\xi)\, d\xi = 1.$$

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

- For any $\delta > 0$ and $\xi \in \mathbb{R}^d$,
$$|K_\delta(\xi)| \lesssim \frac{\delta}{|\xi|^{d+1}}.$$

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 $\hat{\Phi} \in L^1(\mathbb{R}^d)$ and

$\int \widehat{\Phi}(\xi) \, d\xi = 1$, and forms an approximation to the identity if we assume in addition that $\Phi \in C^{d+1}(\mathbb{R}^d)$. 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 $\text{Im}(z) \leqslant 0$,

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

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

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

72

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*

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

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,

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

and so we conclude that

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

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

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

so $g_\xi(z) = O_\xi(1/|z|^2)$ in the half plane above the line $\mathbb{R} - iy$. 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

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

73

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

$$P(\xi) = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}(1 + |\xi|^2)^{\frac{d+1}{2}}} = \frac{2}{|S^d|} \frac{1}{(1 + |\xi|^2)^{\frac{d+1}{2}}}.$$

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 $\mathbb{R}^d$ 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

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

$$= \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} \int_{\mathbb{R}^d} e^{-|\pi x|^2/u} e^{-2\pi i \xi \cdot x} \, dx \, du = \int_0^\infty \frac{e^{-u}}{\sqrt{\pi u}} (Dil_{(\pi^{1/2}/u^{1/2})}\Phi)^\wedge(\xi) \, du$$

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

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 $\mathbb{T}$ is the periodization of the Poisson kernel on $\mathbb{R}$.

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

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

75

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

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

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 $D_R \notin L^1(\mathbb{R})$. 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 $L^p(\mathbb{R})$ for all $p \in (1, \infty]$, and is uniformly bounded in $L^p(\mathbb{R})$ 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 $L^p(\mathbb{R})$.

Thus we now know there are a large examples of functions $\Phi \in C_0(\mathbb{R}^d)$ 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 $|\widehat{f}(\xi) e^{2\pi i \xi \cdot x} \Phi(\delta \xi)| \leq \|\Phi\|_\infty |\widehat{f}(\xi)|$ 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 $L^1(\mathbb{R}^d)$, then for any x in the Lebesgue set of f,*

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

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

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

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

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

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

then $R \in L^1(\mathbb{R}^d)$ 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 $f \in L^2(\mathbb{R}^d)$.

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 $f \in L^1(\mathbb{R})$ is continuous at the origin, and $\hat{f} \geq 0$, then $\hat{f}$ is integrable.*

Proof. This follows because

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

By Fatou's lemma,

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

so $\hat{f}$ 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. Here is a related result. Here, without integrability constraints, the Fourier transform of $m \in L^\infty(\mathbb{R}^d)$ will be taken in the sense of tempered distributions, introduced later on in the notes.

Lemma 4.12. *If $m \in L^\infty(\mathbb{R}^d)$, and $\hat{m} \geq 0$, then $\|m\|_{L^\infty(\mathbb{R}^d)} = \|\hat{m}\|_{L^1(\mathbb{R}^d)}$.*

Proof. Set $k = \hat{m}$. The result is equivalent to showing that if $m \in L^\infty(\mathbb{R}^d)$ and $\hat{m} \geq 0$, then $\|m\|_{L^\infty(\mathbb{R}^d)} = \|k\|_{L^1(\mathbb{R}^d)}$. From this it then follows that for a general p,

$$\|k\|_{L^1(\mathbb{R}^d)} = \|m\|_{M^2} \leq \|m\|_{M^p} \leq \|m\|_{M^1} = \|k\|_{L^1(\mathbb{R}^d)}.$$

Suppose first that m is continuous. If Φ is the Gaussian function, and $\Phi_\delta(\xi) = \delta^{-d}\Phi(\xi/\delta)$ then the multiplication formula implies that

$$\int m(\xi)\Phi_\delta(\xi - \xi_0) = \int k(x)\Phi(\delta\xi)e^{2\pi i \xi_0 \cdot x} \, dx.$$

If m is continuous, taking $\delta \to 0$ gives

$$|m(\xi_0)| \leq \limsup_{\delta \to 0} \left| \int k(x)\Phi(\delta\xi)e^{2\pi i \xi_0 \cdot x} \, dx \right| \leq \int k(x); dx$$

78

and by monotone convergence,

$$m(0) = \lim_{\delta \to 0} \int k(x)\Phi(\delta\xi) = \int k(x)\,dx.$$

Thus $\|m\|_{L^\infty(\mathbb{R}^d)} = \|k\|_{L^1(\mathbb{R}^d)}$.

But now if we let $k_\varepsilon(x) = k(x)\Phi(\varepsilon x)$, and $m_\varepsilon = \hat{k}_\varepsilon = m * \Phi_\varepsilon$, then m_ε converges distributionally to m as $\varepsilon \to 0$. Now k_ε is non-negative, and Young's inequality implies $\|m_\varepsilon\|_{L^\infty(\mathbb{R}^d)} \leqslant \|m\|_{L^\infty(\mathbb{R}^d)}$. Thus

$$\|m_\varepsilon\|_{L^\infty(\mathbb{R}^d)} = \|k_\varepsilon\|_{L^1(\mathbb{R}^d)}.$$

Taking $\varepsilon \to 0$ gives

$$\lim_{\varepsilon \to 0} \|m_\varepsilon\|_{L^\infty(\mathbb{R}^d)} = \|k\|_{L^1(\mathbb{R}^d)}.$$

But now

$$\|m\|_{L^\infty(\mathbb{R}^d)} \geqslant \lim_{\delta \to 0} \delta^{d/2} \|k * \Phi_\delta\|_{L^2(\mathbb{R}^d)}$$
$$= \lim_{\delta \to 0} \lim_{\varepsilon \to 0} \delta^{d/2} \|k_\varepsilon * \Phi_\delta\|_{L^2(\mathbb{R}^d)}.$$

These limits are both monotonic, so we can interchange them, thus getting that

$$\lim_{\delta \to 0} \lim_{\varepsilon \to 0} \delta^{d/2} \|k_\varepsilon * \Phi_\delta\|_{L^2(\mathbb{R}^d)} = \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \delta^{d/2} \|k_\varepsilon * \Phi_\delta\|_{L^2(\mathbb{R}^d)}$$
$$= \lim_{\varepsilon \to 0} \|k_\varepsilon\|_{L^1(\mathbb{R}^d)}$$
$$= \|k\|_{L^1(\mathbb{R}^d)}.$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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

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

The inverse transform is also denoted by $\mathcal{F}^{-1}(f)$. The last theorem says that $\mathcal{F}^{-1}$ really is the inverse operator to the operator $\mathcal{F}$, at least on the set of functions f where $\hat{f}$ is integrable. In particular, this is true if f has 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 4.13. *If $f \in C(\mathbb{R})$ is integrable and $\hat{f} \in L^1(\mathbb{R})$, $S_R f \to f$ uniformly.*

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

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

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. $\qquad\square$

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

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

also converge uniformly to f.

Corollary 4.14. *The map $\mathcal{F} : L^1(\mathbb{R}^d) \to C_0(\mathbb{R}^d)$ 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. $\qquad\square$

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 4.15. *For any δ_1, δ_2,*

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

Proof. We recall that
$$W_{\delta_1+\delta_2} = \mathcal{F}\left(e^{-(\delta_1+\delta_2)|x|^2}\right).$$
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
$$\mathcal{F}\left(e^{-\pi\delta_1|x|^2}e^{-\pi\delta_2|x|^2}\right) = \mathcal{F}\left(e^{-\pi\delta_1|x|^2}\right) * \mathcal{F}\left(e^{-\pi\delta_2|x|^2}\right) = W_{\delta_1} * W_{\delta_2}.$$
Thus $W_{\delta_1} * W_{\delta_2} = W_{\delta_1+\delta_2}$. Similarly, $P_{\delta_1+\delta_2}$ is the Fourier transform of $e^{-(\delta_1+\delta_2)|x|}$, which breaks into a product, whose individual Fourier transforms are P_{δ_1} and P_{δ_2}. $\qquad\square$

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 $f \in L^1(\mathbb{R})$ 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, $L^2(\mathbb{T}^d)$ is contained in $L^1(\mathbb{T}^d)$ and can thus be viewed as a *more regular* family of functions than the square integrable functions. In the noncompact case, $L^2(\mathbb{R}^d)$ is not contained in $L^1(\mathbb{R}^d)$, 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 $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, which is a dense subspace of $L^2(\mathbb{R}^d)$.

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

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

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

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

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

Theorem 4.16. *If $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, then $\|\hat{f}\|_{L^2(\mathbb{R}^d)} = \|f\|_{L^2(\mathbb{R}^d)}$.*

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

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

and

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

This implies

$$\int_{\mathbb{R}^d} |\hat{f}(\xi)|^2 \, d\xi = \int_{\mathbb{R}^d} \hat{f}(\xi)\overline{\hat{f}(\xi)} \, d\xi = \int_{\mathbb{R}^d} f(x)\overline{f(x)} \, dx = \int_{\mathbb{R}^d} |f(x)|^2 \, dx. \qquad \square$$

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

Corollary 4.17. *If $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ for $1 \leqslant p \leqslant 2$, then*

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

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

Another simple interpolation argument yields the following interesting weighted inequality, due to Paley, which allows one to get some information about the L^p norm of the Fourier transform of an L^p function.

Corollary 4.18. *If $1 < p \leqslant 2$, and let $r = (1 - p/p^*)^{-1} = (2 - p)^{-1}$. Then if $f \in L^p(\mathbb{R}^d)$, and $w \in L^{r,\infty}(\mathbb{R}^d)$, then*

$$\|\hat{f}\|_{L^p(\mathbb{R}^d,w)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}\|w\|_{L^{r,\infty}(\mathbb{R}^d)}^{1/p}.$$

Interpolating between the Hausdorff-Young inequality, we conclude that for an exponent $1 < p \leqslant 2$, then for $p \leqslant q \leqslant p^*$, if $r = (1 - q/p^*)^{-1}$, and if $w \in L^{r,\infty}(\mathbb{R}^d)$, then

$$\|\hat{f}\|_{L^q(\mathbb{R}^d,w)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}\|w\|_{L^{r,\infty}(\mathbb{R}^d)}^{1/q}.$$

Proof. We reformulate the bound into a form that is easier to interpolate. Define $g = w^r$. Then the inequality we want to prove is equivalent to the ienquality

$$\left(\int |\hat{f}(\xi)|^p |g(\xi)|^{2-p}\right)^{1/p} \lesssim \|f\|_{L^p(\mathbb{R}^d)}\|g\|_{L^{1,\infty}(\mathbb{R}^d)}^{2/p-1}.$$

If we define the operator
$$T_g f(\xi) = \hat{f}(\xi)/g(\xi),$$
and $w_g(\xi) = g(\xi)^2$, then the problem is just to show that

$$\|T_g f\|_{L^p(\mathbb{R}^d, w_g)} \lesssim \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^{1,\infty}(\mathbb{R}^d)}^{2/p-1}.$$

The case $p = 2$ is just Parsevel's inequality. We claim for $p = 1$ we also have

$$\|T_g f\|_{L^{1,\infty}(\mathbb{R}^d, w_g)} \lesssim \|f\|_{L^1(\mathbb{R}^d)} \|w\|_{L^{1,\infty}(\mathbb{R}^d)}.$$

But this follows since we have a pointwise bound

$$|Tf| \leqslant \|f\|_{L^1(\mathbb{R}^d)}/|g|.$$

Thus the set
$$\{\xi \in \mathbb{R}^d : |Tf(\xi)| \geqslant \alpha\}$$
is contained in
$$\{\xi \in \mathbb{R}^d : |w(\xi)| \leqslant \|f\|_{L^1(\mathbb{R}^d)}/\alpha\},$$
But a simple computation shows that, if $m(s)$ is the measure of the set $\{x : g(x) \geqslant s\}$, then

$$\int_{w_g(\xi) \leqslant \sigma} g(\xi)\, d\xi = -\int_0^\sigma s^2 m'(s)\, ds$$

$$\leqslant [\lim_{s \to 0} s^2 m(s) - \sigma^2 m(\sigma)] + 2\int_0^\sigma sm(s)\, ds$$

$$\leqslant 2\|g\|_{L^{1,\infty}(\mathbb{R}^d)} \cdot \sigma.$$

Setting $\sigma = \|f\|_{L^1(\mathbb{R}^d)}/\alpha$ yields the required result. $\qquad\square$

Though the integral formula of an element of $L^2(\mathbb{R}^d)$ does not make sense, the bounds above provide a canonical way to define the Fourier transform of an element of $L^p(\mathbb{R}^d)$, for $1 \leqslant p \leqslant 2$. The space $L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ is a dense subset of $L^p(\mathbb{R}^d)$, so we can use the Hahn-Banach theorem to define the Fourier transform $\mathcal{F} : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ 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

$f \in L^2(\mathbb{R}^d)$, and $R > 0$, $f \mathbb{I}_{B_R} \in L^1(\mathbb{R}^d)$, 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 $\lim_{R \to \infty} \| \mathbb{I}_{B_R} f - f \|_{L^2(\mathbb{R}^d)} = 0$, the L^2 continuity of the Fourier transform implies that $\mathcal{F}_R f$ converges to $\widehat{f}$ in $L^2(\mathbb{R}^d)$.

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 $f, g \in L^2(\mathbb{R}^d)$,

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

This equality certainly holds by the multiplication formula if $f, \widehat{g} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. 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 $L^2(\mathbb{R}^d)$ can be approximated in the L^2 norm by an element of $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, and since any element g of $L^2(\mathbb{R}^d)$ can be approximated by functions with $\widehat{g} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, the theorem holds in general. In particular, this shows that the extension of the Fourier transform to $L^2(\mathbb{R}^d)$ remains unitary.

Another approximation argument can be used to obtain convergence results in $L^2(\mathbb{R}^d)$ 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 $\Phi \in L^2(\mathbb{R}^d)$, 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 $\Phi \in L^1(\mathbb{R}^d)$ and $\int \widehat{\Phi}(\xi) \, d\xi = 1$, then we conclude that $S_\delta(f, \Phi) \to f$ in $L^2(\mathbb{R}^d)$, and that if $\Phi \in C^{d+1}(\mathbb{R}^d)$, 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 $L^2(\mathbb{T}^d)$ and $L^2(\mathbb{Z}^d)$, in the case of the Fourier transform the Fourier transform gives a unitary operator from $L^2(\mathbb{R}^d)$ 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*

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

and we define $\tilde{H}_n = e^{-\pi x^2} H_n$, then $\widehat{\tilde{H}_n} = (-i)^n \tilde{H}_n$. Thus we obtain an orthonormal basis of eigenfunctions for $L^2(\mathbb{R})$. In higher dimensions, a basis of eigenfunctions for $L^2(\mathbb{R}^d)$ is given by taking tensor products of Hermite polynomials.

4.6 The Hausdorff-Young Inequality

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

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

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

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 $\mathbb{T}$, 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 4.19. *Suppose $T : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is a translation invariant continuous operator, where $q < p$. Then $T = 0$.*

Proof. If T was nonzero, we could pick some $f_0 \in L^p(\mathbb{R}^d)$ such that $Tf_0 \neq 0$. Rescaling T and f_0, we may assume without loss of generality that $\|f_0\|_{L^p(\mathbb{R}^d)} = \|Tf_0\|_{L^q(\mathbb{R}^d)} = 1$. Furthermore, by truncation we may assume that f_0 has compact support on some ball B_R. But then the supports of the functions $\text{Trans}_{2Rn}f_0$ and f_0 are disjoint for $n \in \mathbb{Z}^d$, so for any choice of coefficients $\{a_n\}$,

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

Assume that at most N of the coefficients a_n are nonzero. We cannot necessarily assume that $T f_0$ has compact support but the majority of the mass of $T f_0$ can still be concentrated on a compact set. For any $\varepsilon > 0$ we can choose R large enough that

$$\left(\int_{|x| \geqslant R} |T f_0(x)|^q \right)^{1/q} \leqslant \varepsilon.$$

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

$$\left(\int_{x \in 2Rm + B_R} \left| \sum_{n \in \mathbb{Z}^d} a_n \mathrm{Trans}_{2Rn} T f_0(x) \right|^q dx \right)^{1/q} \geqslant \left(|a_m|^q - \varepsilon \sum_{n \neq m} |a_n|^q \right)^{1/q}.$$

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

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

Then we find

$$\left(\int_{x \in 2Rm + B_R} \left| \sum_{n \in \mathbb{Z}^d} a_n \mathrm{Trans}_{2Rn} T f_0(x) \right|^q dx \right)^{1/q} \geqslant 0.5^{1/q} |a_m|$$

and so summing over all m, we conclude that

$$\left\| \sum_{n \in \mathbb{Z}^d} a_n \mathrm{Trans}_{2Rn} T f_0 \right\|_{L^q(\mathbb{R}^d)} \geqslant 0.5^{1/q} \left(\sum_{n \in \mathbb{Z}^d} |a_n|^q \right)^{1/q}.$$

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

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

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 $t \in \mathbb{R}$. One can easily calcluate using analytic continuation and the Fourier transform for the Gaussian that $\hat{f}_s(\xi) = e^{-\pi s|\xi|^2}$. We calculate*

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

whereas $\|\hat{f}_s\|_q = q^{-d/2}$. 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 $\mathbb{T}^d$ 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 Boundedness in L^p

We restrict ourselves to the study of partial summation on the real line. In particular, we consider the partial summation operators

$$S_R f(x) = \int_{-R}^{R} \hat{f}(\xi)e^{2\pi i \xi x}.$$

These operators are translation invariant, and we have $S_R f = D_R * f$, where

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

is the Dirichlet kernel. We are concerned with determining whether, given a general $f \in L^p(\mathbb{R}^d)$, the functions $\{S_R f\}$ converge to f in the L^p norm as $R \to \infty$. Since the result *is* true if f is a Schwartz function, it follows via the uniform boundedness theorem and a density arguemnt that we must establish a bound of the form

$$\|S_R f\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

which is uniform over all Schwartz f and $R > 0$.

A tensorization argument shows that if we can establish such a bound on the real line, then an analogous result holds for the *square summation* operators on $\mathbb{R}^d$, i.e. the operators

$$S_R f(x) = \int_{\max(|\xi_1|,\ldots,|\xi_d|) \leqslant R} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi.$$

On the other hand, the theory of summation operators $\{S_R\}$ for *other* regions of integration (e.g. the spherical summation operators obtained by integrating over $|\xi| \leqslant R$) is much more subtle.

To start with, we note that

$$\|D_R\|_{L^1(\mathbb{R})} \sim \log R.$$

This implies that the operators $\{S_R\}$ are *not* uniformly bounded in $L^1(\mathbb{R})$. As with the case of the Dirichlet kernel on $\mathbb{T}$, this implies the theory of partila summation for general functions in $L^1(\mathbb{R})$ can behave very erratically, i.e. there exists $f \in L^1(\mathbb{R})$ such that $\|S_R f\|_{L^1(\mathbb{R})} \to \infty$ as $R \to \infty$. By duality, partial summation is also not well behaved on $L^\infty(\mathbb{R})$, or even $C_b(\mathbb{R})$.

On the other hand, the operators $\{S_R\}$ *are* uniformly bounded in $L^p(\mathbb{R})$ for any $1 < p < \infty$, and so for any $f \in L^p(\mathbb{R})$, $\{S_R f\}$ converges in $L^p(\mathbb{R})$ to f as $R \to \infty$. We will not provide a complete proof of this fact here, relegating some parts of the proof to later parts of these notes. But we describe the general principles of the argument.

Marcel Riesz obtained the first proof of the L^p boundedness of the operators $\{S_R\}$ by reducing the study of the operators $\{S_R\}$ to the study of a single operator, the *Hilbert transform*

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

The price we pay for reducing our study to a single operator is that this operator is highly *singular*. The definition of the operator cannot be interpreted as a Lebesgue

integral for general $f \in L^p(\mathbb{R}^d)$, since the Fourier transform of a general f need not be integrable on the real line. To begin the study of the Hilbert transform, we must therefore restrict our study to functions f whose Fourier transforms are more well behaved; one can justify that the integral formula defines a bounded operator from $\mathcal{S}(\mathbb{R})$ to $\mathcal{E}(\mathbb{R})$. Since we can write

$$\mathbb{I}(-R \leqslant \xi \leqslant R) = \frac{\mathrm{sgn}(\xi - R) - \mathrm{sgn}(\xi + R)}{2},$$

for Schwartz f we have

$$S_R = \frac{i}{2\pi} \left(\mathrm{Mod}_{-R} \circ H \circ \mathrm{Mod}_R - \mathrm{Mod}_R \circ H \circ \mathrm{Mod}_{-R} \right).$$

Thus if we could show that H is bounded on $L^p(\mathbb{R})$, i.e. that

$$\|Hf\|_{L^p(\mathbb{R})} \lesssim \|f\|_{L^p(\mathbb{R})},$$

for all Schwartz f, then it follows that we have

$$\|S_R f\|_{L^p(\mathbb{R})} \lesssim \|f\|_{L^p(\mathbb{R})},$$

for all Schwartz f, uniformly in R, and the bound then follows by a density argument for all $f \in L^p(\mathbb{R})$.

We will not prove the boundedness of the Hilbert transform at this time, since it requires some more advanced techniques. The classical method involves some complex analysis, though it is now more conventional to analyze the Hilbert transform via the more modern *Calderon-Zygmund theory of singular integrals*. We will see later on in these notes that, for Schwartz f, we can write the Hilbert transform as a singular integral of the form

$$Hf(x) = \lim_{\varepsilon \to 0} \int \frac{f(y)}{x - y} \, dy.$$

The theory of such integrals will then give the L^p boundedness we require.

4.8   The Poisson Summation Formula

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

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

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

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

i.e. projecting f onto it's Fourier components that are periodic of degree one. 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 4.20. *Suppose $f \in L^1(\mathbb{R}^d)$. Then the series*

$$\sum_{n\in\mathbb{Z}^d} \mathrm{Trans}_n f$$

converges absolutely in $L^1[0,1]^d$ to a function $g \in L^1[0,1]^d$ with the property that

$$\widehat{g}(n) = \widehat{f}(n)$$

for each $n \in \mathbb{Z}^d$.

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

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

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

$$\int_{[0,1]^d} \sum_{m\in\mathbb{Z}^d} (\mathrm{Trans}_n f)(x)e^{2\pi nix}\, dx = \sum_{m\in\mathbb{Z}^d} \int_{[0,1]^d} f(x+m)e^{2\pi ni(x+m)}\, dx$$

$$= \int_{\mathbb{R}^d} f(x)e^{2\pi nix}\, dx = \widehat{f}(n). \qquad \square$$

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

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

90

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

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

In particular, this holds if $f \in \mathcal{S}(\mathbb{R})$. From a distributional standpoint, this result says that the tempered distribution $\sum_{n \in \mathbb{Z}^d} \delta_n$ is an eigendistribution of the Fourier transform, i.e. it is it's own Fourier transform.

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

The Poisson summation formula can often be used as a method to prove that bounds for operators on $\mathbb{R}^d$ imply bounds for analogous operators on $\mathbb{T}^d$. For instance, we will prove in our notes on Fourier multipliers that if every point of m is a Lebesgue point, and if

$$T f(x) = \int_{\mathbb{R}^d} m(\xi) \hat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi$$

is the associated Fourier multiplier satisfying bounds of the form

$$\|T f\|_{L^p(\mathbb{R}^d)} \leqslant A \|f\|_{L^p(\mathbb{R}^d)},$$

then if we consider the analogous Fourier multiplier on $\mathbb{T}^d$, i.e.

$$T f(x) = \sum_{n \in \mathbb{Z}} m(n) \hat{f}(n) e^{2\pi i n \cdot x},$$

for functions $f : \mathbb{T}^d \to \mathbb{C}$, then

$$\|T f\|_{L^p(\mathbb{T}^d)} \leqslant A \|f\|_{L^p(\mathbb{T}^d)}.$$

In particular, this implies that the Hilbert transform

$$H f(x) = \sum_n \operatorname{sgn}(n) \hat{f}(n) e^{2\pi i n x}$$

is bounded on $L^p(\mathbb{T})$ for all $1 < p < \infty$, and the partial Fourier summation operators

$$S_N f(x) = \sum_{|n| \leqslant N} \hat{f}(n) e^{2\pi i n x}$$

are uniformly bounded in N as operators on $L^p(\mathbb{T})$ for $1 < p < \infty$.

4.9 Radial Functions

Suppose $f \in L^1(\mathbb{R}^d)$ is a radial function. Then $\hat{f}$ 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 $\hat{f}(\xi) = \tilde{u}(|\xi|)$. 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 $f_R(x) = \mathbb{I}_{|x| \leqslant R}$. Then*

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

The n TODO

92

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 $\mathbb{R}^d$. To begin with, we study the heat equation on $\mathbb{R}^d$, 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 $f \in L^p(\mathbb{R}^d)$, u is well defined by this formula, lies in $C^\infty(\mathbb{T}^d)$, 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 $x \in \mathbb{R}^d$ 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 $f \in L^1(\mathbb{R}^d)$, 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 $\mathbb{R}$ satisfying a decay condition

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

It follows that $f \in L^p(\mathbb{R}^d)$ for each $1 \leqslant p \leqslant \infty$. Suppose that in addition, $\widehat{f}$ is supported on $[-1/2, 1/2]$. It follows that, $\widehat{f} \in L^p(\mathbb{R}^d)$ 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 $f \in L^1(\mathbb{Z})$ because of it's decay, we find that for each $\xi \in \mathbb{R}$,

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

But now we conclude by the Fourier inversion formula that

$$f(x) = \int_{-1/2}^{1/2} \left(\sum_{n=-\infty}^{\infty} f(n)e^{-2\pi i n\xi} \right) e^{2\pi i \xi x} \, d\xi$$

$$= \sum_{n=-\infty}^{\infty} f(n) \int_{-1/2}^{1/2} e^{2\pi i \xi(x-n)} \, d\xi$$

$$= \sum_{n=-\infty}^{\infty} f(n) \cdot \frac{\sin(\pi(x-n))}{\pi(x-n)}.$$

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 $\hat{f}$ on $[-\lambda/2, \lambda/2]$, noting that

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

so that

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

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

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

Thus it follows from the Fourier inversion formula that

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{f(n/\lambda)}{\lambda} \int_{-\infty}^{\infty} \chi(\xi)e^{2\pi i \xi(x-n/\lambda)} \, dx$$

$$= \sum_{n=-\infty}^{\infty} \frac{f(n/\lambda)}{\lambda} \hat{\chi}(n/\lambda - x)$$

$$= \sum_{n=-\infty}^{\infty} f(n/\lambda) \frac{\cos(\pi(n/\lambda - x)) - \cos(\lambda\pi(n/\lambda - x))}{\pi^2\lambda(\lambda-1)(n/\lambda - x)^2}.$$

95

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 $f_1, f_2 \in \mathcal{S}(\mathbb{R})$, with $\hat{f_1}$ and $\hat{f_2}$ supported on $[-1/2, 1/2]$, such that $f_1(n/\lambda) = f_2(n/\lambda)$ for any $n \in \mathbb{Z}$.*

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

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

One can determine all the coefficients $\{f(n/\lambda)\}$ if one knows the right hand side for all values $\xi \in \mathbb{R}$. Thus if $f_1, f_2 \in \mathcal{S}(\mathbb{R})$ are distinct functions such that $\hat{f_1}$ and $\hat{f_2}$ 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 $n \in \mathbb{Z}$. This is certainly possible if $\lambda < 1$. □

We can also get a discretized L^2 identity.

Theorem 5.2. *Suppose $f \in L^2(\mathbb{R})$ and $\hat{f}$ 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

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

□

5.3 The Uncertainty Principle

The uncertainty principle is a *constraint* placed on the Fourier transform, which prevents a function and it's transform from concentrating too tightly in space. Roughly speaking, the uncertainty principle is the heuristic that if f is a function, and it's Fourier support is supported on a ball around the origin of radius $1/R$, then f should be 'locally constant' at a scale R, because intuitively, f is only composed of waves oscillating at low frequencies, and so in particular, should have large mass on some ball of radius R. This principle is seen in various scenarios, and we describe some basic situations here. This property is fundamental to the transform. For instance, we already see it in the formula for the Fourier transform of the dilation of a function, i.e. that

$$\mathcal{F} \circ \mathrm{Dil}_t = t^d \cdot \mathrm{Dil}_{1/t} \circ \mathcal{F}.$$

Thus dilation which shrinks the support of a function to a scale t, *expands* the support of the Fourier transform of the function by $1/t$. Another, version is the *Paley-Wiener theorem*: an exponentially decaying function must have a Fourier transform which is extendable to a holomorphic function on a neighborhood of the real line. The uncertainty principle appears because holomorphic functions cannot have support concentrated on small sets. Here we consider some other versions of the principle.

5.3.1 The Heisenberg Uncertainty Principle

The first version of the uncertainty principle we give here is the most famous, the *Heisenberg uncertainty principle*, which is an L^2 instance of the heuristic, and has important implications in quantum mechanics.

Theorem 5.3 (Heisenberg). *Suppose $\psi \in \mathcal{S}(\mathbb{R})$. Then for any $x_0, \xi_0 \in \mathbb{R}$,*

$$\left(\int_{-\infty}^{\infty} (x - x_0)^2 |\psi(x)|^2 \, dx \right) \left(\int_{-\infty}^{\infty} (\xi - \xi_0)^2 |\hat{\psi}(\xi)|^2 \, d\xi \right) \geqslant \frac{1}{16\pi^2} \left(\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx \right)^2.$$

Thus if ψ has L^2 norm A, then for any x_0 and ξ_0, the L^2 norm of $(x - x_0)\psi$ and $(\xi - \xi_0)\psi$ cannot simultaneously both be $\lesssim A$.

Proof. Normalizing, we may assume that $\int_{-\infty}^{\infty} |\psi(x)|^2 \, dx = 1$ and that $x_0, \xi_0 = 0$.

Integration by parts shows that

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

$$= -\int_{-\infty}^{\infty} x \frac{d|\psi(x)|^2}{dx} = -\int_{-\infty}^{\infty} \left(x\psi'(x)\overline{\psi(x)} + x\overline{\psi'(x)}\psi(x) \right) \, dx.$$

Thus

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

$$\leqslant 2 \left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx \right)^{1/2} \left(\int_{-\infty}^{\infty} |\psi'(x)|^2 \, dx \right)^{1/2}$$

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

Remark. Taking ψ to be a standard Gaussian function shows the constant $1/16\pi^2$ in the Heisenberg uncertainty principle is tight.

Let us explain the applications of this uncertainty principle in quantum mechanics. Unlike in classical mechanics, 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

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

On the other hand, the *momentum* of the particule is also random, given in terms of a function ϕ such that the distribution of the momentum is given with density $|\varphi|^2$. Thus the variance of the momentum, if ξ_0 is the expectation, is equal to

$$\sigma_\rho^2 = \int_{-\infty}^{\infty} |\rho - \rho_0|^2 |\varphi(\rho)|^2 \, d\xi.$$

The relation between φ and ψ is given by the *De-Broglie* relation, which informally says the momentum of a particle is proportional to it's frequency, and formally says that

$$\varphi(\rho) = h^{-d/2} \int_{\mathbb{R}^d} e^{-2\pi i \rho \cdot x/h} \psi(x) \, dx = h^{-d/2} \hat{\psi}(\rho/h),$$

where $h > 0$ is Planck's constant. Thus Heisenberg's uncertainty principle tells us precisely that

$$\sigma_x \cdot \sigma_\rho \geqslant h/4\pi = \hbar/2.$$

Thus we have a fundamental limitation to the minimum possible uncertainty of the position and momentum of a particle.

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 $f \in \mathcal{S}(\mathbb{R})$,

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

The Heisenberg uncertainty principle thus implies that

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

Thus the operator $L - 1$ is formally positive definite. If we consider the operator

$$Af = \frac{d}{dx} + x \quad \text{and} \quad A^* = -\frac{d}{dx} + x$$

then $A^*A = L - 1$. 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

$$Df = (2\pi i)^{-1} f'(x)$$

both continuous, self adjoint operators on $\mathcal{S}(\mathbb{R})$. Given an arbitrary self-adjoint operator A on $\mathcal{S}(\mathbb{R})$, and $f \in \mathcal{S}(\mathbb{R})$ with $\|f\|_{L^2(\mathbb{R})} = 1$, we define the *standard deviation* of the operator to be

$$\sigma_A^2 f = \left(\langle Af, Af \rangle - \langle Af, f \rangle^2 \right)^{1/2}$$

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

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

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

5.3.2 Bernstein's Inequality

You can measure how 'locally constant' a function is in terms of the smoothness of the function, i.e. if the derivative is small, then the function is roughly speaking, locally constant. Roughly speaking, if a function f is constant at a scale R, then Df should have magnitude on average equal to $1/R$. The Fourier transform of Df is also supported on this ball, so we can iterate this argument, we should expect $D^k f$ to have magnitude on average equal to $1/R^k$. Formulating a precise version of the uncertainty principle here gives Bernstein's inequality. Let's begin with an L^2 version which is easiest to prove, since we have Plancherel.

Lemma 5.4. *If* $f \in L^2(\mathbb{R}^d)$ *and* $\hat{f}$ *is supported on a ball of radius* $1/R$ *about the origin, then for any multi-index* α,

$$\|D^\alpha f\|_{L^2(\mathbb{R}^d)} \leqslant (2\pi/R)^{|\alpha|} \|f\|_{L^2(\mathbb{R}^d)}.$$

100

Proof. Plancherel tells us that

$$\|D^\alpha f\|_{L^2(\mathbb{R}^d)} = \|(2\pi i\xi)^\alpha \hat{f}\|_{L^2(\mathbb{R}^d)}.$$

Since $\hat{f}$ is supported on a ball of radius R, we conclude that on this support, $(2\pi i\xi)^\alpha$ has magnitude at most $(2\pi R)^{|\alpha|}$, and thus

$$\|(2\pi i\xi)^\alpha \hat{f}\|_{L^2(\mathbb{R}^d)} \leqslant (2\pi/R)^{|\alpha|}\|\hat{f}\|_{L^2(\mathbb{R}^d)} = \|f\|_{L^2(\mathbb{R}^d)}. \qquad \square$$

There is a version in more general L^p spaces, but the proof is a little more technical.

Lemma 5.5. *If $f \in L^p(\mathbb{R}^d)$ and $\hat{f}$ is supported on a ball of radius $1/R$ about the origin, then for any multi-index α, and any $q \leqslant p$,*

$$\|D^\alpha f\|_{L^p(\mathbb{R}^d)} \lesssim R^{d(1/p-1/q)-|\alpha|}\|f\|_{L^q(\mathbb{R}^d)}.$$

Proof. Choose a function $\psi \in \mathcal{S}(\mathbb{R}^d)$ which has a compactly support Fourier transform which is equal to one on the ball of radius 1 about the origin. If $\psi_R(x) = R^{-d}\psi(x/R)$, then the Fourier transform of ψ_R is equal to one on a ball of radius $1/R$ about the origin. Now we have smoothness bounds on ψ_R, namely

$$|(D^\alpha \psi_R)(x)| \lesssim_{\alpha,N} R^{-d-\alpha}(x/R)^{-N}.$$

Moreover, $f = f * \psi_R$, so the convolution should show that the smoothness of ψ_R transfers to give the smoothness of f. More precisely, Young's inequality implies that if $1/p + 1 = 1/q + 1/r$, then

$$\begin{aligned}
\|D^\alpha f\|_{L^p(\mathbb{R}^d)} &= \|f * D^\alpha \psi_R\|_{L^p(\mathbb{R}^d)} \\
&\lesssim \|f\|_{L^q(\mathbb{R}^d)}\|D^\alpha \psi_R\|_{L^r(\mathbb{R}^d)} \\
&= R^{d(1/r-d)-\alpha}\|f\|_{L^q(\mathbb{R}^d)} \\
&= R^{d(1/p-1/q)-\alpha}\|f\|_{L^q(\mathbb{R}^d)}. \qquad \square
\end{aligned}$$

Here is another version of the inequality. Recall that for discrete quantities, $l^q \leqslant l^p$ for $p \leqslant q$. The uncertainty principle says that if the Fourier support of a function is concentrated, then it should be locally constant, and thus, roughly speaking, discrete. Thus we should expect that for such functions, we have $L^q \lesssim L^p$ for $p \leqslant q$. Making this precise leads to Bernstein's inequality. The result immediately follows from the last argument by taking $\alpha = 0$ in the last argument,

but we give an independent argument that involves the l^p and l^q norms directly. First we prove a pointwise version of the uncertainty principle. Let B and B^* be *dual ellipsoids*, i.e. ellipsoids with the same axes, and recipricol axis lengths. For a fixed $N > 0$, and a given ellipsoid B with center x_0, axes given by vectors $v_1, \ldots, v_d$, and axis lengths $r_1, \ldots, r_d > 0$, define

$$w_B(x) = \left\langle \sum_j \frac{|(x - x_0) \cdot v_j|^2}{r_j^2} \right\rangle^{-N}.$$

Then w_B is concentrated on B.

Lemma 5.6. *Suppose $\hat{f}$ is supported on a dual of an ellipsoid B. Then for any $x_0 \in B$, and any $x' \in B$,*

$$|f(x')| \lesssim_N \frac{1}{|B|} \int |f(x)| w_B(x) \, dx.$$

If $R > 0$, and $\hat{f}$ is supported on a ball of radius $1/R$, then for $x_0 \in B$, and any x,

$$|(D^\alpha f)(x)| \lesssim_{\alpha, N} R^{-|\alpha|} \left\langle \frac{x - x_0}{R} \right\rangle^d (Mf)(x_0).$$

Proof. Applying rescaling and translation symmetries, assume $x_0 = 0$, and $R = 1$. We have $f = f * \psi$, where ψ is as in the last argument. Thus

$$|f(x')| \leq \int |f(x)| |\psi(x' - x)| \, dx.$$

Since $|x'| \leq 1$, $\langle x' - x \rangle \geq \langle x \rangle$, and so using the bound $|\psi(y)| \lesssim_N w(y)$, we conclude that

$$|f(x')| \lesssim_N \int |f(x)| w(x) \, dx.$$

This proves the first inequality, for the second inequality, we may again assume $R = 1$ and $x_0 = 0$. We perform the same calculation, instead using the bound $\langle x' - x \rangle \geq \langle x \rangle$ for $|x| \geq 2|x'|$, using a dyadic decomposition in x, and for $|x| \leq 2|x'|$, using the trivial bound $\langle x' - x \rangle \geq 1$. $\square$

Remark. This calculation often suffices if f is roughly speaking, a radial function, and the maximal averages of f near x_0 occur on balls of radius $\sim R$. More detailed decompositions and calculations might be necessary if these conditions fail.

The intuitive idea of the version of the proof using the fact that $l^q \leqslant l^p$ for $p \leqslant q$ intuitively precedes as follows. Without loss of generality, suppose that $\hat{f}$ is supported on a ball of radius $1/R$ at the origin. An intuitive proof proceeds as follows: the uncertainty principle tells us that f is locally constant on sidelength $O(R)$ cubes. Thus we can divide $\mathbb{R}^d$ into an almost disjoint union of sidelength cubes $\{I_k : k \in \mathbb{Z}^d\}$, where I_k has centre $R \cdot k$. Roughly speaking, we should be able to find constants $\{c_k\}$ such that

$$f \approx \sum c_k \chi_{I_k}.$$

It then follows that

$$\|f\|_{L^q(\mathbb{R}^d)} \approx R^{d/q} \|c\|_{l^q} \leqslant R^{d/q} \|c\|_{l^p} \approx R^{-(d/p-d/q)} \|f\|_{L^p(\mathbb{R}^d)}.$$

To make this argument work, we use the bound above. This result really says that for any ball B_R of radius R,

$$\|f\|_{L^\infty(B)} \lesssim_N R^{-d} \|f\|_{L^1(w_B)}\|.$$

Interpolation with the trivial inequalities

$$\|f\|_{L^1(B)} \lesssim_N \|f\|_{L^1(w_B)} \quad \text{and} \quad \|f\|_{L^\infty(B)} \lesssim_N \|f\|_{L^\infty(w_B)}$$

gives that for $p \leqslant q$, $\|f\|_{L^q(B)} \lesssim_N R^{d(1/p-1/q)} \|f\|_{L^p(w_B)}$. This is not quite a local version of Bernstein's inequality, since w_B is not compactly supported. But w_B has Schwartz tails, so this is a *pseudolocal* version of the inequality, which is often just as good in harmonic analysis. In particular, these bounds lead to another proof of the following lemma, by decomposing $\mathbb{R}^d$ into a family of almost disjoint cubes $\{I_k\}$ and then writing

$$\|f\|_{L^q(\mathbb{R}^d)} \sim \big\| \|f\|_{L^q(I_k)} \big\|_{l_k^q} \lesssim_N R^{d(1/p-1/q)} \big\| \|f\|_{L^p(w_{I_k})} \big\|_{l_k^p}.$$

The rapid decay of the weights $\{w_{I_k}\}$ allows us to get a bound of the form

$$\big\| \|f\|_{L^p(w_{I_k})} \big\|_{l_k^p} \lesssim \big\| \|f\|_{L^p(I_k)} \big\|_{l_k^p} \sim \|f\|_{L^p(\mathbb{R}^d)},$$

which completes the proof. More precisely, we have

$$w_{I_k} \lesssim \sum_{m=0}^{\infty} 2^{-Nm} \chi_{2^m I_k}.$$

103

The triangle inequality implies that

$$\|f\|_{L^p(w_{I_k})} \lesssim \sum_{m=0}^{\infty} 2^{-Nm} \sum_{m=0}^{\infty} \|f\|_{L^p(2^m I_k)} \lesssim \sum_{m=0}^{\infty} 2^{-Nm} \sum_{|a| \leqslant 2^m} \|f\|_{L^p(I_{k+a})}.$$

Finally, the triangle inequality then implies that

$$\begin{aligned}
\left\| \|f\|_{L^p(w_{I_k})} \right\|_{l_k^p} &\leqslant \sum_{m=0}^{\infty} 2^{-Nm} \sum_{|a| \leqslant 2^m} \left\| \|f\|_{L^p(I_{k+a})} \right\|_{l_k^p} \\
&= \sum_{m=0}^{\infty} 2^{-Nm} 2^{dm} \left\| \|f\|_{L^p(I_{k+a})} \right\|_{l_k^p} \\
&\sim \sum_{m=0}^{\infty} 2^{-Nm} 2^{dm} \|f\|_{L^p(\mathbb{R}^d)}.
\end{aligned}$$

Taking $N > d$ gives the required bound.

There are extensions of these results to *ellipsoids* rather than balls. One can reformulate the uncertainty principle as follows: if $\hat{f}$ is supported in the *dual* of an ellipsoid E (obtained by inverting the ellipse on each axis), then f is locally constant of translates of the ellipsoid E. It is simple to obtain these results from the results above using balls, by using the way the Fourier transforms under changing coordinates linearly by an element of $GL(d)$. One can also use rectangles, and dual rectangles (we snuck in some applications using cubes instead of balls in the proofs above), since these shapes all have comparable size to their round counterparts.

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 $\mathbb{R}^n$, which can be seen as a measure $\mathbf{P}_X$ on $\mathbb{R}^n$, with $\mathbf{P}_X(E) = \mathbf{P}(X \in E)$. Given two independent random vectors X and Y, $\mathbf{P}_{X+Y}$ is the convolution $\mathbf{P}_X * \mathbf{P}_Y$ between the measures $\mathbf{P}_X$ and $\mathbf{P}_Y$, in the sense that

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

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

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

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

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

Lemma 5.7. *Let X and Y be independent 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 5.8. *If X and Y are independent 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 5.9. *Let $X_1, \ldots, X_N$ be independent and identically distributed with mean zero and variance σ^2. If $S_N = X_1 + \cdots + X_N$, 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. $\square$

5.5   The Wirtinger Inequality on an Interval

Theorem 5.10. *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

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

Then $a_0 = 0$, and so

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

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

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

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

Corollary 5.11. *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 $\mathbb{R}^n$ 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 5.12. *If $f \in L^p(\mathbb{R}^n)$, 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 $\|u(\cdot, y)\|_{L^p(\mathbb{R}^n)} \leqslant \|f\|_{L^p(\mathbb{R}^n)}$. 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

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

$\square$

Chapter 6

Partial Derivatives and Harmonic Functions

6.1   Conjugate Poisson Kernel

Recall from our discussion of the Poisson kernel that for $f \in L^p(\mathbb{R})$ 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 $\mathbb{T}$, the functions take the form of the power maps $\phi_n : z \mapsto z^n$, for $n \in \mathbb{Z}$. 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 $\mathbb{T}$, 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 $\mathbb{T}$ 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 $n \in \mathbb{Z}$. 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 $\mathbb{Z}_N$ 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 $\mathbb{T}$ and periodic functions on $\mathbb{R}$. The characters of $\mathbb{Z}_n$ 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 $\mathbb{Z}_N$ form a group under multiplication, denoted $\mathbb{Z}_N^*$. Any character from $\mathbb{Z}_N^*$ is valued on the $\varphi(N)$'th roots of unity, because the order of each element in $\mathbb{Z}_N^*$ divides $\varphi(N)$. The groups are in general non-cyclic. For instance, $\mathbb{Z}_8^* \cong \mathbb{Z}_2^3$. 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 $\mathbb{T}$, and

111

therefore μ_N provides a good finite approximation to $\mathbb{T}$. Functions from μ_N to $\mathbb{C}$ are really just functions from $[n] = \{1, \ldots, n\}$ to $\mathbb{C}$, and since μ_N is isomorphic to $\mathbb{Z}_N$, 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 $\mathbb{Z}_N$.

There is a trick which we can use to obtain quick results about Fourier analysis on μ_N. Given a function $f : [N] \to \mathbb{C}$, 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 $f : [N] \to \mathbb{C}$ 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 $f : [N] \to \mathbb{C}$, 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 $f : [N] \to \mathbb{C}$, we have, if we set $\hat{f}(m) = \langle f, \phi_m \rangle$, then

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

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

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 $\mathbb{C}$. We shall let V be the space of all such functions from G to $\mathbb{C}$, 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. $\qquad\square$

113

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

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

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

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

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

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

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 $T_1, \ldots, T_{k-1}$. Putting this together, we simultaneously diagonalize $T_1, \ldots, T_k$. $\qquad\square$

This theorem enables us to prove the character theory in a much simpler manner. Let V be the space of complex valued functions on G, and define, for $a \in G$, the map $(T_a f)(b) = f(ab)$. V has an orthonormal basic consisting of the $\chi_a(b) = N[a = b]$, for $a \in G$. In this basis, we comcpute $T_a \chi_b = \chi_{ba^{-1}}$, hence T_a is a permutation matrix with respect to this basis, hence unitary. The operators T_a commute, since $T_a T_b = T_{ab} = T_{ba} = T_b T_a$. Hence these operators can be simultaneously diagonalized. That is, there is a family $e_1, \ldots, e_n \in V$ and $\lambda_{an} \in \mathbb{T}$ 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.

114

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 $f : G \to \mathbb{C}$,*

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

In this context, we also have Parseval's theorem

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

7.3 Convolutions

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

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

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

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

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

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

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

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

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

$$= \frac{1}{|G|}\left(\sum_{a} f(a)\overline{e(a)}\right)\frac{1}{|G|}\left(\sum_{b} g(b)\overline{e(b)}\right)$$

$$= \hat{f}(e)\hat{g}(e)$$

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

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

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

$$\sum_{e \in \Gamma(G)} e(a) = 0$$

Now $f * D = f$, because

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

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

7.4   The Fast Fourier Transform

The main use of the fourier series on μ_n in applied mathematics is to approximate the Fourier transform on $\mathbb{T}$, where we need to compute integrals explicitly. If we have a function $f \in L^1(\mathbb{T})$, then f may be approximated in $L^1(\mathbb{T})$ by step functions of the form

$$f_n(t) = \sum_{k=1}^{n} a_k \, \mathbb{I}\big(x \in \big(2\pi(k-1)/n, 2\pi k/n\big)\big)$$

And then $\hat{f}_n \to \hat{f}$ uniformly. The Fourier transform of f_n is the same as the Fourier transform of the corresponding function $k \mapsto a_k$ on $\mathbb{Z}_n$, and thus we can approximate the Fourier transform on $\mathbb{T}$ by a discrete computation on $\mathbb{Z}_n$. Looking at the formula in the definition of the discrete transform, we find that we can compute

the Fourier coefficients of a function $f : \mathbb{Z}_n \to \mathbb{C}$ 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 $\mathbb{Z}_{2n}$. Given $f : \mathbb{Z}_{2n} \to \mathbb{C}$, define two functions $g, h : \mathbb{Z}_n \to \mathbb{C}$, 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 $\mathbb{Z}_{2n}$, we have

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

Because

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

$$= \frac{1}{2n} \sum_{k=1}^{2n} f(k)\nu^{-km}$$

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 $s > 1$,*

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

Proof. The general idea is this – we may write

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

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

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

$\square$

Chapter 8

Harmonic Functions

In this chapter, we illustrate the intimate connection between the Fourier transform on the real line, and complex analysis, as well as connections to the theory of harmonic functions. 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   Harmonic Functions

Fix a connected, open set $\Omega \subset \mathbb{R}^d$. A function $u \in C^2(\Omega)$ is said to be *harmonic* if $\Delta u = 0$ on Ω. Examples include *harmonic polynomials*, like all linear functions, and $u(x, y) = x^2 - y^2$. The mean value property of harmonic functions is essential to their study.

Theorem 8.1. *Let u be harmonic on $\Omega \subset \mathbb{R}^d$. Then for all $z \in \Omega$ and $0 < r < d(z, \partial\Omega)$,*

$$u(z) = \fint_{\mathbb{T}} u(z + re^{it}) \, dt.$$

Proof. One simply takes the derivative of the right hand side, applies the divergence theorem to see the derivative is zero, and then notes that the right hand side converges to $u(z)$ as $r \to 0$. $\qquad\square$

In particular, it implies the maximum principle.

Theorem 8.2. *Let u be harmonic on* Ω*. If u obtains an extremum on* Ω*, then u is constant.*

Proof. Suppose $u(z_0) \geqslant u(z)$ for all $z \in \Omega$. Applying the mean value property around z_0 shows that we must have $u(z) = u(z_0)$ for all z with $|z - z_0| < d(z_0, \partial\Omega)$. Thus we conclude that the set

$$S = \{z \in \Omega : u(z) = u(z_0)\}$$

is open and closed, and thus equal to Ω. $\qquad\square$

Remark. If Ω is bounded, and u extends to a continuous function on the closure of Ω, then the maximum principle says that for any $z \in \Omega$

$$\min_{\partial\Omega} u(z) \leqslant u(z) \leqslant \max_{\partial\Omega} u(z),$$

and either side can only be an equality if u is constant.

Remark. Note also that by replacing u with $-u$ in the theorem, one can obtain a corresponding 'minimum principle'

The maximum principle gives us a uniqueness result for the Dirichlet problem on a bounded domain.

Theorem 8.3. *Let* Ω *be a bounded domain. If u and v are harmonic on* Ω*, and extend to continuous functions on* $\overline{\Omega}$ *in such a way that* $u = v$ *on* $\partial\Omega$*, then* $u = v$*.*

We also have a version of Louiville's theorem for harmonic functions.

Theorem 8.4. *If u is harmonic on* $\mathbb{R}^d$ *and bounded, then u is constant.*

Proof. Fix $x \in \mathbb{R}^d$. Then for any $r > 0$,

$$u(x) = \frac{1}{r^n} \int_{|y| \leqslant r} u(y) \, dy.$$

But this means that for any $x_1, x_2 \in \mathbb{R}^d$, if B_1 and B_2 are the balls of radius r around x_1 and x_2,

$$|u(x_1) - u(x_2)| \lesssim \frac{1}{r^n} \int_{B_1 \Delta B_2} |u(y)| \, dy.$$

As $r \to \infty$, we have $|B_1 \Delta B_2| \lesssim r^{n-1}$, so taking $r \to \infty$ gives that $u(x_1) = u(x_2)$ for any $x_1, x_2 \in \mathbb{R}^d$. $\qquad\square$

Any function satisfying the mean value theorem is harmonic, as we now show.

Theorem 8.5. *Suppose u is continous on* Ω *and satisfies the mean value property on* Ω. *Then u is twice differentiable, and harmonic in* Ω.

Proof. To begin with, we may suppose that u is apriori twice differentiable. To see why, for any x, we can find a neighborhood of x such that u is given by convolution with a smooth bump function in that neighborhood, which actually shows that u is C^∞. Now fix $x \in \Omega$, and for sufficiently small r, let

$$M(r) = \fint_{|y-x|=r} u(y) \, dy.$$

Our proof will be complete if we can show that $M''(0) = \Delta u(x)/n$. But we can do this by taking a Taylor series when r is small, i.e. we find that

$$M(r) = \fint_{|y-x|=r} \left[u(x) + \left(\sum_i \partial_i u(x) \cdot y_i \right) + \left(\sum_{i,j} \partial_{i,j} u(x) \cdot y_i y_j \right) + o(|y|^2) \right] d\sigma(y).$$

By symmetry considerations, we thus find that

$$M(r) = u(x) + \nabla u(x) \cdot x + \sum_i \partial_i^2 u(x) x_i^2 + \fint_{|y|=r} \left(\sum_i \partial_i^2 u(x) y_i^2 \, d\sigma(y) \right) + o(r^2)$$

$$= C + r^2/2 \sum_i \partial_i^2 u(x) \fint_{|y|=1} y_i^2 \, d\sigma(y) + o(r^2).$$

Now for any i and j, we have

$$\fint_{|y|=1} y_i^2 = \fint_{|y|=1} y_j^2.$$

But this means that

$$\fint_{|y|=1} y_i^2 = \frac{1}{n} \fint_{|y|=1} \sum_j y_j^2 = 1/n.$$

Thus we find that $M(r) = C + (r^2/2n)\Delta u(x) + o(r^2)$. Taking second derivatives gives the required claim. $\square$

This condition makes it easy to check (as with the Cauchy integral formula for harmonic functions) that the locally uniform limit of a sequence of harmonic functions is also harmonic.

8.2 Poisson Kernels

Consider the classic Dirichlet problem on a bounded domain $\Omega \subset \mathbb{R}^d$ with C^1 boundary. Given a continuous function b defined on $\partial\Omega$, is it possible to find a harmonic function u on Ω which extends to a continuous function on $\overline{\Omega}$ in such a way that $u = b$ on $\partial\Omega$? We have already seen that u, if it exists, must be unique.

Let us write the relation of u and b in operator form, i.e. writing $u = Lb$ for some operator L. We use the method of *Green's functions* to derive an integral expression for the operator L. Let Φ be the fundamental solution for the Laplacian on $\mathbb{R}^d$, i.e. the distribution such that $\Delta\Phi = \delta$. Applying Green's theorem yields that if u is harmonic in Ω, then for $x \in \Omega$,

$$u(x) = \int_\Omega u(y)\Delta\Phi(y-x) \, dy$$

$$= \int_\Omega [u(y)\Delta\Phi(y-x) \, dy - \Phi(y-x)\Delta u(y)] \, dy$$

$$= \int_{\partial\Omega} [u(y)\partial_\eta\Phi(y-x) - \partial_\eta u(y)\Phi(y-x)] \, dy.$$

This almost gives an expression for u given it's boundary values, expect we must know the normal derivatives $\partial_\eta u$, which is not available to us. To fix this, we find a harmonic function ϕ^x on Ω for each $x \in \Omega$, subject to the boundary conditions that $\phi^x(y) = \Phi(y-x)$ for $y \in \partial\Omega$. It then follows that

$$0 = \int_\Omega u(y)\Delta\phi^x(y) - \Delta u(y)\phi^x(y)$$

$$= \int_{\partial\Omega} u(y)\partial_\eta\phi^x(y) - \partial_\eta u(y)\phi^x(y).$$

But this means that

$$u(x) = \int_{\partial\Omega} u(y) \left[\partial_\eta\Phi(y-x) - \partial_\eta\phi^x(y)\right] \, dy.$$

Thus we conclude that the *Poisson kernel*, i.e. the kernel of the operator L, is given by

$$K(x,y) = \partial_\eta\Phi(y-x) - \partial_\eta\phi^x(y).$$

In general, for a very general family of domains Ω it is possible to find a family of functions $\{\phi^x\}$, though it is only in certain nice, symmetry domains that one can

find explicit formula for these functions, and thus to have an explicit formula for the Poisson kernel K.

The easiest Green's function to find is for the half space

$$\mathbf{H} = \{x \in \mathbb{R}^n : x_n > 0\}.$$

To find the Green's function, we rely on a reflection trick. We define

$$\phi^x(y) = \Phi(y - \tilde{x}),$$

where $\tilde{x}$ is obtained from x by reflecting it across the line defining the half plane. Noting that for $y \in \partial\mathbf{H}$, $|y - \tilde{x}| = |y - x|$, we conclude that the Poisson kernel is given by

$$K(x,y) = \frac{1}{|S^{n-1}|} \left[-\frac{y_n - x_n}{|y - x|^n} + \frac{y_n + x_n}{|y - \tilde{x}|^n} \right]$$

$$= \frac{2}{|S^{n-1}|} \frac{x_n}{|y - x|^n}.$$

Thus we should expect that, for a function b defined on $\mathbb{R}^{n-1}$, we have

$$Lb(x) = \int_{\partial\mathbf{H}} \frac{2x_n}{|S^{n-1}|} \frac{1}{|x - y|^n} b(y) \, dy.$$

This is Poisson's formula for the half plane.

Next, let us find the Green's function for the unit ball B in $\mathbb{R}^n$. Here we use a similar inversion trick. Given x in the unit ball, we let $\tilde{x} = x/|x|^2$ be the *dual point* to x, i.e. the point obtained by *spherical inversion* about the unit sphere. For $y \in \partial B$, we have

$$|y - \tilde{x}|^2 = |y|^2 + |\tilde{x}|^2 - 2y \cdot \tilde{x} = 1 + \frac{1}{|x|^2} - \frac{2y \cdot x}{|x|^2} = \frac{1}{|x|^2} \left(|x|^2 + 1 - 2y \cdot x \right) = \frac{|y - x|^2}{|x|^2}.$$

Thus the function $\phi^x(y) = \Phi(|x|(y - \tilde{x}))$ has the right boundary conditions. Moreover, it is harmonic, for $n \geq 3$ because Φ is homogeneous, so

$$\phi^x(y) = \frac{\Phi(y - \tilde{x})}{|x|^{n-2}},$$

and for $n = 2$ because

$$\phi^x(y) = \Phi(y - \tilde{x}) - \frac{\ln|x|}{2\pi},$$

123

and both choices are harmonic on B. Regardless, since, for $|y| = 1$, we have $|y - \tilde{x}| = |y - x||x|^{-1}$, we conclude that the Poisson kernel for the unit ball is given by

$$
\begin{aligned}
K(x, y) &= \frac{-1}{|S^{n-1}|} \left[-\frac{1 - y \cdot x}{|y - x|^n} + \frac{1}{|x|^{n-2}} \frac{1 - y \cdot \tilde{x}}{|y - \tilde{x}|^n} \right] \\
&= \frac{-1}{|S^{n-1}|} \frac{1}{|y - x|^n} \left[-(1 - y \cdot x) + (|x|^2 - y \cdot x) \right] \\
&= \frac{1}{|S^{n-1}|} \frac{1 - |x|^2}{|y - x|^n}.
\end{aligned}
$$

Thus we should expect that, for a function b defined on S^{n-1}, we have

$$
Lb(x) = \frac{1}{|S^{n-1}|} \int_{S^{n-1}} \frac{1 - |x|^2}{|y - x|^n} b(y) \, dy.
$$

This is Poisson's formula for the unit ball.

8.3   Holomorphic Functions and the Poisson Kernel

If $\Omega \subset \mathbb{R}^2$, and $f : \mathbb{R}^2 \to \mathbb{C}$, viewed as a function with domain $\mathbb{C}$, is holomorphic, and $f = u + iv$ for two real functions u and v, then u and v are both harmonic. Conversely, if Ω is *simply connected*, then for any harmonic function u on Ω, we can find a harmonic v on Ω such that $u + iv$ is harmonic. To do this, we note that the differential form

$$
\frac{\partial u}{\partial x} \, dy - \frac{\partial u}{\partial y} \, dx
$$

is exact, so that there exists a function v such that

$$
\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \quad \text{and} \quad \frac{\partial v}{\partial y} = \frac{\partial u}{\partial x}.
$$

These are precisely the Cauchy-Riemann equations, so that we see $u + iv$ is harmonic. In particular, we see that this implies harmonic functions are automatically analytic, i.e. they lie in $C^\infty(\Omega)$, and their Taylor series expansions converge locally uniformly to u.

Chapter 9

Complex-Variable Methods

9.1 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 9.1. *Let $f : S_a \to \mathbb{C}$ be holomorphic, integrable on each horizontal line in the strip, such that $f(x + iy) \to 0$ as $|x| \to \infty$. Then if $\hat{f}$ is the Fourier transform of the restriction of f to the real line, then for each $b < a$,*

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

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

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

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

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

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

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

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

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

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 9.2. *A function $f : \mathbb{R} \to \mathbb{C}$ 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 $x \in \mathbb{R}$,

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

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 $z \in \mathbb{C}$,

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

126

where $g \geqslant 0$ is integrable on $\mathbb{R}$. 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,

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

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 $x \in \mathbb{R}$, 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

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

127

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

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

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 $|g(z)| \leqslant \max(A, \|f\|_{L^\infty(\mathbb{R})})$ for all z on the quandrant. A similar argument works for the other quadrants. Thus we conclude that for all $z \in \mathbb{C}$

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

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

$\square$

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 $\xi < 0$.

9.2 Classical Theorems by Contours

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

Theorem 9.3. *Let $f : S_b \to \mathbb{C}$ be holomorphic. Then for any $x \in \mathbb{R}$,*

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

128

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

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

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

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

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

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

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

Similarily, another application of Fubini's theorem implies

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

In particular, we conclude that

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

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 $f(x)$. $\qquad\square$

129

We can also prove the Poisson summation formula.

Theorem 9.4. *Let* $f : S_a \to \mathbb{C}$ *be holomorphic. Then*

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

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 $\mathbb{Z}$, with reside equal to $f(n)$ at each $n \in \mathbb{Z}$. 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 $\sum_{n \in \mathbb{Z}} f(n)$ 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

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

130

to conclude that

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

$$= \sum_{n=0}^{\infty} \hat{f}(-n).$$

Combining these two calculations completes the proof. □

We can also prove the boundedness of the Hilbert transform on $L^{2k}(\mathbb{R})$, for all even integers k, which proves L^p bounds for all $2 \leqslant p < \infty$ by interpolation, and then proves L^p bounds for $1 < p \leqslant 2$ by duality.

Theorem 9.5. *For any k, the Hilbert Transform*

$$Hf(x) = \frac{1}{\pi} \int \frac{f(t)}{t-x} \, dt$$

is bounded on $L^{2k}(\mathbb{R})$.

Proof. Without loss of generality, let f be Schwartz. Then the *Cauchy integral*

$$Cf(z) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{f(t)}{t-z} \, dt$$

is well defined for $z \in \mathbb{C} - \mathbb{R}$ and holomorphic. One can prove using an approximation argument that

$$Cf(x+i0) = \frac{f(x) + iHf(x)}{2},$$

and a contour shift argument shows that

$$\int_{-\infty}^{\infty} Cf(x+i0)^{2k} \, dx = 0.$$

H is bounded on $L^2(\mathbb{R})$, i.e. because it is a Fourier multiplier with bounded symbol. To obtain boundedness in $L^{2k}(\mathbb{R})$ for $k > 2$, assume $\|f\|_{L^{2k}(\mathbb{R})} = 1$ and $B = \|Hf\|_{L^{2k}(\mathbb{R})}$. We may assume f is Schwartz, so that both these quantities are easily seen to be finite. If we expand the integral above, and rearrange, we conclude that

$$\left| \int_{-\infty}^{\infty} (Hf)^{2k} \right| \leqslant C_k \sum_{j=0}^{2k-1} \left| \int (Hf)^j f^{2k-j} \right|.$$

131

Applying Hölder's inequality, we obtain that

$$B^{2k} \leqslant C_k \sum_{j=0}^{2k-1} B^j.$$

This is only possible if $B \lesssim_k 1$, i.e. because the polynomial $x^{2k} - C(\sum_{j<2k} x^j)$ is eventually positive for suitably large x. Thus we see $\|Hf\|_{L^{2k}(\mathbb{R})} \lesssim_k 1$. $\qquad \square$

9.3   The Laplace Transform

We now look at things from a dual perspective as in the first section. Instead of looking at whether a function can be extended to a holomorphic function, we look at whether the Fourier transform of a function can be extended to a holomorphic function. For a function $x : \mathbb{R} \to \mathbb{R}$, 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 $\xi \in \mathbb{R}$, $X(i\xi) = \hat{x}(\xi)$ 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 $I \subset \mathbb{C}$, 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 trans-
form 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 deriva-
tive, 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 $w \in \mathbb{C}$, 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, $t_0 \in \mathbb{R}$, 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

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

133

if $y(t) = x'(t)$, and if $z = \xi + i\omega$ for some $\xi \in \mathbb{R}$, 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 $t_1, \ldots, t_N$, such that the left and right-hand limits exist at each t_i. For each $i \in \{1, \ldots, N\}$, 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* $x : [0, \infty) \to \mathbb{R}$ *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*

$$\mathcal{L}(x'')(z) = z^2 X(z) - 6z - 1 \quad \text{and} \quad \mathcal{L}(x')(z) = zX(z) - 6.$$

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.

9.4 Asymptotics via the Laplace Transform

For simplicity, in this chapter we study integrable functions $x : [0, \infty) \to \mathbb{R}$, 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 $t \in \mathbb{R}$,

$$x(t) = \int_{-\infty}^{\infty} X(i\xi) e^{i\xi t} \, d\xi.$$

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 $a \in \mathbb{R}$, we let $\mathcal{E}_a$ be the set of functions whose region of convergence contains $\omega + i\xi$ for all $\omega > a$.

Theorem 9.6. *Suppose $x : [0, \infty) \to \mathbb{R}$ 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

$$x(t) = \lim_{R \to \infty} \frac{1}{2\pi}$$
$$X(z) = \lim$$

$\square$

136

Chapter 10

Spherical Harmonics

One of the main advantages with doing harmonic analysis with $\mathbb{R}^d$ rather than with a general Abelian group G is that $\mathbb{R}^d$ has a much greater family of symmetries then such a general group (which a priori only has translational symmetries). In this chapter, we exploit the *rotational symmetry* of $\mathbb{R}^d$ to obtain further information about the Fourier transform.

Each orthogonal matrix R commutes with the Fourier transform, i.e. for any function f, if $R^* f(x) = f(Rx)$, then we have

$$\widehat{R^* f} = R^* \hat{f}.$$

It follows from this that the Fourier transform of any radial function is radial. If we let $\mathfrak{H}^0(\mathbb{R}^n)$ denote the space of all square integrable radial functions, then $\mathfrak{H}^0(\mathbb{R}^n)$ is an invariant subspace of the Fourier transform on $L^2(\mathbb{R}^n)$ with respect to the Fourier transform. The theory of *spherical harmonics* will allow us to continue this development, writing

$$L^2(\mathbb{R}^n) = \bigoplus_k \mathfrak{H}^k(\mathbb{R}^n),$$

where $\mathfrak{H}^k(\mathbb{R}^d)$ is an invariant subspace of the Fourier transform, and also satisfies good symmetry properties with respect to orthogonal matrices.

The simplest version of this study is on the real line, where $\mathfrak{H}^0(\mathbb{R})$ corresponds to the *even* square integrable functions. The orthogonal complement of this space is precisely the *odd* square integrable functions, i.e. the space $\mathfrak{H}^1(\mathbb{R})$, since one can in general decompose a function in $L^2(\mathbb{R})$ into it's even and odd part, and these functions are orthogonal to one another. There are only two elements of

$O(1)$, namely, the identity, and the reflection operator $R_0 x = -x$, and we find that these spaces are characterized by this operator, i.e.

$$\mathfrak{H}^0(\mathbb{R}) = \{f \in L^2(\mathbb{R}) : Rf = f\} \quad \text{and} \quad \mathfrak{H}^1(\mathbb{R}) = \{f \in L^2(\mathbb{R}) : Rf = -f\}.$$

Thus we have found an orthogonal complement to the space $\mathfrak{H}^0(\mathbb{R})$ of radial functions, which behaves nicely with respect to rotations.

The space $\mathbb{R}^2$, which we identify with $\mathbb{C}$, is also tractable given the techniques we already know. Given $f \in \mathcal{S}(\mathbb{R}^2)$, Fubini's theorem implies that, the function $f_r : S^1 \to \mathbb{C}$ given by $f_r(x) = f(rx)$ is square integrable, and moreover, the map $r \mapsto f_r$ is continous from $(0, \infty)$ to $L^1(S^1)$. It follows that there are continuous functions $\{a_k : (0, \infty) \to \mathbb{C}\}$ such that

$$f(re^{i\theta}) = \sum_k a_k(r) e^{ki\theta},$$

such that for any fixed $r > 0$, the series on the right converges in the L_θ^2 norm, and moreover,

$$\int_{S^1} |f(re^{i\theta})|^2 \, d\theta = \sum_k |a_k(r)|^2.$$

But by monotone convergence, this implies that

$$\int_{\mathbb{R}^2} |f(x)|^2 \, dx = \int_0^\infty \int_{S^1} r|f(re^{i\theta})|^2 \, d\theta = \sum_k \int_0^\infty r|a_k(r)|^2 \, dr.$$

In particular, if we define a sequence $\{f_k\}$ by setting $f_k(re^{i\theta}) = a_k(r)e^{ki\theta}$, then we find that

$$\|f\|_{L^2(\mathbb{R}^2)} = \sum_k \|f_k\|_{L^2(\mathbb{R}^2)}^2,$$

and that $f = \sum f_k$. The boundedness of this decomposition shows that this decomposition exists for any $f \in L^2(\mathbb{R})$, and that for each such function f, we have

$$f_k(re^{i\theta}) = a_k(r)e^{ki\theta},$$

where the function a_k is defined analogously to the Schwartz case. We have therefore found a decomposition

$$L^2(\mathbb{R}^2) = \bigoplus_k \mathfrak{H}^k(\mathbb{R}^2),$$

138

where

$$\mathfrak{H}^k(\mathbb{R}^2) = \{f \in L^2(\mathbb{R}^2) : f(re^{i\theta}) = a(r)e^{ki\theta} \text{ for some function } a\}.$$

If we parameterize $SO(2)$ by defining $R_\theta(x) = e^{i\theta}x$, then $\mathfrak{H}^k$ can also be written as

$$\{f \in L^2(\mathbb{R}^2) : R_\theta^* f = e^{ki\theta} f\}.$$

This implies that $\mathfrak{H}^k$ is preserved under the Fourier transform, i.e. since if $R_\theta^* f = e^{ki\theta} f$, then taking Fourier transforms on both sides of this equation yields that

$$R_\theta^* \widehat{f} = e^{ki\theta} \widehat{f}.$$

Thus we have decomposed $L^2(\mathbb{R}^2)$ into a family of orthogonal subspaces that behave nicely with respect to rotations.

If $f \in \mathfrak{H}^k(\mathbb{R}^2)$, with $f(re^{i\theta}) = a(r)e^{ki\theta}$, then there exists a function b such that $\widehat{f}(\rho e^{it}) = b(\rho)e^{kit}$, and we can determine b in terms of a. Indeed, we have

$$\begin{aligned}
b(\rho) &= (-i)^k \widehat{f}(i\rho) \\
&= (-i)^k \int_0^\infty \int_0^{2\pi} ra(r)e^{ki\theta}e^{-2\pi ir\rho\sin(\theta)} \, d\theta \, dr \\
&= (-i)^k \int_0^\infty \int_0^{2\pi} ra(r)e^{-ki\theta}e^{2\pi ir\rho\sin(\theta)} \, d\theta \, dr.
\end{aligned}$$

If we introduce the Bessel functions

$$J_k(t) = \frac{1}{2\pi} \int_0^{2\pi} e^{it\sin\theta}e^{-ki\theta} \, d\theta,$$

then we conclude that

$$b(\rho) = (2\pi)(-i)^k \int_0^\infty ra(r)J_k(2\pi r\rho) \, dr.$$

A simple change of variables shows that $J_{-k}(t) = (-1)^k J_k(t)$, so we can alternatively write

$$b(\rho) = (2\pi)i^k \int_0^\infty ra(r)J_{-k}(2\pi r\rho) \, dr.$$

Let us now move on to $\mathbb{R}^n$, where we must replace Fourier series with more general *spherical harmonics*.

We obtained an expansion on $\mathbb{R}^2$ by first performing an expansion on the spheres around the origin, which reduced to the study of Fourier series, since in $\mathbb{R}^2$ the sphere S^1 is precisely the torus. In other words, we first decomposed $L^2(S^1)$, and then used this to decompose $L^2(\mathbb{R}^2)$. In higher dimensions, to decompose $L^2(S^{n-1})$, we must rely on a slightly different perspective. To begin with, let us look at the Fourier series from a different perspective. A trigonometric polynomial on $\mathbb{T}$ of the form

$$t \mapsto \sum_{k=-N}^{N} a_k e^{kit}$$

can be identified with a restriction to S^1 of the meromorphic function

$$P(z) = \sum_{k=-N}^{N} a_k z^k.$$

If P is real-valued on S^1, then $a_{-k} = \overline{a_k}$, and thus P can be written as

$$\mathrm{Re}\left(a_0 + \sum_{k=1}^{N} a_k e^{kit}\right).$$

Thus P is the restriction to S^1 of a *harmonic polynomial* on $\mathbb{R}^2$. Any such polynomial can be decomposed as

$$P = P_0 + \cdots + P_N,$$

where $P_j(z) = \mathrm{Re}(a_k z^k)$. In general, we shall call a function $P : S^{n-1} \to \mathbb{R}$ obtained by restricting a homogeneous harmonic polynomial of degree k on $\mathbb{R}^n$ to the unit sphere S^{n-1} a *spherical harmonic of degree k*. In general, we will find that we can decompose $L^2(\mathbb{R}^n)$ into spaces $\mathfrak{H}_k(\mathbb{R}^n)$, where $\mathfrak{H}_k$ is spanned by functions that are products of radial functions, and spherical harmonics of degree k.

Let $\mathcal{P}_k(\mathbb{R}^n)$ be the space of all harmonic homogeneous polynomials of degree k on $\mathbb{R}^n$. A simple counting argument shows this space has dimension

$$\frac{(n+k-1)!}{(n-1)!\,k!}.$$

We define an inner product on $\mathcal{P}_k$ by setting $\langle P, Q \rangle = P(D)\overline{Q}$. Since P and Q are both homogeneous of the same degree, $\langle P, Q \rangle$ is scalar valued. If $P = \sum_{|\alpha|=k} c_\alpha x^\alpha$, then

$$\langle P, P \rangle = \sum_{|\alpha|=k} |c_\alpha|^2 \alpha!$$

and so we see the inner product is positive definite.

Theorem 10.1. *If P is homogeneous of degree k, then there exists homogeneous harmonic polynomials P_j of degree $k - 2j$ for $j \leqslant k/2$, such that*

$$P = P_0 + |x|^2 P_1 + \cdots + |x|^{2l} P_l,$$

where l is the smallest integer smaller than 2l.

Proof. We may assume without loss of generality that $k \geqslant 2$, since any polynomial of degree < 2 is harmonic. Define $\varphi : \mathcal{P}_k \to \mathcal{P}_{k-2}$ by setting $\varphi(P) = \Delta P$. Then φ is onto, since otherwise we could find a nonzero polynomial $Q \in \mathcal{P}_{k-2}$ orthogonal to the range of φ with respect to the inner product defined above. But this means that if $P(x) = |x|^2 Q(x)$, then

$$0 = \langle Q, \Delta P \rangle = Q(D)\overline{\Delta P} = \Delta Q(D)\overline{P} = P(D)\overline{P} = \langle P, P \rangle.\rangle$$

But this implies $P = Q = 0$.

Let $\mathcal{A}_j$ be all harmonic polynomials in $\mathcal{P}_j$. We claim that $\mathcal{P}_j$ is the orthogonal direct sum of $\mathcal{A}_j$ and the space $\mathcal{B}_j = |x|^2 \mathcal{P}_{j-2}$. By induction, this would complete the proof. Certainly $\mathcal{A}_j$ and $\mathcal{B}_j$ are orthogonal, since if Q is harmonic, then

$$\langle |x|^2 P, Q \rangle = \Delta P(D)\overline{Q} = P(D)\overline{\Delta Q} = 0.$$

The fact that φ is onto implies that the dimensions of $\mathcal{A}_j$ and $\mathcal{B}_j$ add to the dimensions of $\mathcal{P}_j$, completing the argument. □

Remark. A consequence of this result is that the restriction of *any* not-necessarily homogeneous polynomial on $\mathbb{R}^n$ to S^{n-1} is a linear combination of homogeneous polynomials.

Let $\mathcal{H}^k$ be the space of all spherical harmonics of degree k. Note that the argument above shows that for $k \geqslant 2$,

$$
\begin{aligned}
\dim \mathcal{H}^k &= \dim \mathcal{A}^k \\
&= \dim \mathcal{P}_k - \dim \mathcal{P}_{k-2} \\
&= \binom{n+k-1}{k} - \binom{n+k-3}{k-2}.
\end{aligned}
$$

We can now calculate that $\dim \mathcal{H}^0 = 1$ and $\dim \mathcal{H}^1 = n$. In the particular case $n = 3$, we have that $\dim \mathcal{H}^k = 2k + 1$ for all $k \geqslant 0$. The space $\mathcal{A}^k$ is sometimes called the space of *solid spherical harmonics*, and $\mathcal{H}^k$ the space of *surface spherical harmonics*. We observe that the result above implies that the span of $\bigcup_k \mathcal{H}^k$

141

contains the restriction of all polynomials to S^{n-1}, and therefore, by an application of Stone-Weirstrass, is dense in $C(S^{n-1})$, and therefore, dense in $L^2(S^{n-1})$. If we can show that each of the spaces $\mathcal{H}^k$ are orthogonal to one another, we therefore have an orthogonal decomposition of $L^2(S^{n-1})$.

Using the inner product we have defined on the space of polynomials, we can also prove the following interesting result.

Theorem 10.2. *The space $\mathcal{A}^k$ is spanned by polynomials of the form $(w \cdot x)^k$, where $\sum w_i^2 = 0$.*

Proof. Certainly any polynomial of the form $(w \cdot x)^k$ is harmonic if $\sum w_i^2 = 0$. Conversely, if $Q \in \mathcal{A}^k$, and suppose that for any polynomial of the form $P_w(x) = (w \cdot x)^k$, where $\sum w_i^2 = 0$, we have $\langle Q, P_w \rangle = 0$. We calculate that

$$\langle Q, P_w \rangle = m! Q(c),$$

so we conclude that Q vanishes on $Z(x_1^2 + \cdots + x_n^2)$. Since the polynomial $x_1^2 + \cdots + x_n^2$ is irreducible, by Hilbert's Nullstellensatz, Q lies in the ideal generated by $x_1^2 + \cdots + x_n^2$. But we have seen that any polynomial in this family is orthogonal to $\mathcal{H}^k$, and so we conclude that $Q = 0$. $\qquad\square$

Next, we note the spaces $\mathcal{H}^k$ is orthogonal.

Theorem 10.3. *If Y^k and Y^l are spherical harmonics of degree k and l, with $k \neq l$, then*

$$\int_{S^{n-1}} Y^k(x) Y^l(x) \, dx = 0.$$

Proof. Let $u(rx) = r^k Y^k(x)$ and $v(rx) = r^l Y^l(x)$ for $|x| = 1$. Then u and v are homogeneous harmonic polynomials of degree k and l on $\mathbb{R}^d$. By Green's theorem,

$$0 = \int_{|x| \leqslant 1} u \Delta v - v \Delta u = \int_{|x| = 1} \left(u \frac{\partial v}{\partial \eta} - v \frac{\partial u}{\partial \eta} \right)$$

$$= \int_{|x| = 1} (l - k) Y^k(x) Y^l(x).$$

If $l \neq k$, we obtain the required result. $\qquad\square$

Alternatively, this follows from the following argument. Since S^{n-1} has a Riemannian structure, it naturally has a Laplace-Beltrami operator $\Delta_{S^{n-1}} f = \nabla \cdot \nabla f$, where the divergence and gradient are taken with respect to the metric on

S^{n-1}. If we identify $\mathbb{R}^n - \{0\}$ with $(0, \infty) \times S^{n-1}$, then it is a simple calculation to argue that

$$\Delta f = \frac{\partial_r (r^n \partial_r f)}{r^n} + \frac{\Delta_{S^{n-1}} f}{r^2}.$$

where $\Delta_{S^{n-1}}$ corresponds to the spherical Laplacian. As a result of this fact, for a spherical harmonic Y^k of degree k on S^{n-1}, we have

$$\Delta_{S^{n-1}} Y^k = -k(n + k - 1) Y^k.$$

Since Δ is self-adjoint, this implies the orthogonality property of the spherical harmonics above.

Another useful representation of S^n follows by identifying it with $S^{n-1} \times (-1, 1)$, such that for $\sigma \in S^{n-1}$, (σ, t) maps to $(\sin(\theta)\sigma, \cos(\theta))$. In this coordinate system, we have

$$\Delta_{S^n} = \frac{1}{\sin(\theta)^{n-1}} \frac{\partial}{\partial \theta} \left(\sin(\theta)^{n-1} \frac{\partial f}{\partial \theta} \right) + \frac{1}{\sin(\theta)^2} \Delta_{S^{n-1}}.$$

A calculation of this is provided in Jean Gallier's notes on Spherical Harmonics.

To review, we have an orthogonal decomposition

$$L^2(S^{n-1}) = \bigoplus_k \mathcal{H}^k.$$

Once a canonical orthonormal basis $Y_1^k, \ldots, Y_l^k$ is fixed, we can then write

$$f = \sum_{k=0}^{\infty} \sum_j a_{kj} Y_j^k,$$

where $a_{kj} = \langle f, Y_j^k \rangle$. The canonical choice when $n = 2$ is

$$Y_1^k(e^{i\theta}) = \frac{\cos(k\theta)}{\sqrt{\pi}} \quad \text{and} \quad Y_2^k(e^{i\theta}) = \frac{\sin(k\theta)}{\sqrt{\pi}},$$

which connects the theory of spherical harmonics on S^1 back to the theory of Fourier series.

We can utilize the theory of surface harmonics to obtain a decomposition of $L^2(\mathbb{R}^n)$ which plays nicely with respect to the rotation group. To see this, define $\mathfrak{H}_k$ to be the space of all linear combinations of functions of the form $a \cdot Y^k$, where

143

Y^k is a solid spherical harmonic of degree k, and a is a radial function on $\mathbb{R}^n$, such that $a \cdot Y^k$ is square integrable, which holds in particular if and only if

$$\left(\int_0^\infty |a(t)|^2 t^{k+d-1} \, dt \right)^{1/2} < \infty,$$

a fact immediately verified by applying polar coordinates and homogeneity.

Theorem 10.4. *The spaces $\{\mathfrak{H}^k\}$ are closed in $L^2(\mathbb{R}^n)$, pairwise orthogonal, and give a decomposition*

$$L^2(\mathbb{R}^n) = \bigoplus_k \mathfrak{H}^k.$$

Proof. Choose a family of solid spherical harmonics $\{Y_l^k\}$ for each k which form an orthonormal basis when restricted to surface harmonics on S^{n-1}. Any element $f \in \mathfrak{H}^k$ can be written as

$$f = \sum_l a_l Y_l^k,$$

for some radial functions $\{a_l\}$, and applying orthogonality and integration in polar coordinates gives that

$$\|f\|_{L^2(\mathbb{R}^n)} = \left(\sum_l \|a_l Y_l^k\|_{L^2(\mathbb{R}^n)}^2 \right)^{1/2}$$

$$= \left(\sum_l \int_0^\infty |a_l(t)|^2 t^{k+d-1} \, dx \right)^{1/2}$$

Thus the L^2 norm of f is an l^2 sum of a weighted L^2 norm of the radial functions $\{a_l\}$. This immediately gives that the space $\mathfrak{H}^k$ is closed.

It is immediate the spaces $\mathfrak{H}^k$ and $\mathfrak{H}^l$ are parwise orthogonal, since they are orthogonal when integrated on any sphere, and we can then integrate in polar coordinates. To show they decompose $L^2(\mathbb{R}^n)$ completely, fix $f \in L^2(\mathbb{R}^n)$ and suppose f is orthogonal to any element of $\mathcal{H}^k$, for any k. For almost every $r > 0$, the function $f_r(x) = f(rx)$ lies in $L^2(S^{n-1})$. But then we can write $f_r = \sum a_{kl}(r) Y_l^k$ for some coefficients $\{a_{kl}(r)\}$ with the property that

$$\|f\|_{L^2(\mathbb{R}^n)} = \left(\sum_{k,l} \int_0^\infty r^{d-1} |a_{kl}(r)|^2 \, dr \right)^{1/2}$$

144

For any radial function b such that bY_l^k is square integrable, we have that

$$\langle f, bY_l^k \rangle = \int_0^\infty r^{d-1+k} a_{kl}(r)\overline{b(r)}\, dr.$$

This can only be possible for all b if $a_{kl} = 0$, and since this holds for all k and l, we conclude that f vanishes almost everywhere on almost every sphere about the origin. This can only be possible if f is equal to zero almost everywhere. Thus we conclude that $\mathfrak{H}$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Using the theory of zonal harmonics introduced in the next section, we will be able to see that the spaces $\mathcal{H}^k$ and $\mathfrak{H}^k$ are *minimal* invariant subspaces of the action of $SO(n)$ on $L^2(\mathbb{R}^n)$. Indeed, the rotation group is transitive on the family of zonal harmonics of a fixed degree k, and these harmonics span $\mathcal{H}^k$, which shows $\mathcal{H}^k$ is a cyclic subspace of the action of $SO(n)$. We will also see later on in these notes that the spaces $\mathfrak{H}^k$ are invariant under the Fourier transform on $\mathbb{R}^n$.

10.1 Zonal Harmonics

For each $x \in S^n$, we have a linear functional L_x on $\mathcal{H}_k$ given by $L_x(f) = f(x)$. Applying the duality theory of inner product spaces, we can find a spherical harmonic Z_x^k in $\mathcal{H}^k$ such that such that for any $Y^k \in \mathcal{H}_k$,

$$Y^k(x) = \int_{S^n} Z^k(x, x')Y^k(x')\, dx'.$$

The function Z_x^k is called the *zonal harmonic of degree k with pole x.*

Theorem 10.5. *Let Z_x^k denote the zonal harmonic of degree k with pole x. Then:*

- *The function Z_x^k is real-valued.*

- *For any $R \in SO(n)$,*
$$Z_{Rx}^k = (R^{-1})^* Z_x^k.$$

- *For any orthonormal basis $\{Y_1^k, \ldots, Y_l^k\}$ of $\mathcal{H}_k$,*

$$Z_x^k(x') = \sum_j \overline{Y_j^k(x)} Y_j^k(x').$$

We thus have

$$Z_x^k(x) = \frac{\dim \mathcal{H}_k}{|S^n|},$$

and moreover, this is the maximum *value of Z_x^k on S^n.*

Proof. We have

$$Z_x^k = \sum \langle Z_x^k, Y_j^k \rangle Y_j^k,$$

but

$$\langle Z_x^k, Y_j^k \rangle = \overline{Y_j^k}(x),$$

and so

$$Z_x^k(x') = \sum \overline{Y_j^k}(x) Y_j^k(x').$$

We can choose $\{Y_1^k, \dots, Y_l^k\}$ to be real-valued, and from this we therefore conclude Z_x^k is real valued. We also calculate that for $R \in SO(n)$ and $Y^k \in \mathcal{H}_k$, $R^* Y^k \in \mathcal{H}_k$, and thus

$$\int_{|x|=1} Z_x^k(R^{-1}x') Y^k(x') \, dx = \int_{|x|=1} Z_x^k(x') Y^k(Rx') \, dx' = Y^k(Rx),$$

But this means that $\langle Z_{Rx}^k - (R^{-1})^* Z_x^k, Y^k \rangle = 0$ for all $Y^k \in \mathcal{H}_k$, which can only be possible if $Z_{Rx}^k = (R^{-1})^* Z_x^k$.

Using the results we have just proved, the quantity $Z_x^k(x)$ is independent of x. Moreover,

$$Z_x^k(x) = \sum_j |Y_j^k(x)|^2.$$

Integrating the right hand side over S^n, we thus obtain that

$$|S^n| Z_x^k(x) = \dim \mathcal{H}_k.$$

But

$$Z_x^k(x') = \langle Z_{x'}^k, Z_x^k \rangle = \int Z_{x'}^k(x'') Z_x^k(x'') \, dx''.$$

Now

$$\|Z_x^k\|_{L^2(S^n)}^2 = \sum |Y_j^k(x)|^2 = \frac{\dim \mathcal{H}_k}{|S^n|},$$

and thus by Cauchy-Schwartz we have

$$|Z_x^k(x')| \leqslant \frac{\dim \mathcal{H}_k}{|S^n|}. \qquad \square$$

146

Any function orthogonal to the family of functions $\{Z_x^k : x \in S^n\}$ must vanish on S^n. It follows that the family $\{Z_x^k\}$ spans $\mathcal{H}^k$.

We can immediately use the zonal harmonics to obtain an integral formula for the projections $P_k : L^2(S^n) \to \mathcal{H}^k$. Indeed, if $f \in L^2(S^n)$, and we can write $f = \sum_k f_k$, where $f_k \in \mathcal{H}^k$, then

$$f_k(x) = \int Z_x^k(x')f(x')\,dx'.$$

Thus in general, we have

$$P_k f(x) = \int Z_x^k(x')f(x')\,dx'.$$

Thus $(x, x') \mapsto Z_x^k(x')$ is the required integral kernel.

Many operators with rotational symmetry can be written in terms of zonal harmonics. For instance, let us consider the solution operator L to the Dirichlet problem on the unit ball in $\mathbb{R}^n$, i.e. such that if b is an appropriately regular function on the boundary of the unit ball, then Lb gives a harmonic function on the interior of the unit ball, which has b as it's boundary values. The operator L must be radially symmetric, i.e. for any rotation R, we must have

$$R^*(Lb) = L(R^*b),$$

We might therefore expect to write L in terms of zonal harmonics.

Indeed, if $Y^k \in \mathcal{H}^k$ is a spherical harmonic of degree k, then clearly for $0 < r < 1$ and $|x| = 1$ we must have

$$L\{Y^k\}(rx) = r^k Y^k(x) = r^k \langle Z_x^k, Y^k \rangle = r^k \int_{S^{n-1}} Z_x^k(x')Y^k(x')\,dx'.$$

In fact, if we set

$$K(rx, x') = \sum_{k=0}^{\infty} r^k Z_x^k(x'),$$

which is a well defined, smooth function on $B \times S^{n-1}$, then we conclude that for any polynomial f restricted to the unit sphere,

$$Lf(x) = \int_{S^{n-1}} K(x, x')b(x')\,dx'.$$

147

A density argument shows that K must be the *Poisson kernel* for the unit ball, which in particular, gives the identity

$$\frac{1}{|S^{n-1}|} \frac{1 - |rx|^2}{|rx - x'|^n} = \sum_{k=0}^{\infty} r^k Z_x^k(x').$$

We notice that the spaces $\mathcal{H}^k$ are all eigenspaces of the operators L_r on the unit sphere given by $L_r b(x) = Lb(rx)$. This is a property of all appropriately regular operators which are rotationally symmetric.

Let us now return to the study of the zonal harmonics. The following Lemma will be very useful in this regard.

Lemma 10.6. *Consider $\sigma_1, \tau_1, \sigma_2, \tau_2 \in S^n$. If $\sigma_1 \cdot \tau_1 = \sigma_2 \cdot \tau_2$, then there exists a rotation matrix $R \in SO(n+1)$ such that $R\sigma_1 = \sigma_2$ and $R\tau_1 = \tau_2$.*

As a result of this fact, and the rotation invariants of the zonal harmonics, we conclude that $Z_x^k(x')$ is a function of $x \cdot x'$. We will later see that we can write $Z_x^k(x')$ as a *polynomial of degree k* in x and x'.

Lemma 10.7. *Let P be a polynomial in n variables such that $P \circ R = P$ for any rotation matrix $R \in SO(n)$. Then there exists constants $\{c_j\}$ such that*

$$P(x) = \sum_j c_j |x|^{2j}.$$

Proof. Write $P = \sum P_j$, where P_j is homogeneous of degree j. Then each of the polynomials $\{P_j\}$ satisfies the assumptions of the lemma. But then $P_j(x)/|x|^j$ is invariant under rotations, and is homogeneous of degree zero, which can only occur if $P_j(x)/|x|^j$ is constant. Since $|x|^j$ is not a polynomial if j is odd, this yields the result above. $\square$

Remark. One consequence of this Lemma is that any constant coefficient differential operator on $\mathbb{R}^n$ invariant under rotations can be written as

$$P(\Delta),$$

a polynomial in the Laplacian Δ.

Thus a polynomial restricted to the sphere which is invariant under rotations is constant. If we fix a pole x, and consider harmonic polynomials restricted to the sphere which are invariant under rotations fixing x (or equivalently, harmonic polynomials that are constant on *parallels* to x, i.e. the intersections of S^n with affine hyperplanes orthogonal to the line from the origin to x), then we introduce the zonal harmonics.

Theorem 10.8. *Fix $x_0 \in S^n$. Then any spherical harmonic Y^k of degree k, which is invariant under rotations fixing x_0, can be written as a constant multiple of the zonal harmonic $Z_{x_0}^k$. Moreover, there exists a polynomial $G_{n,k}$ of degree k, such that*

$$Y^k(x) = Y^k(x_0)G_{n,k}(x \cdot \tau).$$

Proof. TODO: See notes by Jean Gallier, Theorem 1.17. □

The polynomial $G_{n,k}$ is called the *Gegenbauer polynomial*, or *ultraspherical polynomial*, of degree k. The proof in Jean Gallier's notes shows that if $k = 2m$ is even, then

$$G_{n,k} = \sum_{j=0}^{m} c_j t^{2j}(1 - t^2)^{m-j},$$

and if $k = 2m + 1$ is odd, then

$$G_{n,k} = \sum_{j=0}^{m} c_j t^{2j+1}(1 - t^2)^{m-j}.$$

Thus in particular, $G_{n,k}(-t) = (-t)^k G_{n,k}(t)$. If $n = 2$, then $\{G_{n,k}\}$ is precisely the family of *Legendre polynomials*.

10.2 The Funk-Hecke Formula

The Funk-Hecke formula allows one to define a 'zonal convolution' operator on the sphere. Given a measurable function $K : [-1, +1] \to \mathbb{C}$ such that

$$\int_{-1}^{1} |K(t)|(1 - t^2)^{n/2-1} \, dt < \infty,$$

and a function f on S^n, we can define a 'convolution' $K * f$ by setting

$$(K * f)(x) = \int_{S^n} K(x \cdot y)f(y) \, dy.$$

In other words, to calculate $(K * f)(x)$, we average f along parallel circles to x, and then integrate these averages along K. The following result was first shown by Funk in 1916, and by Hecke in 1918.

$$\int_{S^n} |K(x \cdot y)|^p = \int_{-1}^{1} (1 - r^2)^{d/2-1}|K(t)|^p$$

Theorem 10.9. *If K is as above, then there exists a sequence $\{\lambda_k\}$ such that if Y^k is a spherical harmonic of degree k, then*

$$K * f = \lambda_k f.$$

Thus the spaces $\mathcal{H}^k$ are eigenspaces for the zonal convolution operator.

Proof. Since K is as above, the theory of orthogonal polynomials allows us to write

$$K = \sum \lambda_k G_k.$$

But this means that

$$TZ_x^k = \lambda_k Z_x^k,$$

and since $\{Z_x^k\}$ span $\mathcal{H}^k$, it follows that $K * f = \lambda_k f$ for all $f \in \mathcal{H}^k$. TODO: Go over this proof in more detail. $\square$

Essentially all rotation invariant operators arise as zonal convolutions.

Theorem 10.10. *Let $T : \mathcal{D}(S^{n-1}) \to \mathcal{D}^*(S^{n-1})$ be a Schwartz operator such that for any $R \in SO(n)$,*

$$T \circ R^* = R^* \circ T.$$

Then T is given by a zonal convolution operator.

Proof. The operator T is Schwartz, so we can find a Schwartz kernel K, such that for $f \in C^\infty(S^{n-1})$,

$$Tf(x) = \int K(x,y)f(y)\, dy.$$

Then

$$R^*\{Tf\}(x) = Tf(Rx) = \int K(Rx,y)f(y)\, dy$$

and

$$T\{R^*f\}(x) = \int K(x,y)f(Ry)\, dy = \int K(x, R^{-1}y)f(y)\, dy.$$

This can only possible hold for all f if $K(Rx, Ry) = K(x,y)$ for all $R \in SO(n)$. But this holds if and only if we can find a distribution L on $[-1, +1]$ such that $K(x,y) = L(x \cdot y)$, for some distribution L. And this means, in particular, that T is given by zonal convolution. $\square$

10.3 The Fourier Transform of $\mathfrak{H}^k$

The Fourier transform of an even function on $\mathbb{R}$ is an even function, and for an even function f, we can write

$$\hat{f}(\xi) = 2 \int_0^\infty \cos(2\pi\xi x) f(x) \, dx.$$

Similarily, if f is an odd function, then $\hat{f}$ is odd, and

$$\hat{f}(\xi) = 2i \int_0^\infty \sin(2\pi\xi x) f(x) \, dx.$$

Thus the Fourier transform preserves the decomposition $L^2(\mathbb{R}) = \mathfrak{H}^0 \oplus \mathfrak{H}^1$.

We can obtain a similar formula in $\mathbb{R}^2$, and here, the theory of Fourier series emerges. Indeed, here $SO(2)$ is isomorphic to $\mathbb{T}$, since we can parameterize the rotation group by letting for each $\theta \in \mathbb{R}$, the rotation R_θ denote the rotation $R_\theta(z) = e^{i\theta}z$. The space $\mathcal{H}^k$ is two dimensional, spanned by $e^{ki\theta}$ and $e^{-ki\theta}$, since these functions are restrictions of the harmonic functions z^k and $\bar{z}^k$. Let us denote the span of the first function by $\mathcal{H}^k_+$, and the latter by $\mathcal{H}^k_-$, these spaces being orthogonal to one another. By multiplying these functions by radial functions, we can then see that the resulting spaces $\mathfrak{H}^k_+$ and $\mathfrak{H}^k_-$ are preserved by the Fourier transform, and moreover, compute a simple formula for their Fourier transform. Indeed, we can write

$$\mathfrak{H}^k_+ = \{f \in L^2(\mathbb{R}^2) : R_\theta f = e^{ki\theta} f \text{ for all } R_\theta \in SO(2)\}$$

and

$$\mathfrak{H}^k_- = \{f \in L^2(\mathbb{R}^2) : R_\theta f = e^{-ki\theta} f \text{ for all } R_\theta \in SO(2)\}.$$

But these properties immediately imply that $\mathfrak{H}^k_+$ and $\mathfrak{H}^k_-$ are preserved under the Fourier transform, since rotations commute with the Fourier transform, and so, for example, if $f \in \mathfrak{H}^k_+$,

$$R_\theta \hat{f} = \widehat{R_\theta f} = e^{ki\theta} \hat{f},$$

which means $\hat{f} \in \mathfrak{H}^k_+$. We have a formula for this Fourier transform; indeed, if $f(re^{i\theta}) = a(r)e^{ki\theta}$, for some $k \in \mathbb{Z}$, then we can write $\hat{f}(\rho e^{i\theta}) = b(\rho)e^{ki\theta}$, and we

have

$$b(\rho) = \widehat{f}(\rho e^{i0})$$

$$= \int_0^\infty ra(r) \int_0^{2\pi} e^{ki\theta - 2\pi i\rho\xi \cos(\theta)}$$

$$= (-i)^k 2\pi \int_0^\infty ra(r) J_k(2\pi\rho r) \, dr.$$

Thus we can find a formula for the Fourier transform of the elements of $\mathfrak{H}^k$.

On $\mathbb{R}^n$, it is also simple to find an expression for the Fourier transform of a radial function. If $f(x) = a(|x|)$, then we calculate that, if $\widehat{f}(\xi) = b(|\xi|)$, and a simple calculation also shows that

$$b(\rho) = 2\pi\rho^{1-n/2} \int_0^\infty r^{n/2} a(r) J_{n/2-1}(2\pi\rho r) \, dr.$$

Thus we also see the appearance of Bessel functions. We now show we can find such an expression for general spherical harmonics.

Theorem 10.11. *Suppose $f(x) = a(|x|)Y^k(x)$, where Y^k is a solid spherical harmonic of degree k, and a is a radial function chosen such that $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. Then $\widehat{f}(\xi) = b(|\xi|)Y^k(x)$, where*

$$b(\rho) = 2\pi(-i)^k r^{-n/2-k+1} \int_0^\infty r^{n/2+k} a(r) J_{n/2+k-1}(2\pi\rho r) \, dr.$$

Proof. TODO: See Stein and Weiss, Chapter 4, Theorem 3.10. □

Chapter 11

Orthogonal Polynomials

There are many useful series representations of functions, in which the exponentials that occur in the Fourier expansion are replaced with families of polynomials. The study of such expansions is called the theory of *orthogonal polynomials*.

11.1 Basic Setup

Let $\mathcal{P}$ denote the space of all real-valued polynomials in a single variable. We fix a non-negative weight function $w : \mathbb{R} \to \mathbb{R}$, such that for any polynomial f, the product fw lies in $L^1(\mathbb{R})$. We then define an inner product on $\mathcal{P}$ by setting

$$\langle f, g \rangle_w = \int f(x)g(x)w(x) \, dx.$$

By taking the basis $\{1, x, x^2, \dots\}$ for $\mathcal{P}$, and applying the Gram-Schmidt process, we find an orthonormal basis $\{\phi_n\}$ for $\mathcal{P}$, where $\deg(\phi_n) = n$ for all n. This type of basis for $\mathcal{P}$ is unique up to rescaling the individual elements ϕ_n by non-zero multiplicative factors. This is precisely the family of *orthogonal polynomials* we wish to study in this chapter:

- If we take $w(x) = (1-x^2)\mathbb{I}(|x| \leqslant 1)$, then the functions $\{\phi_n\}$ are normalized *Legendre Polynomials*.

  More generally, if $w(x) = (1 - x^2)^m \mathbb{I}(|x| \leqslant 1)$, then $\{\phi_n\}$ are normalized associated Legendre polynomials. These polynomials arise, for instance, in the analysis of spherical harmonics.

153

- If one chooses $w(x) = \mathbb{I}[-1, +1](1 - t^2)^{n/2-1}$, then one obtains the *Gegenbauer polynomials* on S^n.

- If we define $w(x) = (1 - x^2)^{-1/2}\mathbb{I}(|x| \leqslant 1)$, then the functions $\{\phi_n\}$ are normalized *Chebyshev polynomials of the first kind*, and if $w(x) = (1 - x^2)^{1/2}\mathbb{I}(|x| \leqslant 1)$, *Chebyshev polynomials of the second kind*. These polynomials often arise in certain 'extremal' contexts.

- If we define $w(x) = e^{-x^2/2}$, then the functions $\{\phi_n\}$ are obtained by L^2 normalizing the *Probabilists Hermite Polynomials He_n*, and if $w(x) = e^{-x^2}$, the functions $\{\phi_n\}$ are obtained from L^2 normalizations of the *Physicist's Hermite Polynomials H_n*. These polynomials arise in probability and quantum physics.

Since $\phi_1, \ldots, \phi_{n-1}$ must necessarily span the space of all polynomials of degree $n - 1$, we conclude that if $f \in \mathcal{P}$ and $n > \deg(f)$, then $\langle f, \phi_n \rangle_w = 0$.

We can write

$$x\phi_n = \sum_{k=0}^{n+1} c_{nk}\phi_k,$$

where $c_{nk} = \langle x\phi_n, \phi_k \rangle_w$. Since

$$\langle x\phi_n, \phi_k \rangle_w = \langle \phi_n, x\phi_k \rangle_w,$$

the last point implies that $c_{nk} = 0$ unless $k = n - 1$, $k = n$, or $k = n + 1$. The next Lemma uses this property to give a recurrence relation for the sequence $\{\phi_n\}$.

Lemma 11.1. *For any set of orthogonal polynomials $\{\phi_n\}$, let $\{a_n\}$ be the x^n coefficients of $\{\phi_n\}$, and $\{b_n\}$ the x^{n-1} coefficients of the $\{\phi_n\}$. Then if*

$$A_n = a_{n+1}/a_n, \quad B_n = A_n \left(\frac{b_{n+1}}{a_{n+1}} - \frac{b_n}{a_n} \right), \quad \text{and} \quad C_n = \frac{A_n}{A_{n-1}} \frac{\|\phi_n\|^2}{\|\phi_{n-1}\|^2},$$

then we conclude that

$$\phi_{n+1} = (A_n x + B_n)\phi_n - C_n\phi_{n-1}.$$

Thus

$$(y - x)\phi_n(x)\phi_n(y) = A_n^{-1}[\phi_n(x)\phi_{n+1}(y) - \phi_{n+1}(x)\phi_n(y)]$$
$$- A_n[\phi_{n-1}(x)\phi_n(y) - \phi_n(x)\phi_{n-1}(y)].$$

But this means that

$$(y - x) \sum_{n=0}^{N} \phi_n(x)\phi_n(y) = A_n^{-1} [\phi_n(x)\phi_{n+1}(y) - \phi_{n+1}(x)\phi_n(y)].$$

If we consider the partial summation operator $S_N f = \sum_{n=0}^{N} \langle f, \phi_n \rangle \phi_n$, then the kernel K_N of this operator is symmetric, of the form

$$K_N(x, y) = \sum_{n=0}^{N} \phi_n(x)\phi_n(y),$$

the formula above gives us the *Christoffel-Darboux formula*

$$K_N(x, y) = A_N^{-1} \frac{\phi_N(x)\phi_{N+1}(y) - \phi_{N+1}(x)\phi_N(y)}{y - x},$$

which might be compared to the closed forms for the partial summation operators for Fourier series on the line.

If w is an even weight function, then $\langle \phi_n(-x), p \rangle = 0$ for $\deg(p) < n$. This implies $\phi_n(-x) = c_n\phi_n(x)$ for some constant c_n. Since the leading coefficient of $\phi_n(-x)$ is $(-1)^n$ times the leading coefficient of $\phi_n(x)$, we conclude that $c_n = (-1)^n$. Thus $\phi_n(-x) = (-1)^n\phi_n(x)$. In particular, ϕ_n *only contains even powers of x is n is even, and only odd powers of x if n is odd.*

By applying Lagrange interpolation, one can show that, if w is supported on an interval (a, b), then all the zeroes of the polynomials $\{\phi_n\}$ are real, and contained in the interval (a, b). Indeed, if ϕ_n does not change sign at least n times on the interval (a, b), then we can find a polynomial p of degree less than n such that $\text{sgn}(p) = \text{sgn}(\phi_n)$ on (a, b), and then the equation $\langle \phi_n, p \rangle_w = 0$ gives a contradiction.

11.2 The Pearson Differential Equation

Fix an interval I. We want to consider a weight functions w satisfying the *Pearson differential equation*

$$w' = \frac{D + Ex}{Ax^2 + Bx + C}w,$$

such that $(Ax^2 + Bx + C)w$ vanishes at the finite endpoints of the interval I. These weight function split into three classes, depending on, respectively, whether the interval I is finite, infinite, or half finite:

155

- The *Jacobi polyonials* on an interval $(-1, 1)$, given for $\alpha, \beta > -1$ by the weight function $w(x) = (1 - x)^\alpha (1 + x)^\beta$.

- The *Hermite polynomials* on the real line, given (if you're a probabilistc) by the weight function $w(x) = e^{-x^2/2}$, and (if you're a physicist) by the weight function $w(x) = e^{-x^2}$.

- The *Laguerre* polynomials on the interval $(0, \infty)$, given by the weight function $w(x) = x^\alpha e^{-x}$ for $\alpha > -1$.

Their associated orthogonal polynomials then satisfy an associated differential equation of the form

$$(Ax^2 + Bx + C)\phi_n''(x) + (Ux + V)\phi_n'(x) + W_n\phi_n(x) = 0,$$

for some constants U, V which are independent of n, and a sequence of constants $\{W_n\}$ (TODO: See Jackson, Chapter 7, Section 10). In particular:

- The Jacobi polynomials $\{J_n\}$ satisfy

$$(1 - x^2)J_n'' + [\beta - \alpha - (\alpha + \beta + 2)x]J_n' + n(\alpha + \beta + n + 1)J_n = 0.$$

- The probabilist's Hermite polynomials $\{He_n\}$ satisfy

$$He_n'' - xHe_n' + nHe_n = 0,$$

and the phyisicist's Hermite polynomials $\{H_n\}$ satisfy

$$H_n'' - 2xH_n' + 2nH_n = 0$$

- The Laguerre polynomials $\{L_n\}$ satisfy

$$xL_n'' + (\alpha + 1 - x)L_n' + nL_n = 0.$$

Orthogonal polynomials satisfying the Pearson equation also satisfy a *Rodrigues Formula*. Namely, it can be shown that there exists constants $\{\lambda_n\}$ such that

$$\phi_n(x) = \frac{\lambda_n}{w(x)}\frac{d^n}{dx^n}\left\{(Ax^2 + Bx + C)^n w(x)\right\}.$$

Certainly, because of the Rodriguez formula, the right hand side is a polynomial $\lambda_n q_n$ with degree at most n, and for any polynomial p with degree less than n, integration by parts n times gives that

$$\langle p, q_n \rangle_w = \int p(x) \frac{d^n}{dx^n} [(Ax^2 + Bx + C)^n w(x))] \, dx = 0.$$

Thus q_n is a constant multiple of ϕ_n, and it suffices to show that q_n is non-zero. But this follows because if $q_n = 0$, then the nth derivative of $(Ax^2 + Bx + C)^n w(x)$ is equal to zero on an open interval, so that $(Ax^2 + Bx + C)^n w(x)$ is a polynomial of order n on that interval. This is clearly not true in any of the cases of the Pearson equation.

- We usually define the Jacobi polynomials by the formula

$$P_n^{(\alpha,\beta)} = \frac{(-1)^n}{2^n n!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \partial_x^n \left\{ (1 - x)^{\alpha+n} (1 + x)^{\beta+n} \right\}.$$

  They are not orthonormal, i.e.

$$\| P_n^{(\alpha,\beta)} \|_{L^2(w)}^2 = \frac{1}{n!} \frac{2^{\alpha+\beta+1}}{\alpha + \beta + 2n + 1} \frac{\Gamma(\alpha + n + 1)\Gamma(\beta + n + 1)}{\Gamma(\alpha + \beta + n + 1)}.$$

- We define the Hermite polynomials by the formula

$$H_n = (-1)^n e^{x^2} \partial_x^n \{ e^{-x^2} \}$$

  and

$$He_n = (-1)^n e^{x^2/2} \partial_x^n \{ e^{-x^2/2} \}.$$

  Then $H_n(x) = 2^{n/2} He_n(\sqrt{2}x)$. Moreover,

$$\| H_n \|_{L^2(w)}^2 = 2^n \sqrt{2\pi} n! \quad \text{and} \quad \| He_n \|_{L^2(w)}^2 = \sqrt{2\pi} n!.$$

- The Laguerre polynomials are given by

$$L_n^{(w)} = (-1)^n x^{-\alpha} e^x \partial_x^n \{ x^{\alpha+n} e^{-x} \}.$$

  They have

$$\| L_n^{(\alpha)} \|_{L^2(w)}^2 = n! \Gamma(\alpha + n + 1).$$

  We define the normalized Laguerre polynomials

$$l_n = (n!)^{-1/2} \Gamma(\alpha + n + 1)^{-1/2} L_n.$$

157

11.3   Legendre Polynomials

Consider the function

$$H(x,r) = \frac{1}{(1 - 2xr + r^2)^{1/2}},$$

analytic on the line $r = 0$. If we consider the power series expansion around $r = 0$, we can write

$$H(x,r) = \sum_{j=0}^{\infty} P_j(x) r^j.$$

Thus we have

$$P_j(x) = \frac{1}{j!} \frac{\partial^j H}{\partial r^j}\bigg|_{r=0}.$$

One can prove by induction that there exists polynomials $Q_{j,k}$ with $\deg(Q_{j,k}) \leqslant k$ such that

$$\frac{\partial^j H}{\partial r^j} = \sum_{k=0}^{j} \frac{Q_{j,k}(x,r)}{(1 - 2xr + r^2)^{1/2+k}},$$

since, assuming the formula works for some particular j, the product rule implies

$$\frac{\partial^{j+1} H}{\partial r^j}$$

is also of this form. Setting $r = 0$ shows P_j is a polynomial function with degree at most j. The family of polynomials $\{P_j\}$ are called the *Legendre polynomials*. We can explicitly calculate the first few, i.e. we have

$$P_0(x) = 1 \quad P_1(x) = x \quad P_2(x) = \frac{3x^2 - 1}{2}$$

$$P_3(x) = \frac{5x^3 - 3x}{2} \quad P_4(x) = \frac{35x^4 - 30x^2 + 3}{8}.$$

Historically, the Legendre polynomials were introduced by Adrien-Marie Legendre as the coefficients of the Newton potential in $\mathbb{R}^3$, i.e.

$$\frac{1}{|x - y|} = \sum_{l=0}^{\infty} \frac{|y|^l}{|x|^{l+1}} P_l\left(\frac{x \cdot y}{|x||y|}\right),$$

which is useful when integrating the potential over a continuous medium.

   Let us now derive some of the many interrelated properties of the family of Legendre polynomials.

Theorem 11.2. *If j is even, P_j is an even polynomial, and if j is odd, P_j is an odd polynomial.*

Proof. Since $H(-x, -r) = H(x, r)$, we find that

$$\sum_{j=0}^{\infty}(-1)^j P_j(-x)r^j = \sum_{j=0}^{\infty} P_j(x)r^j,$$

which shows $P_j(-x) = (-1)^j P_j(x)$. □

Since
$$(1 - 2xr + r^2)\partial_r H - (x - r)H = 0,$$

substituting in the Legendre polynomials into this equation gives the recurrence relation
$$(j + 1)P_{j+1}(x) - (2j + 1)xP_j(x) + jP_{j-1}(x) = 0.$$

There are several other differential equations that H satisfies, namely

$$(1 - 2xr + r^2)\partial_r H - rH = 0$$

and
$$r\partial_r(rH) - (1 - rx)\partial_x H = 0.$$

This leads to the recurrence relations

$$jP_j - xP'_j + P'_{j-1} = 0$$

and
$$jP_{j-1} - P'_j + xP'_{j-1} = 0$$

respectively. Adding and transposing terms leads to the recurrence

$$P'_{j+1} - P'_{j-1} = (2j + 1)P_j.$$

and
$$(1 - x^2)P''_j - 2xP'_j + j(j + 1)P_j = 0.$$

Thus P_j is a solution to a homogeneous linear differential equation. We now come to the *orthogonality* of the Legendre polynomials.

159

Theorem 11.3. *If $j \neq k$, then*

$$\int_{-1}^{1} P_j(x)P_k(x) \, dx = 0.$$

More generally, if Q is any *polynomial with* $\deg(Q) < j$, *then*

$$\int_{-1}^{1} P_j(x)Q(x) \, dx = 0.$$

Proof. Multiplying the differential equation above gives that

$$(1 - x^2)P_j''P_k - 2xP_j'P_k + j(j+1)P_jP_k = 0$$

and

$$(1 - x^2)P_jP_k'' - 2xP_jP_k' + k(k+1)P_jP_k = 0.$$

Thus an integration by parts justifies that

$$(j(j+1) - k(k+1)) \int_{-1}^{1} P_jP_k = \int_{-1}^{1} 2x[P_j'P_k - P_jP_k'] - (1 - x^2)[P_j''P_k - P_jP_k'']$$

$$= \int_{-1}^{1} 2xw - (1 - x^2)w'$$

$$= \int_{-1}^{1} 2xw - 2xw = 0.$$

Thus we conclude that P_j and P_k are orthogonal. Since $\{P_j\}$ are orthogonal, they are linearly independent. But this means that $P_0, \ldots, P_{j-1}$ span the family of all polynomials with degree less than j. And this yields the second claim, since any Q as above can be written as a linear combination of $P_0, \ldots, P_{j-1}$. $\qquad\square$

Remark. Note that, since these polynomials are orthogonal, and we have $P_0 = 1$ and $P_1 = x$, one can construct the family $\{P_j\}$ by applying the Gram Schmidt process to the linearly independent set $\{1, x, \ldots\}$.

Given this property, it is natural to calculate the quantities

$$C_j = \left(\int_{-1}^{1} |P_j(x)|^2 \, dx \right)^{1/2},$$

since this is necessary to use $\{P_j\}$ as an orthonormal basis.

Theorem 11.4. *We have $C_j = (j + 1/2)^{-1/2}$.*

Proof. TODO: See Page 51 of Jackson, Orthogonal Polynomials $\qquad\square$

Let us define $p_j = (j + 1/2)^{1/2} P_j$ to be the L^2 normalization of P_j. The Weirstrass approximation theorem shows that this is an orthonormal basis for $L^2[-1, 1]$. For $f \in L^2[-1, 1]$, we can now form it's Legendre expansion

$$f = \sum c_j P_j$$

which converges in $L^2[-1, 1]$, where

$$c_j = (j + 1/2) \int_{-1}^{1} f(x) P_j(x) \, dx.$$

are the *Legendre coefficients* of the function f. Given this integral formula, we can now define the expansion of any $f \in L^1[-1, 1]$, though we now no longer have guarantees on the convergence of this expansion.

Let us also introduce some other useful representations of the Legendre polynomials. First is Rodrigues's formula, which states that

$$P_j(x) = \frac{1}{2^j} \frac{1}{j!} \frac{d^j}{dx^j} (x^2 - 1)^j.$$

One can prove this formula by noting that it is a polynomial of degree j, which satisfies the defining differential equation

$$(1 - x^2) P_j'' - 2x P_j' + j(j + 1) P_j = 0,$$

since P_j is the unique polynomial which satisfies this equation up to a constant multiple (TODO: Page 58 of Jackson).

Another useful formula is the integral representation

$$P_j(x) = \frac{1}{\pi} \int_0^{\pi} [x + (x^2 - 1)^{1/2} \cos \phi]^n \, d\phi.$$

This can be proved by showing this integral formula has the same recurrence relation as that which defines $\{P_j\}$ (TODO: See page 59 of Jackson).

In particular, this tells us that the leading coefficient A_j of P_j satisfies the recurrence

$$A_{j+1} = \frac{2j + 1}{j + 1} A_j.$$

Thus
$$A_j = \frac{1 \cdot 3 \cdot \ldots \cdots (2j-1)}{j!} = \frac{1}{2^{j-1}} \frac{(2j-1)!}{j!(j-1)!}.$$

We thus conclude using Stirling's formula that $A_j = 2^{j+O(\log j)}$, and so the leading coefficient grows exponentially in j. On the other hand, if B_j denotes the constant term of P_{2j} (the constant term is zero for odd index Legendre polynomials), then we find that
$$(j+1)B_j + jB_{j-1} = 0,$$

and $B_0 = 1$. Thus we find that $B_j = (-1)^j/(j+1)$, which is rather small. In fact, overall the Legendre polynomials satisfy rather good bounds.

Theorem 11.5. *We have*
$$|P_j(x)| \lesssim \frac{1}{j^{1/2}(1-x^2)^{1/2}}.$$

In particular, we also have $\|P_j\|_{L^\infty[-1,1]} \lesssim j^{-1/2}$.

To study the convergence of the Legendre series, it is natural to introduce the partial summation operators
$$S_N f = \sum_{j=0}^{N} c_j P_j.$$

We can write
$$S_N f(x) = \int_{-1}^{1} K_N(x,y) f(y) \, dy,$$

where
$$K_N(x,y) = \sum_{j=0}^{N} (j + 1/2) P_j(x) P_j(y).$$

In 1858, Elwin Bruno Christoffel found an expression for this operator in a more simple manner, known as *Christoffel's Identity*.

Theorem 11.6. *We have*
$$K_N(x,y) = \frac{N+1}{2} \frac{P_{N+1}(x)P_N(y) - P_N(x)P_{N+1}(y)}{y-x}.$$

Proof. TODO: See Page 55 of Jackson, Orthogonal Polynomials. $\square$

162

Since $S_N 1 = 1$, we conclude that for all $x \in [-1, 1]$,

$$1 = \int K_N(x, y) \, dy.$$

We thus have

$$|S_N f(x) - f(x)| = \left| \frac{N+1}{2} \int_{-1}^{1} \frac{f(y) - f(x)}{y - x} [P_{N+1}(x) P_N(y) - P_N(x) P_{N+1}(y)] \, dy \right|.$$

Taking in absolute values gives that

$$|S_N f(x) - f(x)| \lesssim \int_{-1}^{1} \frac{f(y) - f(x)}{y - x} \, dy.$$

In particular, if f is differentiable at x, then $S_N f(x)$ converges to $f(x)$ as $N \to \infty$, or more generally, a left and right hand derivative at x.

11.4 Linear Quantum Oscillators

The Schrödinger equation on the line is given by

$$i\hbar \partial_t \Psi = \Delta \Psi - V \Psi$$

for a function V which gives the potential energy at a particular location. Applying separation of variables, if $\Psi(x, t) = u(x) T(t)$, then the equation becomes

$$Vu - \Delta u = \lambda u \quad \text{and} \quad T(t) = e^{-i\hbar \lambda t}.$$

We thus needs to find eigenvalues for the operator $Lu = Vu - \Delta u$. If $V(x) = a^2 x^2$, then this is the potential field of a *linear oscillator*. The operator then becomes $L = a^2 x^2 - \Delta$. The operator L is then diagonalized by the Hermite polynomials. Indeed, since $H_n'' - 2x H_n' + 2n H_n = 0$, if we define $u_n(x) = e^{-ax^2/2} H_n(a^{1/2} x)$, then we calculate that

$$Lu_n = a(2n + 1) u_n.$$

The functions $\{u_n\}$ then give an orthogonal basis for $L^2(\mathbb{R})$, diagonalizing L.

163

In the study of the heat equation $\partial_t u = -Lu$, one is lead to consider the heat propogators e^{-tL}. Since

$$e^{-tL} f = (a/2\pi)^{1/2} \sum_{n=0}^{\infty} \frac{1}{2^n n!} \langle e^{-tL} f, u_n \rangle u_n$$

$$= (a/2\pi)^{1/2} \sum_{n=0}^{\infty} \frac{1}{2^n n!} e^{-(2n+1)at} \langle f, u_n \rangle u_n$$

$$= (a/2\pi)^{1/2} \int \left(\sum_{n=0}^{\infty} \frac{1}{2^n n!} e^{-(2n+1)at} e^{-a\left(\frac{x^2+y^2}{2}\right)} H_n(a^{1/2} x) H_n(a^{1/2} y) \right) f(y) \, dy,$$

we can write the kernel of e^{-tL} formally as

$$e^{-tL}(x, y) = (a/2\pi)^{1/2} \sum_{n=0}^{\infty} \frac{1}{2^n n!} e^{-(2n+1)at} e^{-a\left(\frac{x^2+y^2}{2}\right)} H_n(a^{1/2} x) H_n(a^{1/2} y)$$

To simplify this sum, we apply *Mehler's formula*, which says that for all $w \in \mathbb{C}$,

$$\sum_{n=0}^{\infty} \frac{H_n(x) H_n(y)}{n!} (w/2)^n = (1 - w^2)^{-1/2} \exp\left(\frac{2xyw - (x^2 + y^2)w^2}{1 - w^2} \right).$$

(TODO: See Hörmander, Riemannian Manifolds, Section 6.9). Setting $w = e^{-2at}$ gives that

$$e^{-tL}(x, y) = (a/2\pi)^{1/2} (e^{2at} - e^{-2at})^{-1/2} e^{-a\left(\frac{x^2+y^2}{2}\right)} \exp\left(2a \left(\frac{xy e^{-2at} - \left(\frac{x^2+y^2}{2}\right) e^{-4at}}{1 - e^{-4at}} \right) \right).$$

Using hyperbolic trigonometric functions, we can rewrite this formula as

$$e^{-tL}(x, y) = \sqrt{\frac{a}{4\pi \sinh(2at)}} \exp\left(a \left(\frac{xy - \left(\frac{x^2+y^2}{2}\right) \cosh(2at)}{\sinh(2at)} \right) \right).$$

Notice that as $t \to \infty$, $\operatorname{csch}(2at)$ decays exponentially fast, and $\coth(2at)$ converges to 1 exponentially fast. Thus for $t \gg 1$ we have

$$e^{-tL}(x, y) \approx \sqrt{\frac{a}{2\pi \sinh(2at)}} \exp\left(-a \left(\frac{x^2 + y^2}{2} \right) \right)$$

164

Thus a solution to the heat equation with initial conditions given by a Dirac delta function at a point y has total mass $O(e^{-a(y^2/2+2t)})$ as $t \to \infty$.

How do the half-wave propogators behave? Using Mehler's formula again, we can write

$$e^{2\pi itL}(x,y) = \sqrt{\frac{ia}{4\pi \sin(4\pi at)}} \exp\left(ia\left(\frac{xy - \left(\frac{x^2+y^2}{2}\right)\cos(4\pi at)}{\sin(4\pi at)}\right)\right).$$

Notice this equation is periodic, of period $1/2a$, which reflects the periodicity of the harmonic oscillator.

Given an arbitrary, positive definite quadratic form on $\mathbb{R}^d$, the above calculations tensorize to give eigenvalues for the operator $L = Q(x) - \Delta$ on $\mathbb{R}^d$. Indeed, we can find orthonormal eigenvectors $\{v_1, \ldots, v_d\}$ which diagonalize Q, i.e such that $Q(\sum t_i v_i) = \sum a_i^2 t_i^2$. If we consider functions

$$\tilde{H}_n\left(\sum t_i v_i\right) = \prod_j e^{-a_j t_j^2/2} H_{n_j}\left(\sqrt{2a_i} t_i\right)$$

then $L\tilde{H}_n = \left(\sum_j (2n_j + 1)a_j\right)\tilde{H}_n$. Mehler's formula similarily generalizes. We can therefore show that if $Q_{A,t}(x) = \langle A^{1/2}\coth(2A^{1/2}t)x, x\rangle$, then

$$e^{-tL}(x,y) = \det\left(\frac{A^{1/2}}{2\pi \sinh(2A^{1/2}t)}\right)$$
$$\exp\left(-\left(\frac{Q_{A,t}(x) + Q_{A,t}(y)}{2} - \langle A^{1/2}\mathrm{csch}(2A^{1/2}t)x, y\rangle\right)\right).$$

11.5   Bessel Functions

s

Chapter 12

Tauberian Theorems

In many situations in Fourier analysis, one studies a sum $\sum a_n$, to which we associate, via some nonstandard summation method, like that of Cesaro or Abel, a finite quantity a. One often would like conditions that ensure the sum $\sum a_n$ actually converges to this quantity in the classical sense. The first result of this kind was due to Austrian mathematician Alfred Tauber, in 1897, who showed that an Abel summable series with $a_n = o(n^{-1})$ was classically convergent. Hardy and Littlewood showed that it was actually sufficient to assume that $a_n = O(n^{-1})$, and began a series of generalizations, beginning the study of the general theory of these problems, called *Tauberian Theory*. We begin with Tauber's basic result, which follows from a simple calculation.

Theorem 12.1. *If the formal series*

$$\sum_{n=0}^{\infty} a_n$$

is Abel summable, and $a_n = o(n^{-1})$, *then* $\sum a_n$ *converges.*

Proof. Set

$$f(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Set $S_N = \sum_{n=0}^{N} a_n$. Then

$$|S_N - f(x)| = \left| \sum_{n=1}^{N} a_n(1 - x^n) - \sum_{n=N+1}^{\infty} a_n x^n \right|$$

$$\leqslant \sum_{n=1}^{N} n(1-x)|a_n| + \frac{1}{N} \sum_{n=N+1}^{\infty} |na_n||x|^n$$

$$\leqslant (1-x) \sum_{n=1}^{N} n|a_n| + \frac{1}{N(1-x)} \sup_{n>N} |na_n|.$$

Thus we conclude that

$$|S_N - f(1 - 1/N)| \leqslant N^{-1} \sum_{n=1}^{N} n|a_n| + \sup_{n>N} |na_n|.$$

The second term on the right converges to zero as $N \to \infty$. The first also converges to zero, as an average of a sequence converging to zero. But this means that $S_N - f(1 - 1/N)$ converges to zero, and thus S_N converges to the Abel mean of the sequence. $\qquad\square$

Remark. A similar identity shows that if $a_n = O(n^{-1})$, and if the function $f(x) = \sum_{n=1}^{\infty} a_n x^n$ is bounded on $[0, 1)$, then the partial sums S_N are uniformly bounded.

Hardy noticed that this a fortiori implied that a sequence which is *Cesaro summable* and $o(n^{-1})$ is summable. In 1910, he published a Tauberian result for Cesaro summability assuming only the sum was $O(n^{-1})$. That same year, Landau extended the proof to a 'one-sided bound'.

Theorem 12.2. *Let $\{a_n\}$ be a sequence. If the sequence is Cesaro summable, and there exists $C > 0$ such that $a_n \geqslant -C/n$ for all $n > 0$, then $\sum a_n$ converges.*

Proof. We consider a proof method due to Kloosterman, published in 1940. Assume without loss of generality that the sequence is real, and is Cesaro summable to zero. We note that

$$(n + m)\sigma_{n+m} = s_1 + \cdots + s_{n+m}$$
$$= n\sigma_n + ms_n + (ma_{n+1} + (m-1)a_{n+2} + \cdots + a_{n+m}).$$

We have

$$\frac{m(m+1)}{2} \min_{k \in (n,n+m]} a_k \leqslant (ma_{n+1} + (m-1)a_{n+2} + \cdots + a_{n+m}) \leqslant \frac{m(m+1)}{2} \max_{k \in (n,n+m]} a_k,$$

167

and so there exists a quantity $\delta_{n,m}$ lying between $\min_{k\in(n,n+m]} a_k$ and $\max_{k\in(n,n+m]} a_k$ for which
$$(n+m)\sigma_{n+m} = n\sigma_n + ms_n + \frac{m(m+1)}{2}\delta_{n,m}.$$

Under the assumption that $\{a_n\}$ is Cesaro summable to zero, we have $\sigma_n \to 0$ as $n \to \infty$. We now use the identity above to extract the convergence of $\{s_n\}$, i.e. since we have
$$s_n = \frac{(n+m)\sigma_{n+m} - n\sigma_n}{m} - \frac{m+1}{2}\delta_{n,m}.$$

Under the assumption that $a_n \geqslant -C/n$, we have $\delta_{n,m} \geqslant -C/n$. For any $\varepsilon > 0$, if n is suitably large, we also have $|\sigma_{n+m}| + |\sigma_n| \leqslant \varepsilon$, and so
$$s_n \leqslant (2n/m + 1)\varepsilon + C\frac{m+1}{2n}.$$

Provided $m < n$, we have
$$s_n \lesssim (n/m)\varepsilon + (m/n).$$

Thus optimizing for m, i.e. by picking $m = n\varepsilon^{1/2} + O(1)$, we conclude that
$$s_n \lesssim \varepsilon^{1/2}.$$

On the other hand, we have
$$\begin{aligned}(n-m)\sigma_{n-m} &= s_1 + \cdots + s_{n-m} \\ &= n\sigma_n - (s_{n-m+1} + \cdots + s_n) \\ &= n\sigma_n - ms_n + ((m-1)a_n + (m-2)a_{n-1} + \cdots + a_{n-m+2}).\end{aligned}$$

We can find $\delta'_{n,m}$ such that
$$(n-m)\sigma_{n-m} = n\sigma_n - ms_n + \frac{(m-1)m}{2}\delta'_{n,m},$$

where $\delta'_{n,m}$ lies between $\min_{k\in(n-m+1,n]} a_k$ and $\max_{k\in(n-m+1,n]} a_k$. Thus we conclude that
$$s_n = \frac{n\sigma_n - (n-m)\sigma_{n-m}}{m} + \frac{(m-1)m}{2}\delta_{n,m}.$$

But for any $\varepsilon > 0$, if n is chosen sufficiently large so that $|\sigma_n| + |\sigma_{n-m}| \leqslant \varepsilon$, we conclude that
$$s_n \geqslant -2(n/m)\varepsilon - C\frac{m-1}{2n}.$$

168

Again, picking $m = n\varepsilon^{1/2} + O(1)$, we conclude that there is a constant C' such that

$$s_n \geqslant -C\varepsilon^{1/2}.$$

Putting the two bounds above together, we conclude $|s_n| \lesssim \varepsilon^{1/2}$, and we can now take $\varepsilon \to 0$ to obtain the required result. $\qquad\square$

Several results on the convergence of Fourier series follow from this argument:

- If f is a continuous, periodic function on the real line, and $|\widehat{f}(n)| \lesssim |n|^{-1}$, then the Fourier series of f converges pointwise everywhere to f. Indeed, under these assumptions, the result above implies Cesaro summation implies normal summation.

- We obtain a new proof of Dini's Theorem: If f has bounded variation, then the Fourier series of f converges everywhere to $(1/2)[f(t-) + f(t+)]$.

We also note a slight weakening of the assumptions. Namely, the result would follow under the weaker assumption that the quantities

$$w(\rho) = \liminf_{n \to \infty} \inf_{n \leqslant m \leqslant \rho n} |s_m - s_n|$$

converge to zero as $\rho \searrow 0$. A sequence $\{s_n\}$ satisfying these assumptions is called a *slowly varying sequence*.

Hardy asked whether an analogous result held for Abel summation. Littlewood showed in 1911 that such a result was true. His original proof, using repeated differentiation, is quite difficult, and over the past century, many different proof techniques have been devised to demonstratet the result.

TODO: Read more of Korevaar's Book.

Part II

Distributional Methods

Chapter 13

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 analytical issues. For instance, the theory allows us to take the formal derivative of a much more general class of functions than those differentiable pointwise by methods of the classical differential calculus. That such a consistant formal definition of the derivative is possible is hinted at by equations in Fourier analysis such as

$$D^i f = \mathcal{F}^{-1} \left\{ M^i \mathcal{F} \{f\} \right\},$$

where $M^i f(x) = (2\pi i x_i) f(x)$, and $D^i f$ is differentiation in the variable x_i. The quantity $D^i f$ can only be classically interpreted if f is pointwise differentiable in the variable x_i, whereas the right hand side, interpreted in terms of Lebesgue integrals, is defined for a much less regular class, namely the family of integrable functions f such that $M^i \widehat{F} \{f\}$ is also integrable. We will find a much larger domain, a class of generalized functions called tempered distributions, such that both sides of the equation have the same domain, and can be correctly interpreted.

If we consider solutions $u \in C^2_{\text{loc}}(\mathbb{R}^d)$ to elliptic partial differential equations like the Laplace equation $\Delta u = 0$, then the regularity theory of such solutions shows the family of such functions is closed under locally uniform convergence, i.e. in the space $C_{\text{loc}}(\mathbb{R}^d)$. This is no longer true for solutions to certain non-elliptic partial differential equations, such as solutions to the wave equation $\partial_t u = \Delta_x u$ in $C^2_{\text{loc}}(\mathbb{R} \times \mathbb{R}^d)$. For instance, the closure of such a class under uniform convergence would have to contain all functions of the form

$$f(x + t) + g(x - t)$$

for any $f, g \in C_{\text{loc}}(\mathbb{R} \times \mathbb{R}^d)$. To fix this problem, in PDE it is often of interest to enlarge the class of solutions to consider *weak solutions*, i.e. functions $u \in L^1_{\text{loc}}(\mathbb{R}^d)$ such that for any $\phi \in C^\infty_c(\mathbb{R} \times \mathbb{R}^d)$,

$$\int u(t, x)(\Delta \phi - \partial_t \phi)(t, x) \, dt \, dx = 0,$$

an equation true for any classical solution, as verified by integration by parts. This class is then clearly closed under uniform convergence, i.e. in $L^\infty_{\text{loc}}(\mathbb{R}^d)$. The step to distributional solutions to partial differential equations from classical solutions is then not far off, since the theory of generalized functions allows us to apply differentiable operators to $u \in L^1_{\text{loc}}$, and we find that u will be a weak solution to the wave equation if and only if $\partial_t u = \Delta_x u$. We will also find that the family of *distributional* solutions to the wave equation is closed under a very general topology. These properties are one reason why distributions are a core tool to the formulation of many problems in partial differential equations and modern harmonic analysis.

13.1 Distributions

The path of modern analysis has extended analysis from the study of continuous and differentiable functions to the much larger class of 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, e.g. one cannot differentiate elements of $L^1(\mathbb{R}^d)$ in a classical sense. Nonetheless, in pretty much all function spaces we consider on $\mathbb{R}^d$, the space $C^\infty_c(\mathbb{R}^d)$ forms a dense subclass, and in this subclass we can perform pretty much all possible analytical operations. One approach to studying the general class of measurable functions is to prove results for elements of $C^\infty_c(\mathbb{R}^d)$, in which one can apply all necessary analytical operations, 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 $C^\infty_c(\mathbb{R}^d)$ to a larger family of *generalized functions*.

From the perspective of set theory, a function $f : X \to Y$ is a rule for assigning a value 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 very used to

treating a function as defined 'up to a set of measure zero'. This has the peculiar feature that a function often cannot be interpreted pointwise at any particular point. Distribution theory continues this philosophy, taking the point of view that functions are 'integrands', whose properties are understood by integration against a family of 'test functions'.

As an example of this phenomenon, since the dual space of $L^1(\mathbb{R}^d)$ is $L^\infty(\mathbb{R}^d)$, we can think of elements $f \in L^\infty(\mathbb{R}^d)$ as 'integrands', which can be completely understood by 'testing' f against an element $\phi \in L^1(\mathbb{R}^d)$, i.e. via a study of the quantities

$$\int f(x)\phi(x)\ dx.$$

Similarly, if K is a compact topological space, then the dual space to $C(K)$ is the space $M(K)$ of finite Borel measures on K. Thus we can think of the class of measures $M(K)$ as a family of 'generalized integrands' which can be understood by testing these measures against continuous functions.

Notice that as we shrink the family of test functions, the resulting family of 'generalized functions' becomes larger and larger, and so elements are allowed to behave more erratially. A distribution is a 'generalized function' understood by testing against the very regular family of functions in the space $C_c^\infty(\mathbb{R}^d)$, functions which in this situation are classical denoted as $\mathcal{D}(\mathbb{R}^d)$. We obtain a slightly smaller class by considering generalized functions tested against the larger class of test functions $C^\infty(\mathbb{R}^d)$, this family often being denoted $\mathcal{E}(\mathbb{R}^d)$. An intermediate class of test functions is the space of Schwartz functions $\mathcal{S}(\mathbb{R}^d)$, which is often the most convenient for problems studied in harmonic analysis.

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 families of test functions which cannot be appropriately localized, 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*, not discussed here.

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. And indeed, physicists had experimented with using distributions in their calculations before a formal definition was introduced by mathematicians.

13.2 Test Functions and Distributions

Fix an open subset Ω of $\mathbb{R}^d$. Let us review the family of test functions we use to give a theory of distributions:

- The main choice is the space $\mathcal{D}(\Omega) = C_c^\infty(\Omega)$ of compactly supported smooth functions, equipped with the topology of compact convergence, e.g. the LF space consisting of the inductive limit of the Fréchet spaces $C_c^\infty(K)$ of smooth functions on $\mathbb{R}^d$, which are compactly supported on a compact set K, equipped with the topology of uniform convergence of the functions are all of their derivatives. Taking the dual of this space yields the most general space of distributions $\mathcal{D}(\Omega)^*$.

- If $\Omega = \mathbb{R}^d$, we can use the space $\mathcal{S}(\Omega)$ of Schwartz functions on Ω, i.e. the Fréchet space consisting of functions f whose derivatives are all rapidly decreasing, by which we mean that the seminorms

$$\|f\|_{\mathcal{S}^{n,m}(\Omega)} = \sup_{x \in \Omega} \sup_{|\alpha| \leqslant m} |D^\alpha f(x)| \langle x \rangle^n$$

are finite for all integers n and m. $\mathcal{S}(\Omega)$ becomes a Fréchet space if we treat these seminorms as defining a topology on this space. The dual of $\mathcal{S}(\Omega)$ is called the space of *tempered distributions* $\mathcal{S}(\Omega)^*$.

- The space $\mathcal{E}(\Omega) = C_{\mathrm{loc}}^\infty(\Omega)$ consists of all smooth functions on Ω, and is equipped with the topology of locally uniform convergence of the function and all of it's derivatives. Using these functions as test functions gives the space $\mathcal{E}(\Omega)^*$ of *compactly supported distributions*.

- We can use the space $\mathcal{D}^n(\Omega) = C_c^n(\Omega)$ of compactly supported n-times continuously differentiable functions, under the topology of compact convergence, e.g. equipped with the LF topology analogous to that of $\mathcal{D}(\Omega)$. This gives the space $\mathcal{D}^n(\Omega)^*$ of *distributions of order n*. Simiarily, we can use the space $\mathcal{E}^n(\Omega) = C_{\mathrm{loc}}^n(\Omega)$, and the dual is the space of *compactly supported distributions of order n*.

Despite being apriori defined as duals of different vector spaces, we will later see we can identify all these spaces as subspaces of the most general space of distributions $\mathcal{D}(\Omega)$, and so we will begin by studying this most general space. As an LF space, $\mathcal{D}(\Omega)$ is *not metrizable*. This is in some sense necessary; it is *impossible* to make $C_c^\infty(\Omega)$ into a Fréchet space, such that each of the subspaces

$$\mathcal{D}(K) = \{f \in C_c^\infty(\Omega) : \text{supp}(f) \subset K\}$$

are closed in $\mathcal{D}(\Omega)$. The interior of each of these subspaces is empty because no translation of these subspaces is absorbing. The Baire category theorem would then imply that a countable union of such subspaces has nonempty interior. But $\mathcal{D}(\Omega)$ is the countable union of such spaces, which would give a contradiction. Nonetheless, as an LF space, $\mathcal{D}(\Omega)$ is equipped with a topology which makes each of the subspaces $\mathcal{D}(K)$ closed, while still having many of the useful properties that Fréchet spaces come with. In particular, it follows from the general theory of LF spaces that $\mathcal{D}(\Omega)$ is a complete, locally convex space. The space $\mathcal{D}(\Omega)$ is also Montel, since it is the inductive limit of Montel spaces.

We recall some useful properties of the inductive limit. A sequence $\{\phi_n\}$ converges in $\mathcal{D}(\Omega)$ to ϕ if $\bigcup_n \text{supp}(\phi_n)$ is precompact, and if $\{\phi_n\}$ and all derivatives of this sequence converge uniformly to ϕ. The space is Bornological, so that every bounded linear map $T : \mathcal{D}(\Omega) \to Y$ is continuous, and so in particular, all sequentially continuous linear maps are continuous. Moreover, if X is first countable, then for any bounded linear map $T : X \to \mathcal{D}(\Omega)$, there exists a compact set $K \subset \Omega$ such that $T(X) \subset K$. The Arzela-Ascoli theorem implies that for any compact set K, each of the spaces $\mathcal{D}(K)$ has the Heine-Borel property, and so $\mathcal{D}(\Omega)$ also has the Heine-Borel property, i.e. $\mathcal{D}(\Omega)$ is a Montel space.

TODO: Move next lemma to tensor products section.

Theorem 13.1. *For any two open sets Ω_1 and Ω_2, $C_{loc}^\infty(\Omega_1 \times \Omega_2)$ is naturally isomorphic to $C_{loc}^\infty(\Omega_2, C_{loc}^\infty(\Omega_1))$.*

Proof. The correspondence is obtained by mapping $f \in C_{loc}^\infty(\Omega_2, C_{loc}^\infty(\Omega_1))$ to $F \in C_{loc}^\infty(\Omega_1 \times \Omega_2)$ given by setting $F(x, y) = f(y)(x)$. The map is clearly one to one and continuous. If F is an arbitrary element of $C_{loc}^\infty(\Omega_1 \times \Omega_2)$, and we define $f[y](x) = F(x, y)$, then the mean-value theorem implies that the quantities

$$\frac{D_x^\alpha F(x, y + he_i) - D_x^\alpha F(x, y)}{h}$$

converge locally uniformly in x to $D_y^i F$, and thus f lies in $C^1(\Omega_2, C^\infty(\Omega_1))$. But iterating this process shows that f is actually C^∞. Thus the correspondence $f \mapsto F$

175

is a bijection. and since both spaces are Fréchet spaces, the open mapping theorem implies that the correspondence is an isomorphism. □

Theorem 13.2. *Finite sums of tensor functions are dense in $\mathcal{D}(\mathbb{R}^d)$.*

Proof. Recall from the theory of multiple Fourier series that if $f \in C^\infty(\mathbb{R}^d)$ is N periodic, in the sense that $f(x + n) = f(x)$ for all $x \in \mathbb{R}^d$ and $n \in (N\mathbb{Z})^d$, then there are coefficients a_m for each $m \in \mathbb{Z}^n$ such that $f = \lim_{M \to \infty} S_M f$, where the convergence is dominated by the seminorms $\|\cdot\|_{C^n(\mathbb{R}^d)}$, for all $n > 0$, and

$$(S_M f)(x) = \sum_{\substack{m \in \mathbb{Z}^d \\ |m| \leqslant M}} a_m e^{\frac{2\pi i m \cdot x}{N}}.$$

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 $\phi \in \mathcal{D}(\mathbb{R}^d)$ 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 $\psi : \mathbb{R} \to \mathbb{R}$ 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. □

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 $\mathcal{D}(K)$ 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 on $\mathcal{D}(\Omega)$.

- The inclusion $\mathcal{D}(\Omega) \to L_c^p(\Omega)$ is continuous. To prove this, it suffices to prove for each compact K, the inclusion $C_c^\infty(K) \to L^p(K)$ is continuous, and this follows because if $f \in C_c^\infty(K)$, then

$$\|f\|_{L^p(K)} \leqslant |K|^{1/p} \|f\|_{L^\infty(K)}.$$

- Multiplication gives continuous operators

$$\mathcal{D}(\Omega) \times \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$$

176

and

$$C_b^\infty(\Omega) \times \mathcal{D}(\Omega) \to \mathcal{D}(\Omega).$$

However, the operator

$$C_{\mathrm{loc}}^\infty(\Omega) \times \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$$

is sequentially continuous, and thus bounded, but *not* continuous.

- The convolution operators

$$\mathcal{D}(\mathbb{R}^d) \times \mathcal{D}(\mathbb{R}^d) \to \mathcal{D}(\mathbb{R}^d)$$

  and

$$L_c^1(\mathbb{R}^d) \times \mathcal{D}(\mathbb{R}^d) \to \mathcal{D}(\mathbb{R}^d)$$

  are continuous.

Thus $\mathcal{D}(\Omega)$ is an ideal place to study many of the natural operations which occur in analysis.

Let us now discuss the idea of a distribution. If $f \in L_{\mathrm{loc}}^1(\Omega)$, then the linear functional $\Lambda[f]$ on $\mathcal{D}(\Omega)$ defined for each $\phi \in \mathcal{D}(\Omega)$ by setting

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

is continuous. Moreover, $\Lambda[f]$ determines f uniquely, and so we can safely identify f with $\Lambda[f]$. We thus speak of 'the distribution' f. 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 abuse notation by denoting the quantity $\Lambda(\phi)$ by

$$\int_\Omega \Lambda(x)\phi(x)\, dx,$$

even if Λ is not given by integration against some locally integrable function. The space $\mathcal{D}^*(\Omega)$ will be called the space of distributions on Ω. The class of distributions induced by elements of $L_{\mathrm{loc}}^1(\Omega)$ is a fundamental class of distributions, but we will soon see that $\mathcal{D}^*(\Omega)$ contains a much more erratic family of distributions. Nonetheless, using duality, we will still be able to extend many analytical operations to this class, such as the derivative operators. Let us now discuss this extension.

177

One huge advantage of the approach above 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 $A : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ is a continuous operator, then we can consider it's adjoint $A^* : \mathcal{D}(\Omega)^* \to \mathcal{D}(\Omega)^*$. If A^* maps $\mathcal{D}(\Omega)$ continuously into itself, then we can define an extension of A to a map from $\mathcal{D}(\Omega)^* \to \mathcal{D}(\Omega)^*$ by defining, for a distribution Λ, a distribution $A\Lambda$ by the formula

$$\langle A\Lambda, \phi \rangle = \langle \Lambda, A^*\phi \rangle.$$

This definition has the property that $A(\Lambda[\phi]) = \Lambda[A\phi]$ for any $\phi \in \mathcal{D}(\Omega)$, so that we have really constructed an extension of A to a definition defined for all distributions.

Let us use this idea to define the derivative of an arbitrary distribution. For $\phi, \psi \in \mathcal{D}(\mathbb{R})$, 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 behaves on distributions by setting $A^*\psi = -\psi'$. Thus, for a distribution Λ on $\mathbb{R}$, we define it's derivative to be the distribution Λ' such that for $\phi \in \mathcal{D}(\mathbb{R})$,

$$\Lambda'(\phi) = \Lambda(A^*\phi) = -\Lambda(\phi').$$

More generally, a similar calculation allows us to consider, for a distribution Λ defined on an open subset Ω of $\mathbb{R}^d$, and a multi-index α, the partial derivative $\partial^\alpha \Lambda$ of a distribution on $\mathbb{R}^d$, by setting, for $\phi \in \mathcal{D}(\Omega)$,

$$\int \partial^\alpha \Lambda(x)\phi(x) \, dx = (-1)^{|\alpha|} \int \Lambda(x)\partial^\alpha \phi(x) \, dx,$$

i.e. a repeated integration by parts. We can even define the gradient $\nabla\Lambda$, provided we are willing to work with *vector valued distributions* without a proper introduction (such a distribution will act on the space $\mathcal{D}(\Omega, \mathbb{R}^n)$ of vector-valued test functions).

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*

$$\int_{-\infty}^{\infty} H'(x)\phi(x) \, dx = - \int_{-\infty}^{\infty} H(x)\phi'(x) = - \int_{0}^{\infty} \phi'(x) = \phi(0)$$

178

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 $\phi \in \mathcal{D}(\mathbb{R})$,*

$$\int_{-\infty}^{\infty} \delta(x)\phi(x)\,dx = \phi(0).$$

Then

$$\int_{-\infty}^{\infty} \delta'(x)\phi(x)\,dx = -\int_{\mathbb{R}^d} \delta(x)\phi'(x)\,dx = -\phi'(0).$$

This is a distribution that does not arise from integration with respect to a locally integrable function nor integration against a measure. But it occurs in physics, as an appropriate model of certain physical situations, i.e. for the distribution of electrical charge in a polarized point mass with positive charge infinitisimally to the left of the origin, and negative charge infinitisimally to the right of the origin.

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. Applying the Hahn-Banach theorem shows that if $\Lambda \in \mathcal{D}^*(\Omega)$ has order n, then Λ extends uniquely to a continuous linear functional on $\mathcal{D}^n(\Omega)$, since $\mathcal{D}(\Omega)$ is dense in $\mathcal{D}^n(\Omega)$. In fact, we can actually *identify* $\mathcal{D}^n(\Omega)^*$ with the space of distributions of order n.

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 given by integration against a Radon measure, and thus has order zero. To see this, given such a non-negative functional Λ (which is automatically continuous), we define Λf for a compactly supported continuous function $f \geqslant 0$ as

$$\Lambda f = \sup\{\Lambda g : g \in \mathcal{D}(\mathbb{R}^n), g \leqslant f\}$$

179

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 $\phi \in \mathcal{D}(\mathbb{R}^d)$.*

$$\Lambda[\mu](\phi) = \int_{\mathbb{R}^d} \phi(x) d\mu(x)$$

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. The Riesz-Markov-Kakutani representation theorem, shows that any *distribution of order zero is given by a complex-valued Radon measure.*

Example. *Let $U \subset \mathbb{R}^d$ be any open set such that ∂U is a C^1 hypersurface. Then we can find a C^1 function $\rho : \mathbb{R}^d \to \mathbb{R}$ such that $U = \{x : \rho(x) > 0\}$, and $\nabla \rho$ is nonzero on ∂U. Let us calculate the derivatives of the distribution $u = \mathbf{I}_U$. Clearly ∇u is supported on the surface $S = \partial U$. Let us work locally around a point $x_0 \in S$, and without loss of generality, let us rotate so that in a neighborhood of x_0, if we write $x = (x', x_d)$, then S is described by the equation $x_d = \psi(x')$, where ψ is C^1. Pick $f \in C^\infty(\mathbb{R})$ such that $f(t) = 0$ for $t < 0$ and $f(t) = 1$ for $t \geqslant 1$. Then, locally around x_0, u is the distributional limit of the functions u_ε, where*

$$u_\varepsilon(x) = f\left(\frac{x_d - \psi(x')}{\varepsilon}\right).$$

But

$$\nabla u_\varepsilon(x) = \frac{f'\left(\frac{x_d - \psi(x')}{\varepsilon}\right)}{\varepsilon} \begin{pmatrix} 1 \\ -\nabla\psi(x') \end{pmatrix}$$

which implies that for any $\phi \in \mathcal{D}(\mathbb{R}^d)$ supported near x_0, using a change of vari-

ables,

$$\int \phi(x) \nabla u(x) \, dx$$

$$= \lim_{\varepsilon \to 0} \int \phi(x) \frac{f'\left(\frac{x_d - \psi(x')}{\varepsilon}\right)}{\varepsilon} \begin{pmatrix} 1 \\ -\nabla \psi(x') \end{pmatrix}$$

$$= \lim_{\varepsilon \to 0} \int \phi(x', \psi(x') + \varepsilon t) f'(t) \begin{pmatrix} 1 \\ -\nabla \psi(x') \end{pmatrix} \, dt \, dx'$$

$$= \int \phi(x', \psi(x')) \begin{pmatrix} 1 \\ -\nabla \psi(x') \end{pmatrix} \, dx.$$

Since the surface measure on S is given by $dS = (1 + |\psi'|^2)^{1/2} \, dx'$, and the normal vector to the surface is given by

$$n = \frac{1}{(1 + |\psi'(x)|^2)^{1/2}} \begin{pmatrix} 1 \\ -\nabla \psi(x) \end{pmatrix},$$

it follows that

$$\int \phi(x) \nabla u(x) \, dx = \int_S \phi(x) n(x) \, dS.$$

Switching to the global viewpoint, we thus see that $\nabla u = n \cdot dS$. In particular, we have proved that for any smooth, compactly supported vector field $X = (X_1, \ldots, X_n)$ on $\mathbb{R}^d$,

$$\int_S (\nabla \cdot X) = \sum \int u(x) \partial_i X_i(x) \, dx = - \int_S X \cdot n \, dS,$$

so we see we have essentially proved the Gauss-Green divergence theorem. The fact that $n \cdot dS$ is a distribution of order zero now implies that the Gauss-Green formula continues to hold for any C^1 compactly supported vector field.

Looking forward to other topics, perhaps the most general setting in which the distributional derivative calculation continues to hold for any domain U which is a Caccioppoli set, i.e. such that $u = \mathbf{I}_U$ is a function with bounded variation. We can then define a Radon measure analogous to $n \cdot dS$ such that an analogue of the integration formulas continue to hold. In particular, we can still use the calculation for open sets with Lipschitz boundary, in which case the normal vector is only defined almost everywhere on the boundary.

181

Example. *Consider a functional* Λ *defined for functions* $\phi \in \mathcal{D}(\mathbb{R})$ *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* $\mathcal{D}(\mathbb{R})$. *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* $\mathcal{D}(\mathbb{R})$. *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* $\mathcal{D}(\mathbb{R})$, *and thus by the Hahn-Banach theorem, extends uniquely to at least one distribution on the entirety of* $\mathcal{D}(\mathbb{R})$.

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 $\lim_{\varepsilon \to 0}(x + i\varepsilon)^{-1}$, *defined such that*

$$\int_{-\infty}^{\infty} \lim_{\varepsilon \to 0} \frac{1}{x + i\varepsilon} \cdot \phi(x) \, dx = \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{\phi(x)}{x + i\varepsilon} \, dx.$$

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

$$\begin{aligned}
\int_{-\delta}^{\delta} \frac{\phi(x)}{x + i\varepsilon} &= \int_{-\delta}^{\delta} \frac{\phi(0)}{x + i\varepsilon} + O(\delta) \\
&= -i\pi\phi(0) + O(\varepsilon/\delta) + O(\delta) \\
&= -i\pi\phi(0) + O(\varepsilon^{1/4}).
\end{aligned}$$

182

Taking $\varepsilon \to 0$ shows that

$$p.v(1/x) = i\pi\delta + \lim_{\varepsilon \to 0} \frac{1}{x + i\varepsilon},$$

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 $\phi \in \mathcal{D}(\mathbb{R})$, 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 $\phi \in \mathcal{D}(\mathbb{R}^d)$,

$$\int (\log|x|)' \, \phi(x) \, dx = -\int \log|x|\phi'(x) \, dx$$

$$= \lim_{\varepsilon \to 0} \int_{|x| \geqslant \varepsilon} \log|x|\phi'(x)$$

$$= \lim_{\varepsilon \to 0} \left(\log(\varepsilon) \cdot (\phi(x) - \phi(-x)) + \int_{|x| \geqslant \varepsilon} \frac{\phi(x)}{x} \right)$$

$$= p.v \int \frac{\phi(x)}{x} \, dx.$$

An important analysis of these distributions arises in the theory of the Hilbert transform.

Example. *If Ω is an open subset of $\mathbb{C}$, let us calculate $\partial E_{z_0}/\partial\bar{z}$ in Ω, where $E_{z_0} = p.v\{(z - z_0)^{-1}\}$. To begin with, we note that if X is another open subset of $\mathbb{C}$ containing Ω, and if $\partial\Omega$ (the boundary of Ω relative to X) is C^1, then for $\phi \in \mathcal{D}(X)^*$, Green's formula gives*

$$\int_\Omega \frac{\partial\phi}{\partial\bar{z}} = (-i/2) \int_{\partial\Omega} \phi \, dz.$$

This formula implies the product rule, that for $\phi \in \mathcal{D}(\Omega)^*$,

$$\int_{\Omega - B_\varepsilon(z_0)} \frac{1}{z - z_0} \frac{\partial\phi}{\partial\bar{z}} = \int_{\Omega - B_\varepsilon(z_0)} \frac{\partial}{\partial\bar{z}} \left\{ \frac{1}{z - z_0}\phi \right\}$$

$$= (i/2) \int_{\partial B_\varepsilon(z_0)} \frac{1}{z - z_0}\phi \, dz.$$

183

Letting $\varepsilon \to 0$ gives $\int E_{z_0} \partial_{\bar{z}} \phi = -\pi \phi(z_0)$, or in other words, we conclude that in Ω, $\partial_{\bar{z}} E_{z_0} = \pi \delta_{z_0}$.

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 $\phi \in \mathcal{D}(\mathbb{R})$ we have

$$\int_{|x| \geqslant \delta} \frac{\phi(x)}{x^2} = \frac{2\phi(0)}{\delta} + O(\delta).$$

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$(1/x^2)$ obtained 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 $\phi(0) \neq 0$. Note that we do not need to account for $\phi'(0)$ because of cancellation on both sides of the integral.

Another approach to extending Λ 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 $1/x^3$.

184

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 $f' \in L^1(\mathbb{R})$. By Fubini's theorem, if we let μ be the measure defined by $\mu([a, b)) = f(b) - f(a)$, then for any $\phi \in \mathcal{D}(\mathbb{R})$,*

$$\int_{-\infty}^{\infty} \phi(x) d\mu(x) = -\int_{-\infty}^{\infty} \int_{x}^{\infty} \phi'(y) \, dy \, d\mu(x)$$

$$= -\int_{-\infty}^{\infty} \phi'(y) \int_{-\infty}^{y} d\mu(x) \, dy$$

$$= -\int_{-\infty}^{\infty} \phi'(y) f(y) dy$$

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 $f \in C^1(\mathbb{R} - \{0\})$, 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 $\phi \in \mathcal{D}(\mathbb{R})$,

$$\int f'(x)\phi(x) \, dx = -\int f(x)\phi'(x) \, dx$$

$$= \lim_{\varepsilon \to 0} -\int_{-\infty}^{-\varepsilon} f(x)\phi'(x) \, dx - \int_{\varepsilon}^{\infty} f(x)\phi'(x) \, dx$$

$$= \lim_{\varepsilon \to 0} f(\varepsilon)\phi(\varepsilon) - f(-\varepsilon)\phi(-\varepsilon) + \int_{-\infty}^{-\varepsilon} v(x)\phi(x) \, dx + \int_{-\infty}^{-\varepsilon} v(x)\phi(x) \, dx$$

$$= \int v(x)\phi(x) \, dx + [f(0+) - f(0-)]\phi(0).$$

As a particular example of this, the distributional derivative of $|x|$ is $\text{sgn}(x)$, and the distributional derivative of the Heaviside step function H given above is evaluated to be δ_0.

Example. *The boundary values of analytic functions are distributions, in the following manner. Let $\Omega \subset \mathbb{R}^n$ be an open set, and let Γ be a convex open cone in*

$\mathbb{R}^n$. *For $\gamma > 0$, let*

$$Z_\gamma = \{z \in \mathbb{C}^n : Re(z) \in \Omega, Im(z) \in \Gamma, |Im(z)| < \gamma\}.$$

If f is analytic in Z_γ, and there is some $N > 0$ such that $|f(x + iy)| \lesssim |y|^{-N}$. We claim then that, if for each $y \in \Gamma$, we consider the analytic function $f_y(x) = f(x + iy)$, then as $y \to 0$, $\{f_y\}$ converges distributionally to a distribution of order at most $N+1$ on Ω. In fact, we will obtain an explicit formula for this distribution. For simplicity we deal with the case $n = 1$, where the higher dimensional case is similar (see Hörmander Theorem 3.1.15). Then we might as well assume that $Z_\gamma = \{x + iy : x \in \Omega, 0 < y < \gamma\}$. Given $\phi \in C_c^{N+1}(\Omega)$, write

$$\phi_y(x) = \phi(x + iy) = \sum_{\alpha=0}^{N} \partial_x^\alpha \phi(x) \frac{(iy)^\alpha}{\alpha!}.$$

Then

$$2\frac{\partial \phi}{\partial \bar{z}} = \partial^{N+1}\phi(x)\frac{(iy)^N}{N!}.$$

Thus if $0 < y_0 < \gamma$ is fixed, and if $0 < y < \gamma - y_0$, then the Cauchy integral formula implies that

$$\int \phi(x)f(x+iy)\,dx - \int \phi(x+iy)f(x+iy+iy_0)\,dx = 2i\int\int_0^{y_0} f(x+iy+it)\frac{\partial \phi}{\partial \bar{z}}dt\,dx.$$

Thus

$$\int \phi(x)f_y(x)\,dx = \int \phi_{y_0}(x)f_{y+y_0}(x)\,dx + \int\int_0^1 f_{y+ty_0}(x)\partial^{N+1}\phi(x)\frac{(iy_0)^{N+1}t^N}{N!}\,dt\,dx.$$

Thus we conclude that as $y \to 0$,

$$\int \phi(x)f_y(x)\,dx \to \int \phi_{y_0}(x)f_{y_0}(x)\,dx + \frac{1}{N!}\int\int_0^1 f_{ty_0}(x)\partial^{N+1}\phi(x)(iy_0)^{N+1}t^N\,dt\,dx.$$

The right hand side clearly defines a distribution of order $N + 1$. In higher dimensions, we obtain the similar representation formula

$$\int \phi(x)f_y(x)\,dx \to \int \phi_{y_0}(x)f_{y_0}(x)\,dx + \frac{1}{N!}\int\int_0^1 f_{ty_0}(x)\sum_{|\alpha|=N+1}\partial^\alpha\phi(x)(iy_0)^\alpha t^N\,dt\,dx.$$

Thus we have a theory of boundary values of analytic functions.

For n = 1, we denote the distribution obtained by $f(x+i0)$. If we, by a similar process, obtained a distribution as a limit of an analytic function defined on the lower half plane, then we would denote that distribution by $f(x - i0)$. If f is analytic above and below the half plane, and satisfies estimates of the form above, then the distribution u given by

$$\int \left(u(x + iy)\phi(x, y) \ dx \right) \ dy$$

(the order of integration matters) defines a distribution of order N in the complex plane, the derivative $\partial u / \partial \overline{z}$ is supported on the real line, and on that real line, we have

$$\frac{\partial u}{\partial \overline{z}}(x) = (i/2) \left(f(x + i0) - f(x - i0) \right).$$

We can use this, for instance, to calculate the difference beteween the two distributions $(x + i0)^{-1}$ and $(x - i0)^{-1}$. We have calculated above that the distribution $u(z) = 1/z$ has

$$\frac{\partial u}{\partial \overline{z}} = \pi\delta,$$

and so $(x - i0)^{-1} = (x + i0)^{-1} + 2\pi i\delta$.

Remark. Roughly speaking, this result is tight. If an analytic function f has boundary values defining a distribution of order $N + 1$, then it is *necessary* that

$$|f(z)| \lesssim |\text{Im}(z)|^{-N-2}.$$

See Theorem 3.1.11 of Hörmander, Volume 1, for a discussion of this.

There are many other important operations one can apply to distributions. If Ω is a conic subset of $\mathbb{R}^d$, and $\phi, \psi \in \mathcal{D}(\Omega)$, we find

$$\int_\Omega \text{Dil}_\lambda\phi(x)\psi(x) \ dx = \lambda^{-d} \int_\Omega \phi(x) \cdot \text{Dil}_{1/\lambda}\psi(x) \ dx,$$

Thus if Λ is a distribution on Ω, then we define $\text{Dil}_\lambda\Lambda$ by setting

$$\text{Dil}_\lambda\Lambda(\phi) = \lambda^{-d}\Lambda(\text{Dil}_{1/\lambda}\phi).$$

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. Similarly, 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 an appropriately localized 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.

13.3   Topologies on the Space of Distributions

As a dual space to an LF space, we can equip the space of distributions $\mathcal{D}(\Omega)^*$ with several topologies. The most important for our purposes in the strong topology, which makes $\mathcal{D}(\Omega)^*$ into a complete locally convex space, and the weak $*$ topology, which is quasi-complete. These topologies are roughly the same for many purposes. For instance, the uniform boundedness theorem implies that a sequence $\{u_n\}$ of distributions converges in the weak $*$ topology if and only if it converges in the strong dual topology, and we call this convergence *distributional convergence*. We note also that $\mathcal{D}(\Omega)^*$, as the strong dual of a Montel space, is also a Montel space, i.e. it is locally convex, barelled, and satisfies the Heine-Borel property.

Example. *If $u \in \mathcal{D}(\mathbb{R}^d)$, and we set $u_\varepsilon(x) = \varepsilon^{-d} Dil_\varepsilon u(x)$, then for any $\phi \in \mathcal{D}(\mathbb{R}^d)$,*

$$\lim_{\varepsilon \to 0} \int u_\varepsilon(x)\phi(x) \, dx \to \phi(0) \int u(x) \, dx.$$

Thus u_ε converges distributionally to $(\int u(x) \, dx) \cdot \delta_0$. Similarily, if $u \in \mathcal{D}(\mathbb{R}^d)$ and for any multi-index α with $|\alpha| \leqslant k$,

$$\int u(x)x^\alpha \, dx = 0,$$

and we define $u_\varepsilon(x) = \varepsilon^{-d-k} Dil_\varepsilon u(x)$, then for any $\phi \in \mathcal{D}(\mathbb{R}^d)$,

$$\lim_{\varepsilon \to 0} u_\varepsilon(x)\phi(x) \, dx \to \frac{1}{k!} \sum_{|\alpha|=k} \left(\int x^\alpha u(x) \, dx \right) \cdot D^\alpha \phi(0)$$

Thus u_ε converges distributionally to an appropriate linear combination of $D^\alpha \delta_0$.

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* $\phi \in \mathcal{D}(\mathbb{R})$,

$$\int_{-\infty}^{\infty} t^n e^{2\pi itx}\phi(x)\ dx = i^{n+1}t^{-1}\int_{-\infty}^{\infty} e^{itx}\phi^{(n+1)}(x)\ dx,$$

which converges to zero as $t \to \infty$. *Thus* $t^n e^{2\pi itx}$ *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. Note that if we tested against functions that were less smooth (say, viewing these distributions as linear functionals on* $L^1(\mathbb{R}^d)$, *or even* C^∞ *functions that are only of polynomial decrease as they approach* ∞*) 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 $\phi \in \mathcal{D}(\mathbb{R})$, we can apply an integration by parts to write

$$\int_{-\infty}^{\infty} u_t(x)\phi(x)\,dx = \int_{-\infty}^{\infty} t^{1/k} e^{itx^k}\phi(x)\,dx$$
$$= \int_{-\infty}^{\infty} t^{1/k} F'(t^{1/k}x)\phi(x)\,dx$$
$$= -\int_{-\infty}^{\infty} F(t^{1/k}x)\phi'(x)\,dx,$$

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.

It is often of interest to focus on subfamilies of $\mathcal{D}(\Omega)^*$, equipped with a topology which is compatible with the behaviour of distributions. We thus define a *space of distributions* to be a vector subspace X of $\mathcal{D}(\Omega)^*$, equipped with a topology which makes the inclusion map $X \to \mathcal{D}(\Omega)^*_b$ continuous. Most function spaces are spaces of distributions, for instance, $\mathcal{D}(\Omega)$, L^1_{loc}, and so on. We will later encounter the space of *tempered distributions* $\mathcal{S}(\Omega)^*$, which will turn out to be a subspace of $\mathcal{D}(\Omega)^*$ equipped with the relative topology. We can then speak of 'distributional convergence' in these spaces, i.e. convergence in the strong dual.

Since convergence in $\mathcal{D}(\Omega)$ is incredibly strict, a sequence of distributions can very easily converge distributinoally. It is therefore surprising that, since a differential operator $L : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ is continuous, it's extension to a map $L : \mathcal{D}(\Omega)^* \to \mathcal{D}(\Omega)^*$, roughly speaking, a rescaling of it's adjoint, is also continuous, both in the weak $*$ topologies and in the strong topology.

With either the weak $*$ topology or the strong topologies described above, the bilinear multiplication map

$$C^{\infty}_{\text{loc}}(\Omega) \times \mathcal{D}(\Omega)^* \to \mathcal{D}(\Omega)^*$$

is sequentially continuous, but *not* jointly continuous.

13.4 Homogeneous Distributions

An important family of distributions are the *homogenous distributions*. Given any open *conic* subset Ω of $\mathbb{R}^d$, i.e. a set such that $\lambda\Omega \subset \Omega$ for any $t > 0$, we can define a distribution $\mathrm{Dil}_t u$ on Ω for each distribution u on Ω. A distribution u on Ω is *homogeneous* of order α if for any $t > 0$, $\mathrm{Dil}_t u = t^{-\alpha} u$.

Example. *If $f \in L_{1,loc}(\Omega)$ and $f(\lambda x) = \lambda^\alpha f(x)$ for all $x \in \Omega$ and $\lambda > 0$, then integration against $f(x)\, dx$ defines a homogenous distribution of order α on Ω.*

Example. *Let δ be the Dirac-delta distribution at the origin in $\mathbb{R}^d$. Then for $\phi \in \mathcal{D}(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}^d} (Dil_\lambda \delta)(x)\phi(x)\, dx = \lambda^{-d} \int_{\mathbb{R}^d} \delta(x) Dil_{1/\lambda}\phi(x)\, dx = \lambda^{-d}\phi(0).$$

Thus $Dil_\lambda \delta = \lambda^{-d}\delta$, which implies δ is a homogenous distribution of order $-d$.

Example. *For any complex number $a \in \mathbb{C}$, we define a function*

$$x_+^a = \begin{cases} x^a : x > 0, \\ 0 \qquad\quad : x \leqslant 0. \end{cases}$$

This function is locally integrable on $\mathbb{R} - \{0\}$, and so defines a distribution on this set, with the following properties:

- *The map $a \mapsto x_+^a$ is a continuous map.*

- *The distribution x_+^a is homogeneous of degree a.*

- *We have $(x_+^a)' = ax_+^{a-1}$, and $x \cdot x_+^a = x_+^{a+1}$.*

By the Hahn-Banach theorem, these distributions all extend to distributions on $\mathbb{R}^n$. It is of interest whether we can perform this extension so that the resulting distributions are homogeneous, *and that the prpoerties above continue to hold. It is clear that any two such extensions of the distributions x_+^a differ by a linear combination of Dirac delta functions, and their derivatives, so, unless $a \in \{-1, -2, \dots\}$ (since this is the homogeneity of the Dirac delta function and it's derivatives), this homogeneous extension will then be* unique.

Firstly, if $Re(a) > -1$, *then the distribution* x^a *is locally integrable on* $\mathbb{R}$, *and so it is easy to check that* x_+^a *is well defined and homogeneous of order* a *on* $\mathbb{R}$. *For* $Re(a) > 0$, *it also remains true that*

$$(x_+^a)' = a x_+^{a-1}.$$

To continue this extension to a larger range of values $Re(a) \leqslant -1$, *we rely on analytic continuation. For any* $\phi \in \mathcal{D}(\mathbb{R})$, *the function*

$$a \mapsto \langle x_+^a, \phi \rangle$$

is analytic in the region $Re(a) - 1$. *Integration by parts shows that in such a range, we have*

$$\langle x_+^{a+1}, \phi' \rangle = -(a+1)\langle x_+^a, \phi \rangle.$$

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* $a \in \mathbb{C}$ *with* $Re(a) > -2$, *except that we have a pole of order one when* $a = -1$, *which prevents us from defining the distribution at this value. Iterating this integration by parts trick allows us to analytically continuous the function* $\langle x_+^a, \phi \rangle$ *to all complex numbers* a, *except for the negative integers. Using this as a* definition *of a distribution* x_+^a, *it is immediate to check that all of these distributions are homogeneous of order* a, *so we have found the unique homogeneous extension of* x_+^a *required above. The formulas above all continue to hold,* except *that the derivative formula does not hold for* $a = 0$, *since we have not defined the distributions* $\{x_+^{-k}\}$.

As we expected, the negative integers prove to be problematic. Marcel Riesz used some other complex-analytic tricks to define x_+^{-k}, *but this is not without it's problems. For any* $\phi \in \mathcal{D}(\mathbb{R})$, *the function* $a \mapsto \langle x_+^a, \phi \rangle$ *is meromorphic, with simple poles at each integer* $-k$ *for* $k > 0$, *and the residue at* $-k$ *is equal to*

$$(-1)^k \frac{D^{k-1}\phi(0)}{(k-1)!}.$$

Thus we conclude that, as $a \to -k$,

$$\langle (a+k)x_+^a, \phi \rangle \to (-1)^{k-1}\frac{D^{k-1}\phi(0)}{(k-1)!}.$$

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. This gives an extension of the definition of x_+^{-k} on $\mathcal{D}(\mathbb{R} - \{0\})$, because the infinite part of the expansion above vanishes for functions supported away from the origin. For values a close to k, we have*

$$\langle x_+^a, \phi \rangle - \frac{(-1)^{k-1} D^{k-1} \phi(0)}{(k-1)!} \cdot (a+k)^{-1}$$

$$= (-1)^k (a+1)^{-1} \cdots (a+k-1)^{-1} \int_0^\infty \frac{x^{a+k} - 1}{a+k} D^k \phi(x) \, dx$$

$$+ \frac{(-1)^{k-1}}{a+k} \left((-1-a)^{-1} \ldots (1-a-k) - \frac{1}{(k-1)!} \right) D^{k-1} \phi(0)$$

$$\to \frac{-1}{(k-1)!} \int_0^\infty \log(x) D^k \phi(x) \, dx + \frac{1}{(k-1)!} \left(\sum_{j=1}^k 1/j \right) D^{k-1} \phi(0).$$

and thus

$$\langle x_+^{-k}, \phi \rangle = \frac{-1}{(k-1)!} \int_0^\infty \log(x) D^k \phi(x) \, dx + \frac{1}{(k-1)!} \left(\sum_{j=1}^k 1/j \right) D^{k-1} \phi(0).$$

This definition leads to a definition of $\{x_+^a\}$ for all $a \in \mathbb{C}$. The algebraic formula

$$x \cdot x_+^a = x_+^{a+1}$$

continues to hold in this range.

Our extension of x_+^a for all $a \in \mathbb{C}$ continues to satisfy $x \cdot x_+^a = x_+^{a+1}$ for all $a \in \mathbb{C}$. The derivative formula is not *maintained when k is an integer, however. Namely, we have*

$$D x_+^{-k} = -k x_+^{-k-1} + (-1)^k D^k \delta / k!$$

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 analogously define the homogeneous distributions x_-^a, and $|x|^a$, by reflecting and symmetrizing the constructions we have given above.

Example. *Another way to fix the definition of the distributions $\{x_+^{-k}\}$ is by appropriately normalizing the distributions to remove the singularities here when we apply analytic continuation. Namely, set*

$$\chi_+^a = \frac{x_+^a}{\Gamma(a+1)}.$$

Then $D(\chi_+^a) = -\chi_+^{a-1}$, and because the Gamma function has no zeroes, this allows us to extend the definition of $\{\chi_+^a\}$ to a family of homogeneous distributions of order a, for any $a \in \mathbb{C}$. Since χ_+^0 is the Heaviside step function, we conclude that $\chi_+^{-1} = -\delta$, and more generally, that $\chi_+^{-k} = (-1)^k D^{k-1}\delta$, which is somewhat undesirable.

Given any $a \in \mathbb{C}$, and any distribution v on S^{n-1}, we can define a homogeneous distribution u of order a on $\mathbb{R}^n - \{0\}$, such that for $\phi \in \mathcal{D}(\mathbb{R}^n - \{0\})$,

$$\langle u, \phi \rangle = \int_0^\infty r^{a+n-1} \int_{S^{n-1}} v(x)\phi(rx)\, d\sigma(x).$$

For a fixed a, this is actually a *bijective correspondence*. Given any distribution $u \in \mathcal{D}(\mathbb{R}^n - \{0\})^*$, which is homogeneous of degree a, there exists a *unique* distribution v on S^{n-1} with this property. Indeed, if $\eta \in \mathcal{D}((0,\infty))$ is supported on a neighborhood of 1 and has

$$\int_0^\infty \eta(x)\, dx = 1,$$

and if v existed, then for $\psi \in \mathcal{D}(S^{n-1})$, if for $r > 0$ and $|x| = 1$ we set $\phi(rx) = \psi(x)r^{1-n-a}\eta(r)$, then we would have

$$\langle v, \psi \rangle = \langle u, \phi \rangle.$$

One can check that, given any u, we can use this construction to define a distribution v on S^{n-1} that induces u. On the other hand, one can change the order of integration, i.e. we have

$$\langle u, \phi \rangle = \int_{S^{n-1}} v(x) \int_0^\infty r^{a+n-1}\phi(rx)\, d\sigma(x).$$

In particular, this implies that for any $\phi \in \mathcal{D}(\mathbb{R}^n - \{0\})$, such that

$$\int_0^\infty r^{a+n-1}\phi(rx)\, dr = 0$$

for all $x \in \mathbb{R}^d - \{0\}$, we have $\langle u, \phi \rangle = 0$.

Theorem 13.3. *Let u be a homogeneous distribution on $\mathbb{R}^n - \{0\}$ which has order $a \in \mathbb{C}$, where a is* not *an integer less than or equal to $-n$. Then u has a unique extension to a distribution $E(u)$ on $\mathbb{R}^n$, such that for any homogeneous polynomial P,*

$$E(Pu) = PE(u),$$

and if $a \neq -(n-1)$, $E(\partial_i u) = \partial_i E(u)$. Moreover, the map $u \mapsto E(u)$ is continuous from $\mathcal{D}(\mathbb{R}^n - \{0\})^*$ to $\mathcal{D}(\mathbb{R}^n)^*$.

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 apply the theory we discussed before the notes above. Any homogeneous distribution u on $\mathbb{R}^n - \{0\}$ of order a induces a homogeneous distribution v on S^{n-1} such that for $\phi \in \mathcal{D}(\mathbb{R}^n - \{0\})$,

$$\langle u, \phi \rangle = \int_{S^{n-1}} v(x) \int_0^\infty r^{a+n-1} \phi(rx) \, d\sigma(x).$$

The challenge is to interpret the right hand side for $\phi \in \mathcal{D}(\mathbb{R}^n)$. A natural choice, given $\phi \in \mathcal{D}(\mathbb{R}^n)$, is to define $R_a \phi \in \mathcal{D}(S^{n-1})$ by setting

$$R_a \phi(x) = \int t_+^{a+n-1} \phi(tx) \, dt,$$

the right hand side being well defined whenever a is not a negative integer less than or equal to $-n$. But we can now define

$$\langle Eu, \phi \rangle = \int_{S^{n-1}} v(x) R_a \phi(x) \, dx.$$

This clearly gives an extension of u to a distribution on $\mathbb{R}^n$. To show that Eu remains homogeneous of degree a, we have to show that

$$\langle Eu, \mathrm{Dil}_t \phi \rangle = t^{a+n} \langle Eu, \phi \rangle.$$

But we calculate that $R_a \{\mathrm{Dil}_t \phi\} = t^{a+n} R_a \phi$, which gives the required result. Continuity follows from the continuity of the map $u \mapsto v$ and the map $\phi \mapsto R_a \phi$. $\qquad \square$

When a is an integer smaller than or equal to $-n$, one can still use the construction performed above to define an operator E from homogeneous distributions on $\mathbb{R}^n - \{0\}$ to distributions on $\mathbb{R}^n$. *However*, the resulting distribution can

fail to be homogeneous, and may fail to satisfy the identities $E(Pu) = PE(u)$ and $E(\partial_i u) = \partial_i E(u)$. Fortunately, we do see that an extension exists, and the extension *will* be unique, up to a linear combination of Dirac delta functions.

There is a special family of homogeneous distributions that often occurs in harmonic analysis, that being the homogeneous distributions of order a on $\mathbb{R}^n$ that are C^∞ away from the origin, i.e. distributions u that agree with the action of a C^∞ function u_0 as a distribution tested against elements of $\mathcal{D}(\mathbb{R}^n - \{0\})$. Of course, u_0 must be a homogeneous function of order a. If $\operatorname{Re}(a) > -n$, then this implies in fact that u_0 is *locally integrable*, and that, in fact, for any $\phi \in \mathcal{D}(\mathbb{R}^n)$,

$$\langle u, \phi \rangle = \int u_0(x)\phi(x)\, dx.$$

On the other hand, if $\operatorname{Re}(a) \leqslant -n$, there may be more funny business at the origin. Nonetheless, the borderline case when $\operatorname{Re}(a) = -n$ is of interest. In this case, if

$$\int_{|x|=1} u_0(x)\, d\sigma(x) = 0,$$

then we can define a principal value distribution $\mathrm{p.v}(u_0)$ by setting

$$\langle \mathrm{p.v}(u_0), \phi \rangle = \lim_{\varepsilon \to 0} \int_{|x| \geqslant \varepsilon} u_0(x)\phi(x)\, dx.$$

The limit converging. If we consider a radial $\psi \in \mathcal{S}(\mathbb{R}^d)$, we can also write the distribution so

$$\langle \mathrm{p.v}(u_0), \phi \rangle = \int u_0(x)\left(\phi(x) - \phi(0)\psi(x)\right)\, dx,$$

since then the integrand is an element of L^1 by virtue of the rapid decay of the function inside the brackets, and that this same function is $O(|x|)$ as $x \to 0$. The integral also does not depend on the choice of ψ since, given any two choices ψ_1 and ψ_2, we have

$$\int u_0(x)(\psi_1(x) - \psi_2(x))\, dx = 0,$$

because u_0 has cancellation properties on circles about the origin. Clearly $u - \mathrm{p.v}(u_0)$ is a distribution supported at the origin, and thus there exists some constant c such that

$$u = c\delta + \mathrm{p.v}(u_0).$$

More generally, if we have

$$c = \fint_{|x|=1} u_0(x)\, d\sigma(x),$$

then one need only define the distribution $E_\psi r^{-n}$ by setting

$$\langle E_\psi r^{-n}, \phi \rangle = \int |x|^{-n} \left(\phi(x) - \phi(0)\psi(x) \right).$$

Unlike in the principal value case, this assumption *depends* on ψ, i.e. we have

$$E_{\psi_1} r^{-n} - E_{\psi_2} r^{-n} = \left[\int |x|^{-n} (\psi_1(x) - \psi_2(x))\, dx \right] \delta.$$

Moreover, $E_\psi r^{-n}$ is not homogeneous, since one can verify via a calculation that

$$\mathrm{Dil}_t (E_\psi r^{-n}) = t^n E_\psi r^{-n} + (A_{n-1} t^n \log t)\, \delta,$$

where A_{n-1} is the surface area of S^{n-1}. In particular, we see from this computation that if u is a homogeneous distribution of degree n *on* $\mathbb{R}^n$ (rather than just on $\mathbb{R}^n - \{0\}$), then we *must* have

$$\int_{|x|=1} u(x)\, dx = 0,$$

since otherwise u would not be homogeneous.

The study of homogeneous functions of order a when a is an integer less than or equal to $-n$ is often, unfortunately, necessary. A common example occurs when constructing fundamental solutions to the Laplace operator Δ in $\mathbb{R}^2$, i.e. tempered distributions u on $\mathbb{R}^2$ solving the equation $\Delta u = \delta$. Taking Fourier transforms yields that $-4\pi^2 |\xi|^2 \hat{u} = 1$. We are thus lead to conclude that, away from the origin, the tempered distribution $\hat{u}$ agrees with the homogeneous function $-1/4\pi^2 |\xi|^2$. We have

$$\int_{|\xi|=1} \frac{-1}{4\pi^2 |\xi|^2}\, d\xi = \frac{-1}{2\pi}.$$

Thus if we let $\psi(x) = e^{-\pi |x|^2}$, then the distribution $\hat{u}$ agrees with the distribution $(-1/4\pi^2)(E_\psi r^{-2})$ away from the origin. Thus we can write

$$\hat{u} = c\delta - \frac{1}{4\pi^2} \left(E_\psi r^{-2} \right).$$

197

To determine u, we need to take the inverse Fourier transform of $E_\psi r^{-2}$. But this is not too difficult, because $E_\psi r^{-2}$ is a radial distribution, so it's inverse Fourier transform is also radial. Moreover, we have the dilation formula

$$\mathrm{Dil}_t\left(E_\psi r^{-2}\right) = t^2\left[\left(E_\psi r^{-2}\right) + +2\pi \log t\delta\right].$$

Thus if v is the inverse Fourier transform of $E_\psi r^{-2}$, then taking Fourier transforms of both sides yields that

$$v(tx) = v(x) + 2\pi \log t.$$

Since v is radial, away from the origin we see that $v(x) = a + 2\pi \log |x|$, and this function is locally integrable, so we conclude that

$$u(x) = a + 2\pi \log |x|.$$

The fact we were dealing with a homogeneous partial differential equation of degree two on $\mathbb{R}^2$ lead us to construct non-homogeneous fundamental solutions. For $n > 2$, the fundamental solution is a homogeneous function of degree $2 - n$.

13.5   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 $\phi \in \mathcal{D}(U)$. 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 13.4. *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

$$\Lambda(\phi) = \sum \Lambda(\psi_\alpha \phi) = \sum \Psi(\psi_\alpha \phi) = \Psi(\phi)$$

198

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. $\qquad\square$

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 13.5. *If a distribution $\Lambda \in \mathcal{D}^*(\Omega)$ has compact support, then Λ has some finite order n, and extends uniquely to a continuous functional on $C_{loc}^n(\Omega)$.*

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 $C^\infty(X)$. $\qquad\square$

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

$$|\Lambda(\phi)| \leqslant \|\phi'\|_{L^\infty(\mathbb{R}^d)},$$

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.

In general, if Λ is a distribution of order k supported on a compact set K, though we do not have a uniform bound of the form

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant k} \|D^\alpha \phi\|_{L^\infty(K)},$$

what *is* true is that if the right hand side vanishes, so does the left hand side.

Lemma 13.6. *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 $\varepsilon \to 0$. $\qquad\square$

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

$$\Lambda(\phi) = \sum_{|\alpha| \leqslant k} \lambda_\alpha D^\alpha \phi(x_0).$$

This means that Λ is a sum of Dirac delta functions and their derivatives. If we work harder, using the Whitney extension theorem as a black box, we can obtain a similar process for more general supports.

Theorem 13.7 (Whitney). *Let K be a compact set in $\mathbb{R}^d$, 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 $v \in C^k(\mathbb{R}^d)$ 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 13.8. *For* any *compact set K, and any distribution* Λ *of order k supported on K, we have*

$$|\Lambda(\phi)| \lesssim \sum_{|\alpha| \leqslant k} \sup_{x,y \in K} \left| D^\alpha \phi(x) - \sum_{|\beta| \leqslant k-|\alpha|} D^{\alpha+\beta} \phi(y) \cdot (x-y)^\beta/\beta! \right| \cdot |x-y|^{|\alpha|-k}$$
$$+ \sum_{|\alpha| \leqslant k} \|D^\alpha \phi\|_{L^\infty(K)}.$$

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 $\psi \in C^k(\mathbb{R}^n)$ 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. $\qquad\square$

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 13.9. *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

$$|dF_\alpha/ds| \leqslant \sum_{i=1}^d \left| \left(D^{\alpha+i} \phi(\gamma(s)) - \sum_{|\beta| \leqslant k-|\alpha|} D^{(\alpha+i)+(\beta-i)} \phi(y)(\gamma(s) - y)^{\beta-i}/(\beta-1)! \right) \cdot \gamma_i'(s) \right|$$
$$\lesssim s^{k-|\alpha|-1} \sum_{|\beta|=k} \|D^\beta \phi\|_{L^\infty(K)}$$

202

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. $\qquad\square$

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 13.10. *Let $x = (x_0, x_1)$, where $x_0 \in \mathbb{R}^{d_1}$, $x_1 \in \mathbb{R}^{d_2}$, and $d = d_1 + d_2$. Let $H = \{(x_0, x_1) \in \mathbb{R}^d : x_1 = 0\}$. If Λ is a distribution of order k compactly supported on H, then there exists distributions Λ_α of order $k - |\alpha|$ on $\mathbb{R}^{d_1}$ for each $|\alpha| \leqslant k$, where α is a multi-index in the $\mathbb{R}^{d_2}$ variables, and constants γ_α such that*

$$\Lambda(\phi) = \sum \Lambda_\alpha(D^\alpha \phi|_H).$$

Proof. Fix a function $\psi \in \mathcal{D}(\mathbb{R}^{d_1})$ equal to one in a neighborhood of the origin. Given $\phi \in \mathcal{D}(\mathbb{R}^{d_1})$, 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 $\mathbb{R}^{d_1}$ such that for $\psi \in \mathcal{D}(\mathbb{R}^{d_1})$,

$$\Lambda_\alpha(\psi) = \Lambda(\psi \otimes (x_1^\alpha/\alpha!)),$$

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,

$$\begin{aligned}
|\Lambda_\alpha(\psi)| &\lesssim \sum_{|\beta_1| + |\beta_2| \leqslant k} \left\| D^{\beta_1 + \beta_2} \left\{ \psi \otimes (x_1^\alpha/\alpha!) \right\} \right\|_{L^\infty(B \times \{0\})} \\
&= \sum_{|\beta| \leqslant k - |\alpha|} \| D^\beta \psi \|_{L^\infty(B)}.
\end{aligned}$$

This implies Λ_α has order $k - |\alpha|$. $\qquad\square$

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.

13.6   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 13.11. *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 □

Theorem 13.12. *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 13.13. *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.*

13.7 Convolutions of Distributions

Using the convolution of two functions as inspiration, we will define the convolution of a distribution Λ with a test function ϕ, and under certain conditions, the convolution of two distributions. Recall that if $f, g \in L^1(\mathbb{R}^n)$, then their convolution is the function in $L^1(\mathbb{R}^n)$ defined by

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

If we define the translation operators $T_y g(x) = \mathrm{Trans}_y g(x) = g(x - y)$, 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 $\mathbb{R}^n$, and ϕ is a test function on $\mathbb{R}^n$, we can define a function $\Lambda * \phi$ by setting $(\Lambda * \phi)(x) = \Lambda(T_x \phi^*)$. Notice that since

$$\int (T_x f)(y)g(y) \, dy = \int f(y - x)g(y) \, dy = \int f(y)g(x + y) \, dy$$
$$= \int f(y)(T_{-x}g)(y) \, dy,$$

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 13.14. $\Lambda * \phi$ *is* C^∞, *and*

$$D^\alpha(\Lambda * \phi) = (D^\alpha \Lambda) * \phi = \Lambda * (D^\alpha \phi).$$

205

Proof. It is easy to calculate that

$$(D^\alpha \Lambda * \phi)(x) = (D^\alpha \Lambda)(\phi_x^*) = (-1)^{|\alpha|} \Lambda(D^\alpha(T_x \phi^*))$$
$$= \Lambda(T_x(D^\alpha \phi)^*) = (\Lambda * D^\alpha \phi)(x)$$

If $k \in \{1, \ldots, d\}$ and $h \in \mathbb{R}$, we set

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

then $\Delta_h \phi$ converges to $D^k \phi$ in $\mathcal{D}(\mathbb{R}^d)$, and as such

$$\Delta_h(\Lambda * \phi)(x) = \frac{(\Lambda * \phi)(x + he_k) - (\Lambda * \phi)(x)}{h}$$
$$= \Lambda \left(\frac{T_{-x-he_k} \phi^* - T_{-x} \phi^*}{h} \right)$$

As $h \to 0$, in $\mathcal{D}(\mathbb{R}^d)$ 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 $\Lambda * \phi \in C^\infty(\mathbb{R}^d)$. An easy calculation then shows that for each $x \in \mathbb{R}^d$,

$$[(D^\alpha \Lambda) * \phi](x) = (D^\alpha \Lambda)(T_{-x} \phi^*)$$
$$= (-1)^{|\alpha|} \Lambda(T_{-x} D^\alpha \phi^*)$$
$$= \Lambda(T_{-x}(D^\alpha \phi)^*)$$
$$= (\Lambda * D^\alpha \phi)(x). \qquad \square$$

There is a certain duality going on here. Distributions can be viewed as linear functionals on $\mathcal{D}(\mathbb{R}^d)$, but one can also view them as a certain family of linear operators from $\mathcal{D}(\mathbb{R}^d) \to C^\infty(\mathbb{R}^d)$, and the convolution operator uniquely represents the distribution. In fact, any such operator that is translation invariant and continuous can be represented as convolution by a distribution.

206

Theorem 13.15. *Let* $T : \mathcal{D}(\mathbb{R}^d) \to C_{loc}^\infty(\mathbb{R}^d)$ *be a translation invariant continuous operator. Then there exists a distribution* Λ *such that* $T\phi = \Lambda * \phi$ *for all* $\phi \in \mathcal{D}(\mathbb{R}^d)$.

Proof. If we knew $T\phi = \Lambda * \phi$ for some Λ, then we could recover Λ since

$$\int \Lambda(x)\phi(x) \, dx = T\phi(0).$$

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 $\phi \in \mathcal{D}(\mathbb{R}^d)$. $\square$

Example. *A linear differential operator* $P : \mathcal{D}(\mathbb{R}^d) \to \mathcal{D}(\mathbb{R}^d)$ *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.*

Theorem 13.16. *If* $\phi, \psi \in \mathcal{D}(\mathbb{R}^n)$, *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 $x \in \mathbb{R}^d$,

$$(\phi * \psi)^*(x) = \int \phi^*(x+y)\psi(y) \, dy = \int (T_y\phi^*)(x)\psi(y) \, dy$$

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

$$\big(\Lambda * (\phi * \psi)\big)(0) = \Lambda\big((\phi * \psi)^*\big) = \int_K \psi^*(y)\Lambda(T_y\phi^*) \, dy$$

$$= \int_K \psi^*(y)(\Lambda * \phi)(y) \, dy = \big((\Lambda * \phi) * \psi\big)(0)$$

The commutativity in general results from applying the commutativity of the translation operators. $\square$

A sequence $\{\phi_n\}$ in $\mathcal{D}(\mathbb{R}^n)$ is known as an *approximate identity* in the space of distributions if $\Lambda * \phi_n$ converges to Λ as $n \to \infty$, for every distribution Λ, and an approximate identity in the space of test functions if $\psi * \phi_n$ converges to ψ in $\mathcal{D}(\mathbb{R}^n)$ for every $\psi \in \mathcal{D}(\mathbb{R}^n)$.

Theorem 13.17. *If $\{\phi_n\}$ is a family of non-negative functions in $\mathcal{D}(\mathbb{R}^n)$ which are eventually supported on every neighbourhood of the origin, and all integrate to one, then $\{\phi_n\}$ 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 $f * \phi_n$ converges locally uniformly to f as $n \to \infty$. But now we calculate that if $f \in \mathcal{D}(\mathbb{R}^n)$, then $D^\alpha(f * \phi_n) = (D^\alpha f) * \phi_n$ converges locally uniformly to $D^\alpha \phi$, which gives that $f * \phi$ converges to f in $\mathcal{D}(\mathbb{R}^n)$. Now if Λ is a distribution, and ψ is a test function, then continuity gives

$$\Lambda(\psi^*) = \lim_{\delta \to 0} \Lambda(\phi_\delta * \psi) = \lim_{\delta \to 0}(\Lambda * (\phi_\delta * \psi))(0)$$
$$= \lim_{\delta \to 0}((\Lambda * \phi_\delta) * \psi)(0) = \lim_{\delta \to 0}(\Lambda * \phi_\delta)(\psi^*)$$

and ψ was arbitrary. $\qquad \square$

If Λ is a distribution on $\mathbb{R}^n$, then the map $\phi \mapsto \Lambda * \phi$ is a linear transformation from $\mathcal{D}(\mathbb{R}^n)$ into $\mathcal{E}(\mathbb{R}^n)$, 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 $C_{\text{loc}}^\infty(\mathbb{R}^n)$. Both of these spaces are Fréchet, so we can apply the closed graph theorem. Consider a sequence $\{\phi_n\}$ converging in $C_c^\infty(K)$ to some ϕ, and we suppose that $\{\Lambda * \phi_n\}$ converges in $C_c^\infty(K)$ to a function f. It suffices to show that $\Lambda * \phi = f$. But we calculate that for each $x \in \mathbb{R}^d$,

$$f(x) = \lim_n (\Lambda * \phi_n)(x) = \lim \Lambda(T_x \phi_n^*) = \Lambda(\lim T_x \phi_n^*) = \Lambda(T_x \phi^*) = (\Lambda * \phi)(x).$$

Here we have used the fact that $T_x \phi_n^*$ converges to $T_x \phi^*$ in $\mathcal{D}(\mathbb{R}^n)$. Thus $\phi \mapsto \Lambda * \phi$ is a continuous, translation invariant operator from $\mathcal{D}(\mathbb{R}^n)$ to $\mathcal{E}(\mathbb{R}^n)$. Surprisingly, the converse is also true.

Theorem 13.18. *If $L : \mathcal{D}(\mathbb{R}^n) \to \mathcal{E}(\mathbb{R}^n)$ is a continuous linear transformation commuting 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. $\qquad \square$

We now move onto the case where a distribution Λ has compact support. Then Λ extends to a continuous linear functional on $\mathcal{E}(\mathbb{R}^n)$, and we can define the convolution $\Lambda * \phi$ if $\phi \in \mathcal{E}(\mathbb{R}^n)$ as above. This is an extension of the convolution operator above to a continuous operator from $\mathcal{E}(\mathbb{R}^n)$ to itself. The same techniques as before, or a density argument, verify that translations and derivatives are carried into the convolution.

Theorem 13.19. *If ϕ and Λ have compact support on $\mathbb{R}^d$, then $\Lambda * \phi$ has compact support, and moreover,*

$$supp(\Lambda * \phi) \subset supp(\Lambda) + supp(\phi).$$

The map $\phi \mapsto \Lambda * \phi$ is a continuous operator from $\mathcal{D}(\mathbb{R}^d)$ to itself.

Proof. Let ϕ be supported on a compact set K, and Λ be supported on K_0. Then $(\Lambda * \phi)(x) = \Lambda(T_x \phi^*)$. Since $T_x \phi^*$ is supported on $x - K$, $(\Lambda * \phi)(x) = 0$ unless $K_0 \cap (x - K) \neq \varnothing$, i.e. $x \in K_0 + K$.

To obtain continuity of the operator, it suffices to prove that $\phi \mapsto \Lambda * \phi$ is a continuous operator from $C_c^\infty(K)$ to $\mathcal{D}(\mathbb{R}^d)$ for any compact set K. Let Λ be supported on a compact set K_0. Then $\Lambda * \phi$ is supported on $K + K_0$, and so it suffices to show $\phi \mapsto \Lambda * \phi$ is a continuous operator from $C_c^\infty(K)$ to $C_c^\infty(K + K_0)$. But this follows because the map is continuous into $C_{loc}^\infty(\mathbb{R}^d)$. $\qquad\square$

Theorem 13.20. *If Λ and ψ have compact support, and $\phi \in C^\infty(\mathbb{R}^n)$, 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. $\qquad\square$

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

$$L(\phi) = \Lambda * (\Psi * \phi)$$

This gives a continuous, translation invariant linear operator from $\mathcal{D}(\mathbb{R}^d)$ to $\mathcal{E}(\mathbb{R}^d)$; if Ψ is compactly supported, then $\phi \mapsto \Psi * \phi$ is a continuous operator on $\mathcal{D}(\mathbb{R}^d)$, which gives continuity. If Λ is compactly supported, then $\Lambda \mapsto \Lambda * \eta$ is a continuous linear operator on $\mathcal{E}(\mathbb{R}^d)$, which gives continuity. But this means that L corresponds to convolution with a distribution, and we define this distribution to be $\Lambda * \Psi$.

Theorem 13.21. *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 $S_{\Lambda * \Psi} \subset S_\Lambda + S_\Psi$.*

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

$$\Psi * ((\Lambda * \phi) * \psi) = \Psi * (\Lambda * (\phi * \psi)) = \Psi * (\Lambda * (\psi * \phi))$$
$$= \Psi * ((\Lambda * \psi) * \phi) = (\Psi * \phi) * (\Lambda * \psi)$$

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 $\Lambda * \Psi$. □

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 13.22. *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. □

211

13.8 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 $\mathcal{S}(\mathbb{R}^d)$ denote a class of functions under which to study this phenomenon, it must be contained in $L^1(\mathbb{R}^d)$ and $C^\infty(\mathbb{R}^d)$, 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 $\mathcal{S}(\mathbb{R}^d)$ be the space of all functions $f \in C^\infty(\mathbb{R}^d)$ such that for any integer n and multi-index α, $|x|^n D^\alpha f \in L^\infty(\mathbb{R}^d)$. The space $\mathcal{S}(\mathbb{R}^d)$ is then locally convex if we consider the family of seminorms

$$\|f\|_{\mathcal{S}^{n,m}(\mathbb{R}^d)} = \sup_{|\alpha| \leqslant n} \|\langle x \rangle^m D^\alpha f\|_{L^\infty(\mathbb{R}^d)}.$$

Elements of $\mathcal{S}(\mathbb{R}^d)$ are known as *Schwartz functions*, and $\mathcal{S}(\mathbb{R}^d)$ is often known as the *Schwartz space*. The seminorms naturally give $\mathcal{S}(\mathbb{R}^d)$ the structure of a Fréchet space. Sometimes, it is more convenient to use the equivalent family of seminorms $\|f\|_{\mathcal{S}^{\alpha,\beta}(\mathbb{R}^d)} = \|x^\alpha D^\beta f\|_{L^\infty(\mathbb{R}^d)}$, because x^α often behaves more nicely under various Fourier analytic operations. It is obvious that $\mathcal{S}(\mathbb{R}^d)$ is separated by the seminorms defined on it, because $\|\cdot\|_{L^\infty(\mathbb{R}^d)} = \|\cdot\|_{\mathcal{S}^{0,0}(\mathbb{R}^d)}$ is a norm used to define the space. We now show the choice of seminorms make the space complete.

Theorem 13.23. $\mathcal{S}(\mathbb{R}^d)$ *is a complete metric space.*

Proof. Let $\{f_i\}$ be a Cauchy sequence with respect to the seminorms $\|\cdot\|_{\mathcal{S}^{n,\alpha}(\mathbb{R}^d)}$. 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 $L^\infty(\mathbb{R}^d)$. Since $L^\infty(\mathbb{R}^d)$ 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 $\mathcal{S}(\mathbb{R}^d)$. $\qquad\square$

Example. *The Gaussian function* $\phi : \mathbb{R}^d \to \mathbb{R}$ *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* $x \in \mathbb{R}^d$*,*

$$|(1 + |x|^m) D^\alpha \phi| \lesssim_{\alpha,m} 1.$$

Since m *and* α *were arbitrary, this shows* ϕ *is Schwartz.*

Example. *Any compactly supported smooth function is Schwartz. In particular, the inclusion map*

$$\mathcal{D}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$$

is bounded. The proof is left to the reader. An important consequence is that elements of $\mathcal{S}(\mathbb{R}^d)^*$, which we will call tempered distributions*, can be viewed as a subspace of the space $\mathcal{D}(\mathbb{R}^d)^*$ of all distributions.*

To show that an operator T on $\mathcal{S}(\mathbb{R}^d)$ is bounded, it suffices to show that for each n_0 and m_0, there is n_1, m_1 such that

$$\|Tf\|_{\mathcal{S}^{n_0,m_0}(\mathbb{R}^d)} \lesssim_{n_0,m_0} \|f\|_{\mathcal{S}^{n_1,m_1}(\mathbb{R}^d)}.$$

For a functional $\Lambda : \mathcal{S}(\mathbb{R}^d) \to \mathbb{R}$, it suffices to show that there exists n and m such that $|\Lambda f| \lesssim \|f\|_{\mathcal{S}^{n,m}(\mathbb{R}^d)}$. 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. Here are some examples:

- Multiplication gives a continuous bilinear map

$$\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d).$$

  More generally, the bilinear multiplication map

$$C_b^\infty(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$$

  is continuous.

- For each $x \in \mathbb{R}^d$, the translation operator

$$\mathrm{Trans}_x : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$$

  is an isomorphism. For each $\xi \in \mathbb{R}^d$, the modulation operator

$$\mathrm{Mod}_\xi : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$$

  is an isomorphism.

- The L^p norms are continuous.

- The Fourier transform is an isomorphism of $\mathcal{S}(\mathbb{R}^d)$.

The last point follows by an application, e.g. of the compatibility of differentiation and polynomial multiplication with the Fourier transform.

Corollary 13.24. *Convolution is a continuous operator*

$$\mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d).$$

Proof. We have

$$f * g = \mathcal{F}^{-1}\{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}\},$$

and the result then follows from the previous examples. $\qquad\square$

Remark. This result allows us to define the convolution of any two tempered distributions in a natural way extending the theory of convolution of distributions defined before.

Now we get to the interesting part of the theory of Schwartz functions. We have defined a homeomorphic linear transform from $\mathcal{S}(\mathbb{R}^d)$ to itself. The theory of functional analysis then says that we can define a dual map, which is a homeomorphism from the dual space $\mathcal{S}(\mathbb{R}^d)^*$ to itself. Note the inclusion map $\mathcal{D}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ is continuous, and $\mathcal{D}(\mathbb{R}^d)$ is dense in $\mathcal{S}(\mathbb{R}^d)$. This implies that we have an injective, continuous map from $\mathcal{S}(\mathbb{R}^d)^*$ to $\mathcal{D}(\mathbb{R}^d)^*$, 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 $\mathcal{D}(\mathbb{R}^d)$ which have a continuous extension to $\mathcal{S}(\mathbb{R}^d)$. Intuitively, this corresponds to having limited growth at infinity.

Example. *Recall that for any $f \in L^1_{loc}(\mathbb{R}^d)$, 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 $f \in L^p(\mathbb{R}^d)$ for some p, then, applying Hölder's inequality, we obtain that

$$|\Lambda[f](\phi)| \leqslant \|f\|_{L^p(\mathbb{R}^d)} \|\phi\|_{L^q(\mathbb{R}^d)}.$$

Since $\| \cdot \|_{L^q(\mathbb{R}^d)}$ is a continuous norm on $\mathcal{S}(\mathbb{R}^d)$, this shows $\Lambda[f]$ is bounded. More generally, if $f \in L^1_{loc}(\mathbb{R}^d)$, and $f(x)(1 + |x|)^{-m}$ is in $L^p(\mathbb{R}^d)$ 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 $\|\Lambda[\mu](\phi)\| \leqslant \|\mu\| \|\phi\|_{L^\infty(\mathbb{R}^d)}$ gives boundedness. More generally, if μ is a measure such that for some n,

$$\int_{\mathbb{R}^d} \frac{d|\mu|(x)}{1 + |x|^n} \, dx < \infty$$

then μ is known as a tempered measure*, and acts as a tempered distribution since*

$$|\Lambda[\mu](\phi)| \leqslant \int_{\mathbb{R}^d} |\phi(x)| \, d|\mu|(x)$$

$$\leqslant \left(\int_{\mathbb{R}^d} \frac{d|\mu|(x)}{1 + |x|^n} \, dx \right) \cdot \|\phi\|_{\mathcal{S}^{0,n}(\mathbb{R}^d)}.$$

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 $C^\infty(\mathbb{R}^d)$. We then find*

$$|\Lambda(\phi)| \lesssim \|\phi\|_{C^n(\mathbb{R}^d)} \leqslant \|\phi\|_{\mathcal{S}^{0,n}(\mathbb{R}^d)}.$$

Example. *The distribution Λ on $\mathbb{R}$ given by*

$$\Lambda(\phi) = p.v \int_{-\infty}^{\infty} \frac{\phi(x)}{x} \, dx$$

is tempered, since

$$\int_{|x| \geqslant 1} \frac{\phi(x)}{x} \lesssim \|\phi\|_{\mathcal{S}^{1,0}(\mathbb{R}^d)}$$

and

$$p.v \int_{-\infty}^{\infty} \frac{\phi(x)}{x} \, dx \lesssim \|\phi\|_{C^1(\mathbb{R}^d)} = \|\phi\|_{\mathcal{S}^{0,1}(\mathbb{R}^d)}$$

and so Λ is tempered of order 1. This distribution is called the (Cauchy) principal value *of $1/x$, often denoted $p.v(1/x)$.*

The derivative of a tempered distribution is tempered, and gives a continuous operator on $\mathcal{S}(\mathbb{R}^d)^*$. Multiplication gives a continuous map

$$C_b^\infty(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d)^* \to \mathcal{S}(\mathbb{R}^d)^*.$$

215

Furthermore, multiplication by a polynomial is also a continuous operator on $\mathcal{S}(\mathbb{R}^d)$, or more generally, by any smooth function whose derivatives are all slowly increasing.

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

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

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 $f \in \mathcal{S}(\mathbb{R}^d)$, the distribution $\widehat{\Lambda}$ corresponding to $\widehat{f}$, then for any Schwartz $\phi \in \mathcal{S}(\mathbb{R}^d)$,

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

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* $\mathbb{R}^d$*. Then for any* $\phi \in \mathcal{S}(\mathbb{R}^d)$*,*

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

Thus for any $\phi \in \mathcal{S}(\mathbb{R}^d)$*,*

$$\widehat{1}\left(\widehat{\phi}\right) = 1(\phi) = \int \phi(\xi)\,d\xi = \widehat{\phi}(0).$$

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 $\widehat{u} \in C^\infty(\mathbb{R}^d)$ 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 $f \in L^p(\mathbb{R}^d)$, 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 $f \in L^p(\mathbb{R}^d)$ such that $\widehat{f}$ is* not *a locally integrable function. Otherwise, we could define an operator $T : L^p(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$ given by*

$$Tf = \widehat{f}\mathbf{I}_{|\xi|\leqslant 1}.$$

If a sequence of functions $\{f_n\}$ converges to f in $L^p(\mathbb{R}^d)$, and Tf_n converges to g in $L^1(\mathbb{R}^d)$, then Tf_n converges distributionally to g, which implies $Tf = g$. The closed graph theorem thus implies that T is a continuous operator from $L^p(\mathbb{R}^d)$ to $L^1(\mathbb{R}^d)$, so there exists $M > 0$ such that

$$\int_{|\xi|\leqslant 1} |\widehat{f}(\xi)| \leqslant M\|f\|_{L^p(\mathbb{R}^d)}.$$

If $f_\alpha(x) = e^{-\pi\alpha|x|^2}$, then $\widehat{f_\alpha}(\xi) = \alpha^{-d/2}e^{-\pi|x|^2/\alpha}$. We have

$$\|f_\alpha\|_{L^p(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} e^{-\pi\alpha p|x|^2}\,dx\right)^{1/p}$$

$$= (\alpha p)^{-d/2p}\left(\int_{\mathbb{R}^d} e^{-\pi|x|^2}\,dx\right)^{1/p} \lesssim_d (\alpha p)^{-1/2p}.$$

On the other hand, for $|\xi| \leqslant 1$, $|\hat{f}_\alpha(\xi)| \geqslant \alpha^{-d/2}e^{-\pi/\alpha}$, so

$$\int_{|\xi| \leqslant 1} |\hat{f}_\alpha(\xi)| \gtrsim_d \alpha^{-d/2}e^{-\pi/\alpha}.$$

Thus we conclude that $\alpha^{-d/2}e^{-\pi/\alpha} \lesssim_d M(\alpha p)^{-d/2p}$, or equivalently,

$$\alpha^{d/2(1/p-1)}e^{-\pi/\alpha} \lesssim_d Mp^{-d/2p}.$$

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

$$\hat{g}_\delta(\xi) = \prod_{k=1}^{d} e^{-\pi(1+i\delta)\xi_k^2}.$$

We have

$$\int_{|\xi| \leqslant 1} |\hat{g}_\delta(\xi)| = \int_{|\xi| \leqslant 1} e^{-\pi|\xi|^2} \gtrsim 1.$$

On the other hand, for $\delta \geqslant 1$,

$$\begin{aligned}
\|g_\delta\|_{L^p(\mathbb{R}^d)} &= \left(\int |g_\delta(x)|^p \, dx\right)^{1/p} \\
&= |1 + i\delta|^{-d/2} \left(\int_{-\infty}^{\infty} e^{-p\pi x^2/(1+\delta^2)} \, dx\right)^{d/p} \\
&\lesssim_d \delta^{-d/2}\delta^{d/p}p^{-d/p} = \delta^{d(1/p-1/2)}p^{-d/p}.
\end{aligned}$$

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 $\mathbb{R}^d$, for each $\alpha \in \mathbb{C}$ 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 $p = d/Re(\alpha)$, $K_\alpha \in L^{p,\infty}(\mathbb{R}^d)$. The Marcinkiewicz interpolation theorem implies that if $d/2 < Re(\alpha) < d$, then K_α can be decomposed as the sum of a $L^1(\mathbb{R}^d)$ function and a $L^2(\mathbb{R}^d)$ function, and so we can intepret the Fourier transform of $\widehat{K_\alpha}$ using techniques in $L^1(\mathbb{R}^d)$ and $L^2(\mathbb{R}^d)$, and moreover, the Marcinkiewicz interpolation theorem implies that

$$\|\widehat{K_\alpha}\|_{L^{q,\infty}(\mathbb{R}^d)} \leqslant \|K_\alpha\|_{L^{p,\infty}(\mathbb{R}^d)}.$$

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 $x \in \mathbb{R}^d$, $K_\alpha(\varepsilon x) = \varepsilon^{-\alpha} K_\alpha(x)$. 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 $\mathbb{R}^d$, then

$$\begin{aligned}
\int_{\mathbb{R}^d} K_\alpha(x)e^{-\pi|x|^2} \, dx &= \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}} \int_{\mathbb{R}^d} |x|^{-\alpha}e^{-\pi|x|^2} \, dx \\
&= A_d \frac{\Gamma(\alpha/2)}{\pi^{\alpha/2}} \int_0^\infty r^{d-1-\alpha}e^{-\pi r^2} \, dr \\
&= A_d \frac{\Gamma(\alpha/2)}{2\pi^{d/2}} \int_0^\infty s^{(d-\alpha)/2-1}e^{-s} \, ds \\
&= A_d \frac{\Gamma(\alpha/2)\Gamma((d-\alpha)/2)}{\pi^{d/2}}.
\end{aligned}$$

But this is also the value of

$$\int_{\mathbb{R}^d} K_{d-\alpha}(x)e^{-\pi|x|^2},$$

219

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 $s \in \mathbb{C}$ with $0 < Re(s) < d$, and for each Schwartz $\phi \in \mathcal{S}(\mathbb{R}^d)$ we define

$$A(s) = \int K_s(\xi)\hat{\phi}(\xi) \, d\xi = \frac{\Gamma(s/2)}{\pi^{s/2}} \int |\xi|^{-s/2}\hat{\phi}(\xi) \, d\xi.$$

and

$$B(s) = \int K_{d-s}(\xi)\hat{\phi}(\xi) \, d\xi = \frac{\Gamma((d-s)/2)}{\pi^{(d-s)/2}} \int |\xi|^{(d-s)/2}\hat{\phi}(\xi) \, d\xi.$$

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 $T : \mathbb{R}^d \to \mathbb{R}^d$, 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 $\phi \in \mathcal{S}(\mathbb{R}^d)$,*

$$e^{-i\pi\sigma/4}|\det(T)|^{-1/2} \int_{\mathbb{R}^d} e^{i\pi(T^{-1}\xi \cdot \xi)}\hat{\phi}(\xi) \, d\xi = \int_{\mathbb{R}^d} e^{-i\pi(Tx \cdot x)}\phi(x) \, dx.$$

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

$$\int_{-\infty}^{\infty} e^{-(\pi/z)\xi^2}\hat{\phi}(\xi) \, d\xi = \sqrt{z} \int_{-\infty}^{\infty} e^{-\pi z x^2}\phi(x) \, dx,$$

220

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.

Example. *The* Airy function *on $\mathbb{R}$ is the tempered distribution defined to be the inverse Fourier transform of $e^{2\pi i \xi^3/3}$, i.e.*

$$Ai(x) = \int e^{2\pi i(\xi^3/3 + \xi x)} \, d\xi.$$

In fact, Ai is an analytic function on $\mathbb{R}$. For $\varepsilon > 0$, let $\zeta = \xi + i\varepsilon$. Then

$$2\pi i \zeta^3/3 = (2\pi)\left(i(\xi^3/3 - \varepsilon^2 \xi) - (\varepsilon \xi^2 - \varepsilon^3/3)\right).$$

This shows that for each $\varepsilon > 0$, $e^{2\pi i(\xi + i\varepsilon)^3/3}$ is a tempered distribution. Since

$$e^{2\pi i(\xi + i\varepsilon)^3/3} - e^{2\pi i \xi^3/3} = e^{2\pi i \xi^3/3}\left(e^{2\pi i \varepsilon^2 \xi} e^{-2\pi \varepsilon \xi^2 + (2\pi/3)\varepsilon^3} - 1\right).$$

Thus these functions converge locally uniformly to zero, and are uniformly bounded, and thus $e^{2\pi i(\xi + i\varepsilon)^3/3}$ converges in the sense of tempered distributions to $e^{2\pi i \xi^3/3}$. But since $e^{2\pi i(\xi + i\varepsilon)^3/3}$ decreases exponentially, the Fourier transforms of these distributions are holomorphic in a strip of width $O(\varepsilon)$. TODO: Look up Section 7.6.8 of Hörmander, Volume 1.

Example. *We know $((-2\pi i x)^\alpha)^\wedge = ((-2\pi i x)^\alpha \cdot 1)^\wedge = D^\alpha \delta$, 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 $\hat{\Lambda}$ is $-2\pi i\delta$, where δ is the Dirac delta function at the origin. But this means there exists A such that $\hat{\Lambda}(\xi) = A - 2\pi i \cdot \mathbf{I}(\xi > 0)$. But $\Lambda(-x) = -\Lambda(x)$, implying that $\hat{\Lambda}(-\xi) = -\hat{\Lambda}(\xi)$, and thus $A - 2\pi i = -A$, i.e. $A = i\pi$. Thus*

$$\hat{\Lambda}(\xi) = \pi i - 2\pi i \cdot \mathbf{I}(\xi > 0) = -i\pi \cdot sgn(\xi).$$

Theorem 13.25. *If μ is a finite measure, $\hat{\mu}$ is a uniformly continuous bounded function with $\|\hat{\mu}\|_{L^\infty(\mathbb{R}^d)} \leqslant \|\mu\|$, and*

$$\hat{\mu}(\xi) = \int e(-2\pi i x \cdot \xi) d\mu(x)$$

The function $\hat{\mu}$ is also smooth if μ has moments of all orders, i.e. $\int |x|^k d\mu(x) < \infty$ for all $k > 0$.

Proof. Let $\phi \in \mathcal{S}(\mathbb{R}^d)$. We must understand the integral

$$\int_{\mathbb{R}^d} \hat{\phi}(x) \, d\mu(x).$$

Applying Fubini's theorem, which applies since μ has finite mass, we conclude that

$$\int_{\mathbb{R}^d} \hat{\phi}(x) \, d\mu(x) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(\xi) e^{-2\pi i \xi \cdot x} d\mu(x) \, d\xi = \int_{\mathbb{R}^d} \phi(\xi) f(\xi) \, d\xi,$$

where

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

Thus $\hat{\mu}$ is precisely f, and it suffices to show that $\|f\|_{L^\infty(\mathbb{R}^d)} \leqslant \|\mu\|$, 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,

$$|f(\xi + \eta) - f(\xi)| = \left| \int_{\mathbb{R}^d} e^{-2\pi i \xi \cdot x} (e^{-2\pi i \eta \cdot x} - 1) \, d\mu(x) \right|$$

$$\leqslant \int_{\mathbb{R}^d} |e^{-2\pi i \eta \cdot x} - 1| \, d|\mu|(x).$$

222

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

$$\frac{f(\xi + \varepsilon e_i) - f(\xi)}{\varepsilon} = \int_{\mathbb{R}^d} e^{-2\pi i \xi \cdot x} \frac{(e^{-2\pi \varepsilon i x_i} - 1)}{\varepsilon} d\mu(x).$$

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

$$f_i(\xi) = (-2\pi i) \int_{\mathbb{R}^d} e^{-2\pi i \xi \cdot x} x_i d\mu(x),$$

which is the Fourier transform of $x_i \mu$. Higher derivatives are similar. $\qquad \square$

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 $\phi \in \mathcal{D}(\mathbb{R}^d)$. Moreover, $\Lambda * \phi$ is a slowly increasing function; to see this, we know there is n such that

$$|\Lambda \phi| \lesssim \|\phi\|_{\mathcal{S}^{n,m}(\mathbb{R}^d)}.$$

Now for $|y| \geqslant 1$,

$$\|T_y \phi\|_{\mathcal{S}^{n,m}(\mathbb{R}^d)} \leqslant |x - y|^n \leqslant 2^n(1 + |y|^n)\|\phi\|_{\mathcal{S}^{n,m}(\mathbb{R}^d)},$$

and so

$$(\Lambda * \phi)(x) = \Lambda(T_{-x}\phi^*) \lesssim_n (1 + |x|^n)\|\phi\|_{\mathcal{S}^{n,m}(\mathbb{R}^d)},$$

which gives that $\Lambda * \phi$ is slowly increasing. In particular, we can take the Fourier transform of $\Lambda * \phi$. Now for any $\psi \in \mathcal{S}(\mathbb{R}^d)$ with $\hat{\psi} \in \mathcal{D}(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \widehat{\Lambda * \phi}(\xi)\psi(\xi) \, d\xi = \int_{\mathbb{R}^d} (\Lambda * \phi)(x)\hat{\psi}(x) \, dx$$

$$= \int_{\mathbb{R}^d} \Lambda(\hat{\psi}(x) \cdot T_{-x}\phi^*) \, dx$$

$$= \Lambda \left(\int_{\mathbb{R}^d} \hat{\psi}(x) T_{-x}\phi^* \, dx \right)$$

$$= \Lambda \left(\hat{\psi} * \phi^* \right) = \Lambda \left(\hat{\psi} * \hat{\hat{\phi}} \right)$$

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

223

We therefore conclude that $\widehat{\Lambda * \phi} = \hat{\phi}\hat{\Lambda}$.

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 $\mathbb{R}^d$ which is homogeneous of degree σ is a homogeneous distribution of degree $-d - \sigma$.

Lemma 13.26. *If u is homogeneous and sing-supp(u) $\subset$ {0}, then $\hat{u}$ is homogeneous, and sing-supp($\hat{u}$) $\subset$ {0}.*

Proof. Suppose first that u is homogeneous of order a, with $\mathrm{Re}(a) < -n$. Then we can write $u = u_0 + u_1$, where u_0 is supported in a neighborhood of the origin, and u_1 is an integrable function. But the Fourier transform of both of these terms is continuous. Thus $\hat{u}$ is continuous in this case.

To upgrade this fact, given any homogeneous function u with sing-supp$(u) \subset$ {0}, then $D^\alpha u$ is homogeneous with real part less than $-n$ for sufficiently large α, and sing-supp$(D^\alpha u) \subset$ {0}, which implies that $|\xi|^\alpha \hat{u}$ is continuous, and thus $\hat{u}$ is continuous away from the origin. But since $x^\beta u$ is homogeneous if u is homogeneous, we conclude that $\widehat{x^\beta u} = D^\beta \hat{u}$ is continuous in $\mathbb{R}^n - \{0\}$ for all β. Thus sing-supp$(\hat{u}) \subset$ {0}. $\square$

13.9 Kernel Operators

We have seen that all translation-invariant operators $T : \mathcal{D}(\mathbb{R}^d) \to \mathcal{E}(\mathbb{R}^d)$ are given by convolution by a distribution. Thus convolutions by a distribution are suitably general to represent all continuous translation-invariant operations. Surprisingly, when studying non translation invariant operators T from $\mathcal{D}(\mathbb{R}^d)$ to $\mathcal{E}(\mathbb{R}^d)$, or more generally, from $\mathcal{D}(\Omega_1)$ to $\mathcal{D}(\Omega_2)^*$ for some open subsets Ω_1 and Ω_2 of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, we can obtain a similar characterization. Instead of studying operators of the form

$$T\phi(x) = \int \Lambda(x - y)\phi(y) \, dy$$

we instead study *kernel* operators of the form

$$T\phi(x) = \int K(x, y)\phi(y) \, dy$$

where K is a distribution on $\Omega_2 \times \Omega_1$, acting as a continuous linear operator from $\mathcal{D}(\Omega_1)$ to $\mathcal{D}(\Omega_2)^*$ by defining, for $\psi \in \mathcal{D}(\Omega_2)$,

$$\int \psi(x)(T\phi)(x) \, dx = \int \psi(x)K(x, y)\phi(y) \, dx \, dy,$$

224

i.e. by testing K against $\psi \otimes \phi$. The behaviour of the operator T uniquely determines the distribution K, and we call K the *Schwartz kernel* of T. In 1953, Schwartz found the surprising result that kernels define all continuous operators from $\mathcal{D}(\Omega_1)$ to $\mathcal{D}(\Omega_2)^*$, which explains the prevalence of kernel operators in analysis.

Theorem 13.27. *Let $T : \mathcal{D}(\Omega_1) \to \mathcal{D}^*(\Omega_2)$ be a continuous linear operator. Then there exists a unique distribution $K \in \mathcal{D}^*(\Omega_2 \times \Omega_1)$ such that for $\phi \in \mathcal{D}(\Omega_1)$ and $\psi \in \mathcal{D}(\Omega_2)$,*

$$\int T\phi(x)\psi(x)\ dy = \int K(x,y)\psi(x)\phi(y)\ dx\ dy.$$

Remark. There are many variants of Schwartz's result:

- Every continuous map $T : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^m)^*$ is given by a distribution kernel $K \in \mathcal{S}(\mathbb{R}^m \times \mathbb{R}^n)^*$. Indeed, the vanilla version of Schwartz's theorem applies in this situation to produce a distribution kernel $K \in \mathcal{D}(\mathbb{R}^m \times \mathbb{R}^n)$, and the fact that $\langle K, \phi \otimes \psi \rangle$ is well defined and bilinearly continuous for any $\phi \in SW(\mathbb{R}^n)$ and $\psi \in \mathcal{S}(\mathbb{R}^m)$ implies that K is actually a tempered distribution.

- Any continuous map $T : \mathcal{E}(\Omega_1)^* \to \mathcal{E}(\Omega_2)$ is given by a kernel $K \in C^\infty(\Omega_2 \times \Omega_1)$. Any such kernel certainly induces a continuous map of this form. Conversely, if T is given, then the continuity of the map implies that the distribution kernel K induced by the vanilla Schwartz theorem implies that $K \in C(\Omega_1, \mathcal{E}(\Omega_2))$ since as $y \to y_0$, δ_y converges in the strong topology to δ_{y_0}, and so $T\delta_y$ converges to $T\delta_{y_0}$ in $\mathcal{E}(\Omega_2)$. A similar argument shows that K actually lies in $C^\infty(\Omega_1, C^\infty_{\mathrm{loc}}(\Omega_2))$, where

$$D_x^\alpha D_y^\beta K(x,y) = D^\alpha T\{D^\beta \delta_y\}.$$

  Thus K is smooth.

One could also view the characterization of translation invariant operators as convolution operators as a special case of Schwartz's result. When $\Omega_1 = \Omega_2 = \mathbb{R}^d$ it is sometimes more elegant to work with a representation of kernel operators in *convolution form*, i.e. writing

$$T\phi(x) = \int k(x,z)\phi(x-z)\ dz$$

which is just another way to represent the operators above by considering the change of variables $z = x - y$, i.e. we have $k(x, z) = K(x, x - z)$, or $K(x, y) = k(x, x - y)$. We call k the *Schwartz convolution kernel* associated with the operator T. If $k(z)$ is independent of the x-variable, then we have $T\phi = k * \phi$, i.e. the operator T is really given by convolution.

Proof. Since Schwartz kernels act via a tensor product, understanding the Schwartz kernel theorem requires an understanding of these tensor products. The key property here is that $\mathcal{D}^*$ is *nuclear*. If X and Y are complete locally convex spaces, X is barrelled, and X_b^* is nuclear, then $L_b(X, Y)$ is isomorphic to the topological tensor product

$$X^* \otimes^\wedge Y.$$

But this is *precisely* what the theorem of Schwartz is saying. $\qquad\qquad\square$

Remark. The fact that, e.g. the spaces $L^2(\Omega)$ are *not* nuclear spaces (no infinite dimensional norm space can be nuclear) is one reason why, for instance, there are continuous linear operators $T : L^2(\Omega_1) \to L^2(\Omega_2)$ for which there does not exist a kernel $K \in L^2(\Omega_2 \times \Omega_1)$ such that

$$T f(x) = \int K(x, y) f(y) \, dy,$$

i.e. the class of Hilbert-Schmidt operators does not constitute the class of all continuous operators in the L^2 norm.

Given an operator $T : \mathcal{D}(\Omega_1) \to \mathcal{D}(\Omega_2)^*$ with associated kernel $K(x, y)$, we can consider the transpose operator $T^t : \mathcal{D}(\Omega_2) \to \mathcal{D}(\Omega_1)^*$ given by the kernel $K(y, x)$. Then

$$\langle T\phi, \psi \rangle = \int K(x, y) \psi(x) \phi(y) \, dx \, dy = \langle \phi, T^*\psi \rangle.$$

Thus T^t is formally adjoint to T, in some sense. In particular, if we consider the adjoint $T^* : \mathcal{D}(\Omega_2)^{**} \to \mathcal{D}(\Omega_1)^*$, then $T^*\phi = T^t\phi$ for each $\phi \in \mathcal{D}(\Omega_2)$.

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. A basic result along this form is that an operator T has Schwartz kernel K, then for any test function ϕ, $\text{supp}(T\phi) \subset \text{supp}(K) \circ \text{supp}(\phi)$. Let us consider some more general properties of operator that can be defined in terms of the kernel K of the operator.

Theorem 13.28. *Let* $T : \mathcal{D}(\Omega_1) \to \mathcal{D}(\Omega_2)^*$ *be an operator with Schwartz kernel* $K \in \mathcal{D}(\Omega_2 \times \Omega_1)^*$:

(i) *T maps $\mathcal{D}(\Omega_1)$ into $\mathcal{E}(\Omega_2)$ if and only if $K \in C^\infty(\Omega_2, \mathcal{D}(\Omega_1)^*)$.*

(ii) *T extends to a continuous linear operator from $\mathcal{E}(\Omega_1)^*$ to $\mathcal{D}(\Omega_2)^*$ if and only if $K \in C^\infty(\Omega_1, \mathcal{D}(\Omega_2)^*)$.*

(iii) *K is called* separately regular *if it is an element of*

$$C^\infty(\Omega_1, \mathcal{D}(\Omega_2)^*) \cap C^\infty(\Omega_2, \mathcal{D}(\Omega_1)^*).$$

Then K is completely regular if and only if T maps $\mathcal{D}(\Omega_1)$ into $\mathcal{E}(\Omega_2)$, and extends to a map from $\mathcal{E}(\Omega_1)$ to $\mathcal{D}(\Omega_2)^*$.

(iv) *Suppose $\Omega_1 = \Omega_2$ is the same open set Ω. Then K is called* very regular *if it is separately regular, and agrees with a C^∞ function away from the diagonal*

$$\Delta = \{(x, x) : x \in \Omega\}.$$

If K is very regular, then T is pseudolocal, *i.e. for any compactly supported distribution u, sing-supp$(Tu) \subset$ sing-supp(u).*

To make a remark, an operator $T : \mathcal{D}(\Omega) \to \mathcal{E}(\Omega)^*$ is called local *if supp$(Tu) \subset$ supp(u) for any $u \in \mathcal{D}(\Omega)^*$. It is a result of Peetre that any local operator (continuous or non continuous) is a partial differential operator with coefficients in $\mathcal{D}(\Omega)^*$.*

(v) *We say T is* smoothing, *or* regularizing, *if T extends to a map from $\mathcal{E}(\Omega_2)^*$ to $C^\infty(\Omega_1)$. Then T is smoothing if and only if $K \in C^\infty(\Omega_2 \times \Omega_1)$.*

(vi) *We say T is* proper *if T maps $\mathcal{D}(\Omega_1)$ into $\mathcal{E}(\Omega_2)^*$, and T^t maps $\mathcal{D}(\Omega_2)$ into $\mathcal{E}(\Omega_1)^*$. This holds if and only if for any $\phi \in \mathcal{D}(\Omega_1)$ and $\psi \in \mathcal{D}(\Omega_2)$, $K\phi$ and ψK are in $\mathcal{E}(\Omega_2 \times \Omega_1)^*$, or equivalently, if for any precompact open set $U \Subset \Omega_1$, there is a precompact open set $V \Subset \Omega_2$ such that if supp$(\phi) \subset U$, then supp$(T\phi) \subset V$, and conversely, for each such V, there is U such that if supp$(\psi) \subset V$, then supp$(T^t\psi) \subset U$.*

Under the assumption that both (ii) and (vi) holds, i.e. $K \in C^\infty(\Omega_1, \mathcal{D}(\Omega_2)^*)$ and T is proper, the operator T extends to a map from $\mathcal{D}(\Omega_1)^*$ to $\mathcal{D}(\Omega_2)^*$

Proof. Let us begin with the proof of (i). Suppose $K \in C^\infty(\Omega_2, \mathcal{D}(\Omega_1)^*)$. If $\phi \in \mathcal{D}(\Omega_1)$, then the function

$$u(x) = \int K(x, y)\phi(y) \, dy$$

then lies in $\mathcal{E}(\Omega_2)$, since one calculates that

$$D^\alpha u(x) = \int (D^\alpha K)(x, y)\phi(y) \, dy.$$

Conversely, if T maps $\mathcal{D}(\Omega_1)$ into $\mathcal{E}(\Omega_2)$, then for any ϕ,

$$\int \frac{K(x + \delta e_i, y) - K(x, y)}{\delta}\phi(y) \, dy = \int \Delta_{y_i,\delta}K(x, y)\phi(y) \, dy$$

converges locally uniformly in x to $D^i T\phi(x)$. Thus $\{\Delta_{y_i,\delta}K\}$ converges in the weak $*$ topology to $D^i T$, and the uniform boundedness theorem implies that it actually converges in the strong topology to $D^i T$, and that $D^i T : \mathcal{D}(\Omega_1) \to \mathcal{D}(\Omega_2)^*$ is continuous for all i. The Schwartz kernel of $D^i T$ is clearly the derivative $D^i K$, where K is viewed as a map from $\mathcal{D}(\Omega_1)$ to $\mathcal{D}(\Omega_2)^*$. Iterating this argument shows that $K \in C^\infty(\mathcal{D}(\Omega_1), \mathcal{D}(\Omega_2)^*)$.

Now let us prove (ii). Suppose $K \in C^\infty(\Omega_1, \mathcal{D}(\Omega_2)^*)$. Then (i) implies that T^t maps $\mathcal{D}(\Omega_2)$ into $\mathcal{E}(\Omega_1)$. Taking adjoints shows that the adjoint of T^t is an extension of T to a continuous map from $\mathcal{E}(\Omega_1)^*$ to $\mathcal{D}(\Omega_2)^*$. Conversely, if T extends to a continuous map from $\mathcal{E}(\Omega_1)^*$ to $\mathcal{D}(\Omega_2)^*$, then it is simple to prove that if K is the kernel of T, then $K(x, y) = T\delta_y(x)$, and thus lies in $C^\infty(\Omega_1, \mathcal{D}(\Omega_2)^*)$.

Result (iii) follows immediately from (i) and (ii).

To prove result (iv), we note that if K is very regular, then T maps $\mathcal{D}(\Omega)$ into $\mathcal{E}(\Omega)$. Furthermore, if $\tilde{K} \in C^\infty(X^2 - \Delta)$ is the smooth function agreeing with K away from the origin, then for any $u \in \mathcal{E}(\Omega)^*$, and any $x \notin \operatorname{supp}(u)$,

$$Tu(x) = \int K(x, x')u(x') \, dy = \int \tilde{K}(x, x')u(x') \, dy.$$

The right hand side defines a smooth function of x for $x \in \Omega - \operatorname{supp}(u)$. But this gives us the initial estimate sing-supp$(Tu) \subset \operatorname{supp}(u)$. To prove the full statement, if $u \in \mathcal{E}(\Omega)^*$, and $x_0 \notin$ sing-supp(u), then there is $\phi \in \mathcal{D}(\Omega)$ which equals one in a neighborhood U of x_0, such that $\phi u \in \mathcal{D}(\Omega)$. Thus $T(\phi u) \in \mathcal{E}(\Omega)$. But $\operatorname{supp}((1 - \phi)u)$ does not contain x_0, which implies that x_0 does not lie in

sing-supp$(T((1 - \phi)u))$. But this means that both $T(\phi u)$ and $T((1 - \phi)u)$ are both smooth in a neighborhood of x_0, which means the same is true of Tu. Thus $x_0 \notin$ sing-supp(Tu).

We have already proved (v) in a remark following the statement of the Schwartz kernel theorem.

The first part of the proof of (vi) is left to the reader. The assumptions of (ii) imply T^t maps $\mathcal{D}(\Omega_2)$ to $\mathcal{E}(\Omega_1)$ continuously, and the assumptions of (vi) imply that T^t maps $\mathcal{D}(\Omega_2)$ to $\mathcal{E}(\Omega_1)^*$ continuously. Taking both these properties into account, we conclude that T^t maps $\mathcal{D}(\Omega_2)$ into $\mathcal{D}(\Omega_1)$ continuously. But taking adjoints gives an extension of T as a continous map from $\mathcal{D}(\Omega_1)^*$ to $\mathcal{D}(\Omega_2)^*$. □

The pseudodifferential operators we will study later are very regular, so in particular, pseudolocal. In fact, they are *microlocal*, in the sense that they never expand the *wavefront set* of an operator, a subtle refinement of the singular support of a function.

13.10   Test Functions on a Manifold

For a general smooth manifold M, it is unnatural to define the space of test functions $\mathcal{D}(M)$ to be $C_c^\infty(M)$. This is because there is no natural inclusion map from $C^\infty(M)$ to this dual, given that there is no natural definition of integration on a manifold M. Thus one cannot naturally think of the dual of $C_c^\infty(M)$ as being a space of 'generalized functions' on M.

We remedy this, we work with *scalar densities* rather than scalar-valued functions. Recall that on any manifold M, we can define a line bundle $\mathrm{Vol}(M)$, and scalar densities are sections of this line bundle. 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 $U \cap V$,

$$\omega_y = \omega_x \cdot \left| \det \left(\frac{\partial y^i}{\partial x^j} \right) \right|.$$

We define $\mathcal{D}(M)$ to be the space of all compactly supported smooth scalar densities, which we can equip with a natural locally convex structure analogous to the topology on the space of test functions in Euclidean space. Given any $f \in C^\infty(M)$, and any $\omega \in \mathcal{D}(M)$, the quantity

$$\langle f, \omega \rangle = \int_M f \cdot \omega$$

is well defined, where the integral on the right is defined by working in local coordinates. Moreover, the map $\omega \mapsto \langle f, \omega \rangle$ is then a continuous linear functional on $\mathcal{D}(M)$. Thus we have a continuous inclusion $C^\infty(M) \to \mathcal{D}(M)^*$, which means it makes sense to define $\mathcal{D}(M)^*$ be the space of generalized functions on M.

It sometimes helps to think more generally, given a smooth line bundle E over M, we define $\mathcal{D}(M, E)$ to be the space of all smooth, compactly supported sections of E^*, the *dual* line bundle to E. Then the space $\Gamma(E)$ of all smooth sections of E is naturally included in $\mathcal{D}(M, E)^*$, which we call the space of *generalized sections* of E. As an example, we can consider the line bundle $\mathrm{Vol}^\alpha(TM)$ for each $0 \leqslant \alpha \leqslant 1$, whose sections ω consist of *scalar densities of order α*, which transform in coordinates via the relation

$$\omega_y = \omega_x \cdot \left| \det \left(\frac{\partial y^i}{\partial x^j} \right) \right|^\alpha .$$

Thus $\mathrm{Vol}(TM) = \mathrm{Vol}^1(TM)$, and $\mathrm{Vol}^0(TM) = \varepsilon^1(M)$ is the trivial bundle. Sections of $\mathrm{Vol}^1(TM)$ thus correspond to scalar densities, and sections of $\mathrm{Vol}^0(TM)$ thus correspond to smooth functions. More generally, the dual bundle to $\mathrm{Vol}^\alpha(TM)$ can naturally be identified with $\mathrm{Vol}^{1-\alpha}(TM)$ since if ω is a scalar density of order α, and η is a scalar density of order $1 - \alpha$, then $\omega\eta$ is a scalar density of order one, and thus can be integrated on M in an invariant manner. Thus we can define the space $\mathcal{D}^\alpha(M)^*$ of generalized scalar densities of order α to be the dual to the space of compactly supported scalar densities of order $1 - \alpha$.

Once we fix a volume form on M, all these spaces can be naturally identified with one another. But in the absense of such a measure, when doing L^p estimates, for instance, it is natural work with scalar densities ω of order $1/p$, since then $|\omega|^p$ is a density of order one, and can therefore be integrated on the manifold. And when doing L^2 estimates with spaces of order $1/2$ lead us to consider scalar densities of order $1/2$, we can consider a self-duality theory, because $\mathrm{Vol}^{1/2}(M)$ is self-dual.

Much of the basic theory of distributions follows for distributions on manifolds. In particular, a version of the Schwartz kernel theorem holds, i.e. for any smooth manifolds M and N, and any continuous linear operator $T : \mathcal{D}(M) \to \mathcal{D}(N)^*$, there exists a generalized section K of $\mathrm{Vol}(M) \oplus TN$ such that for any smooth, compactly supported function f on M, we formally have

$$T f(x) = \int K(x, y) f(y) \, dy,$$

in the sense that for any smooth, compactly supported scalar density ω on N,

$$\langle Tf, \omega \rangle = \int \omega(x) K(x, y) f(y) \, dx \, dy.$$

The idea of the proof is to work locally in coordinates, applying the Schwartz kernel theorem in Euclidean space, and then noting that the kernels patch together if we viwe them as sections of $\mathrm{Vol}(M) \oplus TN$.

There is an abstract way of defining $\mathrm{Vol}^\alpha(M)$, which is helpful for defining the pullback in a coordinate independent manner. Given an n-dimensional vector space V, we can define the space $\mathrm{Vol}^\alpha(V)$ to be the space of all homogeneous functions $\rho : \Lambda^n(V) \to \mathbb{C}$ of order α, where we recall $\Lambda^n(V)$ is the space of all alternating tensors. This is a smooth functor, and so can be used to construct vector bundles on manifolds, which is just the space of scalar densities; a scalar density ω, which is given by a smooth function ω_x in a given coordinate system x, has the property that

$$\omega(\lambda(dx^1 \wedge \cdots \wedge dx^n)) = \lambda^\alpha \omega_x.$$

We can then define the pullback of densities by a smooth immersion $F : M \to N$ between two manifolds of the same dimension such that for a scalar density ω on N, and an alternating form η on $\Lambda^n(T_p M)$, we can define $F_* \eta \in \Lambda^n(T_{F(p)} N)$, and then

$$(F^* \omega)(\eta) = \omega(F_* \eta).$$

In coordinates, i.e. for $F : \mathbb{R}^n \to \mathbb{R}^n$, F^* is just given by multiplication by $|\det(F)|^\alpha$. If F is a diffeomorphism, we let F_* denote $(F^{-1})^*$. We can then also define the pullback of distributional scalar densities ω of order α.

Right now, I don't see an elegant way to study differential operators for distributions on manifolds. For a differential operator L on a manifold M, and a distribution u on M, it seems best to compute Lu locally in coordinates.

It is more difficult to see how one would build a canonical definition of the space $\mathcal{S}^*(M)$ of tempered distributions on a manifold M, because there is no natural direction to measure the rate of decay of a function. Thus it is often wiser to do the things one does in Schwartz space (e.g. Fourier-type arguments) by applying a partition of unity so that we may assume our distributions are compactly supported.

13.11 Paley-Wiener Theorem

TODO: See Rudin, Functional Analysis, or Treves, Chapter 29, or Hörmander Vol 1, Section 7.3.

Theorem 13.29. *Let $K \subset \mathbb{R}^n$ be a convex, compact subset with supporting function H. If u is a distribution of order m, supported on K, then the Fourier transform of u is pointwise defined by*

$$\widehat{u}(\xi) = \langle u, e^{-2\pi i \xi \cdot x} \rangle = \int u(x) e^{-2\pi i \xi \cdot x}.$$

Moreover, this integral formula gives an extension of $\widehat{u}$ to an entire function on $\mathbb{C}^n$, and

$$|\widehat{u}(\xi)| \lesssim (1 + |\xi|)^N e^{2\pi H(Im(\xi))}.$$

Conversely, any entire function $f : \mathbb{C}^n \to \mathbb{C}$ satisfying estimates of the form

$$|f(\xi)| \lesssim (1 + |\xi|)^N e^{2\pi H(Im(\xi))}$$

is tempered, and it's inverse Fourier transform is supported on K. In particular, if f satisfies estimates of the form

$$|f(\xi)| \lesssim_N (1 + |\xi|)^{-N} e^{2\pi H(Im(\xi))}$$

for all $N > 0$, then the inverse Fourier transform of f is smooth and compactly supported on K.

Proof. Suppose u is a distribution of order N supported on K. For each $\delta > 0$, define a smooth function χ_δ supported on K_δ, equal to one on $K_{\delta/2}$, and with $\|\partial^\alpha \chi_\delta\| \lesssim_\alpha \delta^{-|\alpha|}$. Then

$$
\begin{aligned}
|\widehat{u}(\xi)| &= \int u(x) \chi_\delta(x) e^{-2\pi i \xi \cdot x} \\
&\lesssim \sup_{|\alpha| \leq N} \|\partial_x^\alpha \{ \chi_\delta e^{-2\pi i \xi \cdot x} \}\| \\
&\lesssim e^{2\pi H(Im(\xi)) + \delta |Im(\xi)|} \sum_{i=0}^{N} \delta^{-i} (1 + |\xi|)^{N-i}.
\end{aligned}
$$

Taking $\delta = (1 + |\xi|)^{-1}$ gives the result.

Conversely, let us begin with the case where f satisfies the decay estimates for all $N > 0$, and show this it's Fourier transform is smooth and compactly supported on K. Then we can define the inverse Fourier transform pointwise via the integral formula

$$u(x) = \int f(\xi) e^{2\pi i \xi \cdot x}.$$

To prove that u is supported on K, we note that we can perform a contour shift, writing

$$u(x) = \int f(\xi + i\eta) e^{2\pi i (\xi + i\eta) \cdot x}.$$

Applying the required decay estimate with $N = n + 1$ gives

$$|u(x)| \leqslant e^{2\pi (H(\eta) - \eta \cdot x)} \int (1 + |\xi|)^{-n-1} \, d\xi$$

Thus if, for a given x, there exists η with $H(\eta) \leqslant \eta \cdot x$, then scaling η gives $u(x) = 0$. But if $H(\eta) \geqslant \eta \cdot x$ for all $\eta \in \mathbb{R}^d$, it follows that $x \in K$.

Now we consider the general case. Consider an entire function f satisfying estimates of the form above for a fixed N, and let u denote it's Fourier transform. Consider a smooth function ϕ supported in the unit ball, with $\int \phi(x) \, dx = 1$, and let $\phi_\delta(x) = \delta^{-n} \phi(x/\delta)$. Then the inverse Fourier transform of $f_\delta(\xi) = f(\xi) \hat{\phi}(\delta \xi)$ is $u * \phi_\delta$. Now because ϕ is supported in a unit ball, it's Fourier transform has the estimates proved above, and so in particular,

$$|f_\delta(\xi)| \lesssim_M (1 + |\xi|)^{N-M} e^{2\pi (H(\text{Im}(\eta)) + \delta |\text{Im}(\xi)|)}.$$

Thus by the last paragraph, we conclude that $u * \phi_\delta$ is supported on K_δ, and taking $\delta \to 0$ completes the proof since $u * \phi_\delta$ converges to u distributionally. $\qquad \square$

13.12 Hyperfunctions

TODO: Treves Chapter 22, and Hörmander 1.3 (Extending Cauchy Data + Convolution Inequality and Quasi-Analytic Functions), 7.3,7.4,7.5,8.4,8.5,8.6,8.7,Chapter 9, (4.1.6-4.1.15), and Section 4.3.

Chapter 14

Microlocal Analysis of Singularities

Suppose u is a distribution on $\mathbb{R}^d$. The *singular support* of u is the set of points $x_0 \in \mathbb{R}^d$ 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 $\phi \in C^\infty(\mathbb{R}^d)$ with $\phi(x) = 1$ in a neighbourhood of some point $x_0 \in \mathbb{R}^d$. 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 $\phi u \in \mathcal{D}(\mathbb{R}^d)$ 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 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 $\phi \in C^\infty(\mathbb{R}^d)$ 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 $N > 0$,

$$|\widehat{u\phi}(\xi)| \lesssim_N |\xi|^{-N}. \tag{14.1}$$

the set of values ξ_0 which do *not* satisfy (14.1) for any choice of bump function ϕ about x_0 forms a closed conical subset of $\mathbb{R}^d$, and we call this the *wavefront* of u

about the singularity x_0. The set

$$\mathrm{WF}(u) = \{(x_0, \xi_0) : \xi_0 \text{ is in the wavefront of } u \text{ at } x_0\}$$

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 wave-front set. This is the beginning of *microlocal analysis*, localization not only in space, but also localization in a conic subset of space / frequency, e.g. in both the x and ξ variable. This chapter will show the advantage of a microlocal mindset.

Remark. The term *wavefront set* was used to bring analogy to the to Huygens theory of wave propogation, i.e. that solutions to the wave equation propogate at a unit velocity in a direction normal to their wavefront. We will later see that if the principal symbol of a constant coefficient linear partial differential operator is real, then the wavefront set of linear partial differential equations is invariant under the bicharacteristic flow of this operator, which in the case of the wave equation, gives Huygen's theory as a special case.

Let us now discuss the wavefront set a little more precisely. If u is a compactly supported distribution on $\mathbb{R}^d$, we define $\Gamma(u)$ to be the set of $\xi_0 \in \mathbb{R}^d$ which have no conical neighbourhood U such that for each $N > 0$ and $\xi \in U$,

$$|\widehat{u}(\xi)| \lesssim_N |\xi|^{-N}. \tag{14.2}$$

A compactness argument justifies that if $\Gamma(u) = \varnothing$, then $u \in C^\infty(\mathbb{R}^d)$.

Lemma 14.1. *If u is a compactly supported distribution and $\phi \in \mathcal{D}(\mathbb{R}^d)$, then*

$$\Gamma(\phi u) \subset \Gamma(u).$$

Proof. Suppose $\xi_0 \notin \Gamma(u)$, so ξ_0 has a conical neighbourhood U such that (14.2) holds. Then there exists $\varepsilon > 0$ such that U contains

$$\left\{ \eta \in \mathbb{R}^d : \frac{\xi_0 \cdot \eta}{|\xi_0||\eta|} \geq 1 - 2\varepsilon \right\}$$

Let V be the conical neighbourhood of ξ_0 defined by setting

$$V = \left\{ \eta \in \mathbb{R}^d : \frac{\xi_0 \cdot \eta}{|\xi_0||\eta|} \geq 1 - \varepsilon \right\}.$$

We claim V satisfies (14.2). Fix $\xi \in V$. Then

$$|\widehat{\phi u}(\xi)| = (\hat{\phi} * \hat{u})(\xi) = \int_{\mathbb{R}^d} \hat{\phi}(\eta)\hat{u}(\xi - \eta) \, d\xi.$$

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 $N > 0$, $\hat{u}(\eta) \lesssim_N 1/(1 + |\eta|)^N$. Since $\phi \in L^\infty(\mathbb{R}^d$, we conclude

$$\int_{|\eta| \leqslant 0.25\varepsilon|\xi|} \hat{\phi}(\eta)\hat{u}(\xi - \eta) \, d\xi \lesssim_\phi \int_{|\eta| \leqslant 0.25\varepsilon|\xi|} \frac{1}{1 + |\xi - \eta|^N}$$

$$\lesssim_{\varepsilon,d} \frac{|\xi|^d}{(1 + 2|\xi|^N)} \lesssim \frac{1}{1 + |\xi|^{N-d}}.$$

On the other hand, since u is compactly supported, $\hat{u}$ is slowly increasing, i.e. there exists $m > 0$ such that

$$|\hat{u}(\xi)| \leqslant 1 + |\xi|^m.$$

Since $\phi \in \mathcal{D}(\mathbb{R}^d)$, we have $|\hat{\phi}(\eta)| \lesssim_M 1/(1 + |\eta|^M)$ for all $M > 0$ and thus we conclude that if $M > m + d$

$$\int_{|\eta| \geqslant 0.25\varepsilon|\xi|} \hat{\phi}(\eta)\hat{u}(\xi - \eta) \lesssim_M \int_{|\eta| \geqslant 0.25\varepsilon|\xi|} \frac{1 + |\xi - \eta|^m}{1 + |\eta|^M}$$

$$\lesssim_{\varepsilon,m} \int_{|\eta| \geqslant 0.25\varepsilon|\xi|} \frac{1 + |\eta|^m}{1 + |\eta|^M}$$

$$\lesssim_{\varepsilon,d} \frac{1}{1 + |\xi|^{M-m-d}}.$$

Choosing the parameter M and N appropriately, we obtain the required bound which shows that $\xi_0 \notin \Gamma(\phi u)$. $\qquad\square$

This fact means that for each distribution u, and any pair of distributions $\phi_1, \phi_2 \in \mathcal{D}(\mathbb{R}^d)$ such that $\mathrm{supp}(\phi_2)$ is contained in the interior of the support of ϕ_1, then $\phi_2/\phi_1 \in \mathcal{D}(\mathbb{R}^d)$, 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 $x \in \mathbb{R}^d$, then we define $\Gamma_x(U)$ to be equal to

$$\bigcap \left\{ \Gamma(\phi u) : \phi \in \mathcal{D}(\mathbb{R}^d), x \in \mathrm{supp}(\phi) \right\}.$$

If $\{\phi_n\}$ is a sequence in $\mathcal{D}(\mathbb{R}^d)$ such that $\mathrm{supp}(\phi_{n+1})$ is compactly supported in $\mathrm{supp}(\phi_n)$ for each n, and if $\bigcap \mathrm{supp}(\phi_n) = \{x\}$, then $\Gamma_x(u) = \lim_{n \to \infty} \Gamma(\phi_n u)$. Finally, we define

$$\mathrm{WF}(u) = \{(x, \xi) : \xi \in \Gamma_x(u)\}.$$

This is the *wavefront set* of u.

Remark. The theory of wavefront sets can also be defined via a sheaf theoretic framework. The distributions on a manifold form 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 $\mathcal{D}^*$. Thus we can consider the quotient sheaf $\mathcal{F} = \mathcal{D}^*/C^\infty$. The support of a distribution in $\mathcal{F}$ is then precisely the singular support of that distribution. Given a manifold M, we can consider a natural sheaf on T^*M. For each open set $U \subset T^*M$, we consider the family of all distributions on M, modulo the space of distributions u with $\mathrm{WF}(u) \cap U = \varnothing$. This is called the *sheaf of microdistributions on M*. The support of a distribution in this sheaf is precisely the wavefront set of the distribution.

Lemma 14.2. *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 $x_0 \in \mathbb{R}^d$. If $(x_0, \xi_0) \notin \mathrm{WF}(u)$ for all $\xi_0 \in \mathbb{R}^d$, then there exists $\phi \in \mathcal{D}(\mathbb{R}^d)$ such that $\phi(x_0) \neq 0$ and $\Gamma(\phi u) = \varnothing$. But this means $\phi u \in \mathcal{D}(\mathbb{R}^d)$, so x_0 is not in the singular support of u. This shows $\pi_x(\mathrm{WF}(u))$ 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 $(x_0, \xi_0) \notin \mathrm{WF}(u)$ for any $x_0 \in \mathbb{R}^d$ since $\Gamma(\phi u) \subset \Gamma(u)$ for any $\phi \in \mathcal{D}(\mathbb{R}^d)$. But if $(x_0.\xi_0) \notin \mathrm{WF}(u)$ for any $x_0 \in \mathbb{R}^d$ we can cover the support of u by a partition of unity $\phi_1, \ldots, \phi_N \in \mathcal{D}(\mathbb{R}^d)$ such that $\xi_0 \notin \Gamma(\phi_i u)$ for each i, and summing up shows $\xi_0 \notin \Gamma(u)$. $\qquad\qquad\square$

Example. *Suppose u is a homogenous distribution which is C^∞ away from the origin. Then $\hat{u}$ is homogenous and C^∞ away from the origin, and we claim that*

$$WF(u) = \{(0, \xi) : \xi \in supp(\hat{u})\} = \{0\} \times \Gamma(u).$$

Since the singular support of u is $\{0\}$, we know $WF(u) \subset \{0\} \times \mathbb{R}^d$, and so it suffices to calculate $\Gamma_0(u)$. Fix a non-negative radial function $\phi \in \mathcal{D}(\mathbb{R}^d)$ such

that $\widehat{\phi}$ is a non-negative Schwartz function with total mass one. Let $\phi_t(x) = \phi(tx)$, let $\psi(\xi) = \widehat{\phi}(\xi)$, and let $\psi_t(\xi) = \widehat{\phi}_t(\xi) = t^{-d}\psi(\xi/t)$. Set $u_t = \psi_t u$. Then

$$\Gamma_0(u) = \lim_{t \to \infty} \Gamma(u_t).$$

If v is the homogeneous distribution given by the Fourier transform of u, and v_t is the Fourier transform of u_t, then

$$v_t = \widehat{\phi_t u} = \psi_t * v.$$

Since $\phi_t u$ is no longer homogeneous, neither are the distributions $\{v_t\}$. But they do satisfy the homogeneity relations

$$v_t(\xi) = (t/r)^s Dil_{t/r} v_r$$

for any $t, r > 0$. In particular, we see from this that $\Gamma(u_t) = \Gamma(u_r)$ for all $t, r > 0$. Thus $\Gamma_0(u) = \Gamma(u_t)$ for any $t > 0$. If v vanishes in a neighborhood of ξ_0, then it vanishes in a conical neighborhood of ξ_0. It thus follows that v_t vanishes in a conical neighborhood of ξ_0 if we pick t to be sufficiently small, and thus $\xi_0 \notin \Gamma(u_t) = \Gamma_0(u)$. Conversely, if $\xi_0 \notin \Gamma_0(u)$, then for all $t > 0$, there is an open cone $U_t \subset \mathbb{R}^d$ containing ξ_0 such that $|v_t(\eta)| \lesssim_{t,N} |\eta|^{-N}$ for all $\eta \in U_t$. The homogeneity relation above implies that we actually have

$$|v_t(\eta)| \lesssim_N t^{N+|s|}|\eta|^{-N}.$$

for all $\eta \in U_1$. As $t \to 0$, $|v_t(\eta)| \to |v(\eta)|$, and if N is taken larger than $|s|$ in the inequality above, this implies that

$$|v(\eta)| \lesssim_N \limsup_{t \to 0} t^{N+|s|}|\eta|^{-N} = 0.$$

Thus $v(\eta) = 0$ for all $\eta \in U_1$, so in particular, $\xi_0 \notin supp(v)$.

Here are some special cases of the above example:

- If δ is the Dirac delta distribution at the origin in $\mathbb{R}^d$, then $\widehat{\delta}$ is the constant function, and so
$$\mathrm{WF}(\delta) = \{0\} \times \mathbb{R}^d.$$

- If $u(x) = \mathrm{p.v}(1/x)$, then $\widehat{u}(\xi) = -i\pi\mathrm{sgn}(\xi)$, and so
$$\mathrm{WF}(u) = \{0\} \times \mathbb{R}.$$

- If $H \subset \mathbb{R}^d$ is a plane, and σ_H is the distribution given by integration against the surface measure on H, then σ_H is homogeneous, and we claim that $\hat{\sigma}_H = \sigma_{H^\perp}$, from which the argument above implies that

$$\Gamma_0(\sigma_H) = H^\perp.$$

  Since σ_H is invariant under translation by elements of H, it follows that for any $x \in H$, $\Gamma_x(\sigma_H) = H^\perp$. But σ_H is C^∞ away from H, so

$$\mathrm{WF}(u) = H \times H^\perp.$$

Thus one sees that the wavefront set is often 'conormal' to a distribution supported on a singular set.

Example. *Let f_0 be a distribution of order $N+1$ defined on an open set $\Omega \subset \mathbb{R}^n$ by taking the boundary values of an analytic function f defined on the set $\Omega \times \Gamma \subset \mathbb{C}^n$, where Γ is a convex open cone in $\mathbb{R}^n$, and $|f(x + iy)| \lesssim |y|^{-N}$. Let*

$$\Gamma^\circ = \{\xi \in \mathbb{R}^n : y \circ \xi \geqslant 0 \text{ for all } y \in \Gamma\}$$

denote the dual cone of Γ. Then $\mathrm{WF}(f_0) \subset \Omega \times (\Gamma^\circ - \{0\})$. To figure this out, we have that for any $\nu \geqslant N$,

$$\widehat{f_0 \phi}(\xi) = \int \phi_{y_0}(x) f_{y_0}(x) e^{-2\pi i(x + iy_0) \cdot \xi} \, dx$$
$$+ \frac{1}{\nu!} \int \int_0^1 f_{ty_0}(x) e^{-2\pi i(x + ity_0) \cdot \xi} \sum_{|\alpha| = \nu + 1} \partial^\alpha \phi(x) (iy_0)^\alpha t^\nu \, dx \, dt.$$

If $y_0 \cdot \xi < 0$, the first quantity decays exponentially as ξ scales, and the second term is $O((1 + |\xi|)^{N-\nu})$. Taking $\nu \to \infty$ yields the claim.

Example. *Let u be a distribution defined on an open set $\Omega \subset \mathbb{R}^n$, and let u^* be the adjoint of u, i.e. defined by setting*

$$\langle u^*, \phi \rangle = \langle u, \phi^* \rangle,$$

where $\phi^*(x) = \overline{\phi(x)}$. Then $\mathrm{WF}(u^*) = \mathrm{WF}(u)^*$, where

$$\mathrm{WF}(u)^* = \{(x, \xi) : (x, -\xi) \in \mathrm{WF}(u)\}.$$

This follows simply from the fact that for $\phi \in \mathcal{D}(\Omega)$, $\widehat{\phi u^*}(\xi) = \widehat{\phi^* u}(-\xi)$.

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 $\phi \in \mathcal{D}(U_0)$, and any $N > 0$,

$$\int_{\mathbb{R}^d} u(x)\phi(x)e^{-2\pi i \lambda \xi \cdot x} \, dx \lesssim_N \langle \xi \rangle^{-N}.$$

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 14.3. *Let u be a distribution, and let $(x_0, \xi_0) \notin WF(u)$. Let U be an open subset of $\mathbb{R}^d$ containing x_0, let V be an open subset of $\mathbb{R}^p$ containing a_0, and let $\psi : U \times V \to \mathbb{R}$ 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 $\phi \in \mathcal{D}(U_0)$, 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 $\phi \in \mathcal{D}(U_0)$, consider $\tilde{\phi} \in \mathcal{D}(U_0)$ with $\tilde{\phi}\phi = \phi$. Then

$$\int u(x)\phi(x)e^{-2\pi i \lambda \psi(x,a)} \, dx = \int u(x)\phi(x)\phi_1(x)e^{-2\pi i \lambda \psi(x,a)} \, dx$$

$$= \int \widehat{u\phi}(\xi) \left(\int \phi_1(x)e^{-2\pi i(\lambda \psi(x,a) - \xi)} \, dx \right) d\xi$$

$$= \lambda^d \int \widehat{u\phi}(\xi) \left(\int \phi_1(x)e^{-2\pi \lambda i(\psi(x,a) - \xi)} \, dx \right) d\xi$$

$$= \lambda^d \int \widehat{u\phi}(\lambda \xi) J(\lambda, \xi, a) \, d\xi.$$

Let $\eta \in \mathcal{D}(\mathbb{R}^d)$ 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

$$J_1(\lambda, \xi, a) = \eta\left(\frac{\xi - \xi_0}{\varepsilon} \right) \int_{\mathbb{R}^d} \phi_1(x)e^{-2\pi \lambda i(\psi(x,a) - \xi)} \, dx$$

and

$$J_2(\lambda, \xi, a) = \left(1 - \eta\left(\frac{\xi - \xi_0}{\varepsilon} \right) \right) \int_{\mathbb{R}^d} \phi_1(x)e^{-2\pi \lambda i(\psi(x,a) - \xi)} \, dx.$$

240

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

$$\left| \int \widehat{u\phi}(\lambda\xi) J_1(\lambda, \xi, a) \, d\xi \right| \lesssim_N \lambda^{-N}.$$

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. $\qquad\square$

For a smooth function $\phi \in \mathcal{D}(V)$ and a smooth diffeomorphism $f : U \to V$, we can define $f^*\phi \in \mathcal{D}(U)$ by setting $f^*\phi(x) = \phi(f(x))$. Then for $\psi \in \mathcal{D}(V)$,

$$\int f^*\phi(x)\psi(x) = \int \phi(f(x))\psi(x) = \int \phi(y)\psi(f^{-1}(y)) \cdot \frac{1}{|f'(f^{-1}(y))|} \, dy.$$

Thus for a distribution u on V, to define a distribution f^*u on U such that for $\psi \in \mathcal{D}(\mathbb{R}^d)$,

$$\int (f^*u)(x)\psi(x) = \int u(y)\phi(f^{-1}(y)) \cdot \frac{1}{|f'(f^{-1}(y))|} \, dy.$$

There is a simple relation between the wavefront set of u and f^*u. We consider $f^* : V \times \mathbb{R}^d \to U \times \mathbb{R}^d$ 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 $V \times \mathbb{R}^d$ and $U \times \mathbb{R}^d$ with the cotangent bundle T^*V and T^*U.

Theorem 14.4. *For any distribution u on V, $WF(f^*u) = f^*(WF(U))$.*

Proof. Assume $(y_0, \eta_0) \notin WF(u)$, let $(x_0, \xi_0) = f^*((y_0, \eta_0))$, and then define $\psi(y, \xi) = f^{-1}(y) \cdot \xi$. Then

$$\nabla_y \psi(y_0, \xi_0) = (f^{-1}(y_0)')^T (\xi_0) = \eta_0.$$

241

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. $\qquad\square$

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.

14.1 Singular Operations on Distributions

A subset Γ of $\Omega \times \mathbb{R}^d$ 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 $\text{WF}(u) \subset \Gamma$, 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 14.5. *Let $\mathcal{U}$ be a family in $\mathcal{D}^*(\Omega)$ such that $\sup_{u \in \mathcal{U}} |u(\phi)| < \infty$ for all $\phi \in \mathcal{D}(\Omega)$. 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. For any compact set $K \subset \Omega$, the assumptions imply there exists m such that for $u \in \mathcal{U}$ and $\phi \in C_c^\infty(K)$,

$$|u(\phi)| \lesssim \|\phi\|_{C^m(K)}.$$

This is a result we have proven earlier following from the uniform boundedness principle. If ϕ is supported on K, then

$$|u(\phi e^{-2\pi i \xi \cdot x})| \lesssim \|\phi e^{-2\pi i \xi \cdot x}\|_{C^m(K)} \lesssim (1 + |\xi|)^m.$$

A similar argument shows the result for a convergent sequence. $\qquad\square$

Theorem 14.6. *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 $\mathcal{D}_\Gamma^*(U)$. $\qquad\square$

Theorem 14.7. $\mathcal{D}(\Omega)$ *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 $N > 0$,

$$\sup_{\xi \in V} \sup_n \lambda^N |\widehat{\psi(u * \phi_n)}(\lambda\xi)| < \infty.$$

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. $\qquad\square$

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 $\psi \in \mathcal{D}(\mathbb{R}^d)$ 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 $\phi, \psi \in \mathcal{D}(\mathbb{R}^d)$, we might expect us to be able to take Fourier transforms, so that

$$\int u(x)v(x)\phi(x)\psi(x)\,dx = \int (\phi u)(x)(\psi v)(x)\,dx$$

$$= \int \int (\widehat{\phi u} * \widehat{\psi v})(\xi)e^{2\pi i \xi \cdot x}\,d\xi\,dx$$

$$= \int \int \widehat{\phi u}(\eta)\widehat{\psi v}(\xi - \eta)e^{2\pi i \xi \cdot x}\,d\xi\,dx.$$

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 14.8. *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* $\mathcal{D}^*_{\Gamma_1}(\Omega) \times \mathcal{D}^*_{\Gamma_2}(\Omega) \rightarrow \mathcal{D}^*_{\Gamma_3}(\Omega)$ *which agrees with multiplication for elements of* $C^\infty(\Omega)$.

Example. *Consider the distributions* Λ_1 *and* Λ_2 *on* $\mathbb{R}^2$, *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 $0_{\mathbb{R}^2}$, *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 $\delta \rightarrow 0$, $\phi_{x,\delta} \rightarrow \Lambda_1$ *and* $\phi_{y,\delta} \rightarrow \Lambda_2$. *We find that*

$$\phi_{x,\delta}\phi_{y,\delta} = (1/4\delta^2)\mathbf{I}(|x| \leqslant \delta, |y| \leqslant \delta).$$

As $\delta \rightarrow 0$, $\phi_{x,\delta}\phi_{y,\delta}$ *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(Λ) = $\{(0, \xi) : \xi < 0\}$. this means WF(Λ) + WF(Λ) 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 $\mathcal{D}^*(\mathbb{R})$, and it requires only a simple calculation to show that Λ is also the limit in $\mathcal{D}_\Gamma^*(\mathbb{R})$, 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 $\psi \in \mathcal{D}(\mathbb{R}^d)$, 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.

246

Theorem 14.9. *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. *Given a smooth, proper map $f : \Omega \to \mathbb{R}^d$, with a regular value 0, then set set N above is disjoint from $\{0\} \times \mathbb{R}^d$, and so we can define the pullback of the Dirac delta function at the origin, $f^*\delta$, also denoted $\delta(f(x))$.*

As examples of this construction, we can consider the distributions $\delta(x)$ and $\delta(y)$ on $\mathbb{R}^2$. This are equal to $\pi_x^*\delta$ and $\pi_y^*\delta$, where $\pi_x : \mathbb{R}^2 \to \mathbb{R}$ and $\pi_y : \mathbb{R}^2 \to \mathbb{R}$ 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 $\delta(x) = \pi_x^*\delta$ is the distribution given by integration on the y-axis. Similarily, one can calculate that $\delta(y) = \pi_y^*\delta$ is the distribution given by integration on the x-axis. It is simple to calculate explicitly, or using the properties of pullback, that

$$WF(\pi_x^*\delta) \subset \{(0, y; \xi, 0) : \xi \neq 0\}$$

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

$$\mathrm{WF}(u_1 \otimes u_2) \subset \left[\mathrm{WF}(u_1) \times \mathrm{WF}(u_2)\right] \cup \left[\mathrm{WF}(u_1) \times \{0\}\right] \cup \left[\{0\} \times \mathrm{WF}(u_1)\right].$$

obtained by isolating each variable separately with a bump function and then tensoring the Fourier transform.

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.

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. One way to resolve this is to consider only the pushforwards of compactly supported distributions. Another way to solve the problem is to assume 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

$$\int f_*(u\phi)(y)e^{-2\pi i\eta \cdot y}\ dy = \int u(x)\phi(x)e^{-2\pi i\eta \cdot f(x)}\ dx.$$

248

We have already show that for such an oscillatory integral, provided that $(f(x_0), Df(x_0)^T \eta) \notin$ WF(u), this integral converges. Thus

$$\text{WF}(f_* u) \subset \{(y, \eta) : \text{There is } (x, \xi) \in \text{WF}(u) \text{ and } Df(x)^T \eta = \xi\}.$$

Now we have defined pushforward, pullback, and tensoring, let us see how they can be used to define useful operations on distributions.

Example. *Let M be a manifold with a C^∞ immersed submanifold M_0. Then we have a smooth inclusion map $f : M_0 \to M$. That f is proper is equivalent to the assumption that M_0 is a closed submanifold of M. If $u \in \mathcal{D}(M_0)^*$, then $f_* u$ is precisely the distribution on M given by $\langle f_* u, \phi \rangle = \langle u, \phi|_{M_0} \rangle$. The wavefront set WF$(\phi_* u)$ is equal to the* conormal bundle *of M_0, i.e. the set*

$$N^*(M_0) = \{(x, \xi) \in T^* M : x \in M_0 \text{ and } f^*(x)(\xi) = 0\}.$$

Let us consider an important example which occurs in the theory of kernel operators. Recall that the Schwartz kernel theorem says that if $T : \mathcal{D}(Y) \to \mathcal{D}^*(X)$ is any continuous linear map, then there exists a distribution $K \in \mathcal{D}^*(X \times Y)$ such that, formally speaking, if $\phi \in \mathcal{D}(Y)$ and $\psi \in \mathcal{D}(X)$,

$$\int T\phi(x)\psi(x) \, dx = \int \psi(x) K(x, y)\phi(y) \, dy.$$

In other words, if $\pi(x, y) = x$ and $\Delta(x, y) = (x, y, y)$, then, unwinding the definition, we find

$$T\phi = \pi_*(\Delta^*(K \otimes \phi)).$$

Now define the *canonical relation*

$$C(K) = \{(x, \xi; y, \eta) : (x, \xi; y, -\eta) \in \text{WF}(K)\},$$

Working through the definition shows that for $\phi \in \mathcal{D}(Y)$, WF$(T\phi)$ is contained in

$$C(K) \circ 0_Y = \{(x, \xi) : (x, \xi; y, 0) \in C(K) \text{ for some } y \in Y\}.$$

We can also use the pushforward equation 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 $\pi_*(\Delta^*(K \otimes u))$ is well defined provided that WF(u) is disjoint from the set

$$0_X \circ C(K) = \{(y, \eta) : (x, 0, y; \eta) \in C(K) \text{ for some } x \in X\}.$$

249

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 such that

$$(0_X \circ C(K^*)) \cap \Gamma = \varnothing.$$

If, in addition, the projection map $\pi(x, y) = x$ is proper on $\mathrm{supp}(K)$, then this can be extended to a sequentially continuous map from $\mathcal{D}^*_\Gamma(\Omega)$ to $\mathcal{D}^*(\Omega)$. Again, working through the definitions shows that

$$\mathrm{WF}(Tu) \subset C(K) \circ \{\mathrm{WF}(u) \cup 0_Y\}.$$

A simple way to remember the results of this construction is that K can be applied to any compactly supported distribution u such that

$$C(K) \circ \{\mathrm{WF}(u) \cup 0_\Omega\}$$

contains no zero vector, and then $\mathrm{WF}(Ku)$ is equal to this composition.

Example. *Consider a pseudodifferential operator T given by a symbol a, i.e.*

$$T\phi(x) = \int a(x,\xi)\hat{\phi}(\xi)e^{2\pi i \xi \cdot x}\, d\xi = \int a(x,\xi)\phi(y)e^{2\pi i \xi \cdot (x-y)}\, d\xi\, dy.$$

We can also think of T as a kernel operator with kernel

$$K(x,y) = \int a(x,\xi)e^{2\pi i \xi \cdot (x-y)}\, d\xi.$$

The kernel is a distribution defined by an oscillatory integral distribution, and our calculations for such distributions show that

$$WF(K) \subset \{(x, -\xi; x, \xi) : x \in \Omega, \xi \in \mathbb{R}^n - \{0\}\}.$$

Thus

$$C(K) \subset \{(x, \xi; x, \xi) : x \in \Omega, \xi \in \mathbb{R}^n - \{0\}\}.$$

In particular, $C(K)$, viewed as a relation, contains no zero vectors, and so T extends to a continuous operator from $\mathcal{E}(\mathbb{R}^d)^*$ to $\mathcal{D}(\mathbb{R}^d)^*$, and if a is compactly supported in the x variable, T extends to a continuous operator from $\mathcal{D}(\mathbb{R}^d)^*$ to itself. For any distribution u, we find that $WF(Tu) \subset WF(u)$. 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, the operator given by convolution by u has kernel*

$$K(x, y) = u(x - y).$$

In other words, $K = f^*u$, where $f : \mathbb{R}^{2d} \to \mathbb{R}^d$ is given by $f(x, y) = x - y$. Since f is surjective, the pullback f^*u is always well defined, and moreover,

$$WF(K) \subset f^*WF(u) = \{(x_1, -\xi, x_2, \xi) : (x_2 - x_1, \xi) \in WF(u)\},$$

and therefore

$$C(K) \subset \{(x + a, \xi; x, \xi) : (a, \xi) \in WF(u)\}.$$

We actually have equality here. To see this, for $a \in \mathbb{R}^d$, and let $g : \mathbb{R}^d \to \mathbb{R}^{2d}$ such that $g(x) = (x + a, a)$. Then $u = g^*K$, and so it follows that

$$WF(u) \subset g^*WF(K) = \{(x, \xi) : (x + a, \xi; a, \eta) \in WF(K)\}.$$

It follows from this that we have equality. Thus the operation $v \mapsto u * v$ is a continuous operator from $\mathcal{E}^*(\mathbb{R}^d)$ to $\mathcal{D}(\mathbb{R}^d)^*$, and

$$WF(u * v) \subset \{(x + a, \xi) : (a, \xi) \in WF(u), (x, \xi) \in WF(v)\}.$$

If u is a distribution with singular support only at the origin, then $C(K)$ is a subset of the diagonal in $T^*\mathbb{R}^d \times T^*\mathbb{R}^d$, and so $WF(u * v) \subset WF(v)$ for any distribution v for which the convolution can be computed.

How about the *composition* of kernel operators? Intuitively, if $A : \mathcal{D}(Y) \to \mathcal{D}^*(X)$ and $B : \mathcal{D}(Z) \to \mathcal{D}^*(Y)$, with kernels $K_A(x, y)$ and $K_B(y, z)$, then, if we could define a kernel operator $C = A \circ B : \mathcal{D}(Z) \to \mathcal{D}^*(X)$, it should have kernel

$$K_C(x, z) = \int K_A(x, y)K_B(y, z) \, dy.$$

This would be possible and well defined, for instance, if the kernel of B lay in $\mathcal{D}(Y \times Z)$. The integral of K_C would be well defined, the composition of operators would be well defined, for if $\phi \in \mathcal{D}(Z)$, then $B\phi \in \mathcal{D}(Y)$, and so $A(B\phi)$ is well defined, and one finds that K_C is the kernel of this operator. To generalize this composition, we might want to define $K_C = \pi_*\Delta^*(K_A \otimes K_B)$, for the appropriate maps Δ and π. Now this operation is well defined provided that $(x, 0; y, \eta) \in C(K_A)$ and $(y, \eta; z, 0) \in C(K_B)$ for some $x \in X$, $y \in Y$, $z \in Z$, and $\eta \in \mathbb{R}^d$, and the projection map $\pi(x, y, z) = (x, z)$ is proper on the support of $\Delta^*(K_A \otimes K_B)$, which

would hold, for instance, if either $\pi(x, y) = x$ was proper on the support of K_A, or $\pi(y, z) = z$ was proper on the support of K_B. In this situation, we have

$$C(K_C) \subset \{C(K_A) \circ C(K_B)\} \cup \{(C(K) \circ 0_Y) \times 0_Z\} \cup \{0_X \times (0_Y \circ C(K_B))\}.$$

Thus we have defined a fairly general composition operator on distributions, which is the unique bilinear continuous extension of the composition of two operators with kernels in $\mathcal{D}(X \times Y)$ and $\mathcal{D}(Y \times Z)$ to the pair $\mathcal{D}^*_{\Gamma_A}(X \times Y)$ and $\mathcal{D}^*_{\Gamma_B}(Y \times Z)$ for any pair of conic sets Γ_A and Γ_B for which there does not exist any $(x, 0; y, -\eta) \in \Gamma_A$ and $(y, \eta; z, 0) \in \Gamma_B$.

Example. *Let us use this construction to define the composition of two pseudodifferential operators P and Q on some domain U, such that one of these operators is proper in the sense above. Indeed, the wave front sets of P and Q are never incompatible, so we can always define $P \circ Q$ using the theory above, and we find that*

$$C(P \circ Q) = C(P) \circ C(Q).$$

In particular, since $C(P)$ and $C(Q)$ are both subsets of the diagonal in $T^*U \times T^*U$, it follows that $P \circ Q$ is also pseudolocal in the sense that $WF((P \circ Q)(u)) \subset WF(u)$. Of course, in the theory of pseudodifferential operators one shows that $P \circ Q$ is also a pseudolocal operator for any two psuedolocal operators P and Q, and then this result follows from that theory.

Example. *Suppose $T : \mathcal{D}(Y) \to \mathcal{D}^*(X)$ is continuous, and thus has a Schwartz kernel $K \in \mathcal{D}^*(X \times Y)$. Then if we consider the operator $T^* : \mathcal{D}(X) \to \mathcal{D}(Y)^*$ induced by the kernel $K^*(y, x) = \overline{K(x, y)}$, then we find that*

$$\langle T\phi, \psi \rangle = \langle \phi, T^*\psi \rangle$$

for any $\phi \in \mathcal{D}(Y)$ and $\psi \in \mathcal{D}(X)$. Thus T^* is the formal adjoint of T. It is simple to verify that

$$C(K^*) = \{(y, \eta; x, \xi) : (x, \xi; y, \eta) \in C(K)\}.$$

Thus if $\pi(x, y) = y$ is a proper map on $\mathrm{supp}(K)$ and $C(K) \circ 0_Y = \varnothing$, we can define $T^* \circ T$, and we have

$$C(T^* \circ T) \subset \{(y_1, \eta_1; y_2, \eta_2) : (x, \xi; y_1, \eta), (x, \xi; y_2, \eta_2) \in C(K)\}.$$

In particular, if $C(K)$ is contained in the graph of a function, i.e. there exists a function f such that

$$C(K) \subset \{(x, \xi; y, \eta) : (y, \eta) = f(x, \xi)\},$$

then $C(K^* \circ K)$ is a subset of the diagonal of $T^*Y \times T^*Y$, and thus we might expect $T^* \circ T$ to behave roughly pseudolocally. This idea comes up in the theory of Fourier integral operators, for then if T is a Fourier integral operator defined by a Lagrangian distribution that is locally a graph of a function, then T^*T will be a pseudodifferential operator.

Chapter 15

Distributional Methods to PDEs

Distribution theory was originally invented to provide a more amenable setting to the theory of existence for linear partial differential equations. Let us use the theory we have now established to solve some differential equations in the language of distributions. We begin with the most basic differential equation, namely solutions to the transport equation $D^i u = 0$.

Theorem 15.1. *If $u \in \mathcal{D}^*(\mathbb{R}^d)$ and there exists an index i such that $D^i u = 0$, then there exists $v \in \mathcal{D}^*(\mathbb{R}^{d-1})$ such that*

$$\int_{\mathbb{R}^d} u(x)\phi(x) \, dx = \int_{\mathbb{R}^{d-1}} v(x) \left(\int_{-\infty}^{\infty} \phi(x) \, dx^i \right) \, dx,$$

i.e. u is 'constant' in the direction i. In particular, if $d = 1$, and $Du = 0$, then u is a constant.

Proof. Suppose without loss of generality that $i = d$. Suppose $\phi \in \mathcal{D}(\mathbb{R}^d)$ and for each $x \in \mathbb{R}^{d-1}$,

$$\int_{-\infty}^{\infty} \phi(x, t) \, dt = 0.$$

Then the function

$$\psi(x, t) = \int_{-\infty}^{t} \phi(x, s) \, ds = 0$$

has compact support and $D^i \psi = \phi$. Thus

$$\int_{-\infty}^{\infty} u(x, t)\phi(x, t) \, dx \, dt = \int_{-\infty}^{\infty} u(x, t) D^i \psi(x, t) \, dx \, dt$$

$$= - \int_{-\infty}^{\infty} D^i u(x, t)\psi(x, t) \, dx \, dt = 0.$$

Now fix $\phi_0 \in \mathcal{D}(\mathbb{R})$ with $\int_{-\infty}^{\infty} \phi_0(x) = 1$. Then given any $\phi \in \mathcal{D}(\mathbb{R}^d)$,

$$\int_{-\infty}^{\infty} u(x,t)\phi(x,t) \, dx \, dt = \int_{-\infty}^{\infty} u(x,t)\phi_0(t) \left(\int_{-\infty}^{\infty} \phi(x,s) \, ds \right) \, dx \, dt.$$

Thus it suffices to set

$$v(x) = \int_{-\infty}^{\infty} u(x,t)\phi_0(t) \, dt. \qquad \square$$

Remark. Applying this result repeatedly shows that if $u \in \mathcal{D}^*(\mathbb{R}^d)$ is a distribution, and $\nabla u = 0$, then u is a constant.

A change of variables gives versions of this result for any transport equation of the form $w \cdot \nabla u = 0$ for a fixed vector $w \in \mathbb{R}^d$. A corollary is a regularity result for distributional solutions to the equation $w \cdot \nabla u(x) + a \cdot u = f$, where $f \in C(\mathbb{R})$, and $a \in C^{\infty}(\mathbb{R})$. We begin with the case $d = 1$.

Lemma 15.2. *If $u \in \mathcal{D}(\mathbb{R})^*$ and $Du + au = f$, for $f \in C(\mathbb{R})$, and $a \in C^{\infty}(\mathbb{R})$, then $u \in C^1(\mathbb{R})$, 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 $F \in C^1(\mathbb{R})$, and $D(u - F) = 0$, so u differs from F by a constant, and is therefore also in $C^1(\mathbb{R})$. For $a \neq 0$, let A be an antiderivative of a, and set

$$E(x) = e^{A(x)}.$$

Then $E \in C^{\infty}(\mathbb{R})$. Thus if u is a distribution solving $Du + au = f(x)$, and if $v = Eu$ then the product rule shows that $Dv = Ef$. The $a = 0$ case implies that $v \in C^1(\mathbb{R})$, and so $u \in C^1(\mathbb{R})$. $\qquad \square$

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 if u is a distribution satisfying the equation

$$D^m u + a_{m-1} D^{m-1} u + \cdots + au = f,$$

for $a_i \in C^{\infty}(\mathbb{R})$, and $f \in C(\mathbb{R})$, then u actually lies in $C^m(\mathbb{R})$, and satisfies this equation pointwise in the classical sense.

Higher dimensional analogues of these results are not as strong. Indeed, we have already seen that distributional solutions to $D^i u = 0$ may fail to be classical solutions 'normal to the direction i'. On the other hand, we can 'almost' obtain such a result if we assume apriori that u is a continuous function.

Lemma 15.3. *Suppose u and f are continuous functions in $C(\mathbb{R}^d)$, and u, viewed as a distribution, satisfies the equation $D^i u = f$. Then $D^i u$ exists pointwise, in the classical sense, for all $x \in \mathbb{R}^d$, and $D^i u(x) = f(x)$ for all $x \in \mathbb{R}^d$.*

Proof. Assume $i = d$ without loss of generality, and write $x = (x_0, t)$, for $x_0 \in \mathbb{R}^{d-1}$ and $t \in \mathbb{R}$. 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 $w \in \mathcal{D}(\mathbb{R}^{d-1})^*$ 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. $\square$

Theorem 15.4. *Fix $a_1, \ldots, a_n \in C^1(\mathbb{R}^d)$, and $b \in C^0(\mathbb{R}^d)$, and suppose that $u \in C(\mathbb{R}^d)$ is pointwise differentiable, and in a pointwise sense, the equation $a_1 D_1 u + \cdots + a_n D_n u + bu = f$, where $f \in C(\mathbb{R}^d)$. Then the same equation holds in a distributional sense (which makes sense because $D_i u$ is a distribution of order one, and thus can be multiplied against C^1 functions).*

Proof. We adapt the proof of Cauchy's theorem due to Goursat. Our goal is to prove, given the assumptions in the theorem, that for any $\phi \in C_c^\infty(\mathbb{R}^d)$,

$$\int (b\phi - D_1(a_1\phi) + \cdots + D_n(a_n\phi)) \cdot u = \int f\phi.$$

For any cube I, let

$$A_I = \int_I (f\phi - (b\phi - D_1(a_1\phi) - \cdots - D_n(a_n\phi)) \cdot u) + \int_{\partial I} u(a \cdot n) \, dS.$$

For any fixed $x_0 \in \mathbb{R}^d$, we claim that

$$\lim_{\substack{x_0 \in I \\ |I| \to 0}} \frac{|A_I|}{|I|} = 0.$$

To prove this, we may replace u in the formula above with it's first order Taylor expansion, and f with the resultant formula $a_1 D_1 u + \cdots + a_n D_n u + bu$, without loss of generality, since the error term here disappears in the limit. But then the formula is easy to verify. But now it follows that we must have $A_I = 0$ for all I, because otherwise we could find a nested family of cubes A_I with $|A_I| \gtrsim |I|$. But if we now take I large enough that it contains the support of ϕ, the theorem is proved. $\qquad\square$

Using techniques from the next section, we obtain a higher dimensional variant of the ODE result above.

Theorem 15.5. *Let $U = V_x \times I_t$, where $V \subset \mathbb{R}^n$ is open, and $I \subset \mathbb{R}$ is an open interval. If $u \in \mathcal{D}(U)^*$ satisfies*

$$\partial_t^m u + L_{m-1}\{\partial_t^{m-1} u\} + \cdots + L_0\{u\} = f,$$

where $f \in C(I, \mathcal{D}(V)^*)$, and $L_0, \ldots, L_{m-1}$ are differential operators with coefficients in $C^\infty(U)$, then $u \in C^m(I, \mathcal{D}(V)^*)$.

Proof. Assuming we are working with vector-valued inputs reduces us to the study of $m = 1$, i.e. an equation of the form $\partial_t u + Lu = f$, and it suffices to show that $u \in C^1(I, \mathcal{D}(V)^*)$. The case $n = 0$ has already been considered above. Localizing if necessary, write $u = \sum_{\beta_1, \beta_2} \partial_t^{\beta_1} D_x^{\beta_2} u_\beta$, where $u_\alpha \in C(U)$, and write $f = \sum_{|\alpha| \leq N} D_x^\alpha f_\alpha$, where $f_\alpha \in C(U)$. If $u_\beta = 0$ for any β with $\beta_1 > 0$, then $u \in C(I, \mathcal{D}(V)^*)$. Otherwise, let M be the smallest integer such that $u_\beta = 0$ for any β with $\beta_1 > M$. Then

$$\partial_t u = f - Lu = \sum D^\alpha f_\alpha - \sum LD^\alpha u_\alpha.$$

But, antidifferentiating, we can write

$$\sum D^\alpha f_\alpha - \sum LD^\alpha u_\alpha = \partial_t \left\{ \sum_\beta \partial_t^{\beta_1} D^{\beta_2} g_\beta \right\},$$

where $g_\beta = 0$ for $\beta_1 > M - 1$. But we have already seen from the fact that these quantities are equal, that

$$u - \sum_\beta \partial_t^{\beta_1} D^{\beta_2} g_\beta$$

lies in $C(I, \mathcal{D}(V)^*)$. Applying induction on M, we conclude $u \in C(I, \mathcal{D}(V)^*)$. But $\partial_t u = u - Lf$ also lies in $C(I, \mathcal{D}(V)^*)$, and so $u \in C^1(I, \mathcal{D}(V)^*)$. $\qquad\square$

Remark. If f extends to an element of $C(\bar{I}, \mathcal{D}(V)^*)$, and u extends to a distribution defined in an open neighborhood of $\bar{I} \times V$, then the proof actually shows that $u \in C^m(\bar{I}, \mathcal{D}(V)^*)$.

We can also consider some Fourier analytic techniques.

Theorem 15.6. *Let P be a polynomial with constant coefficients. For a compactly supported distribution $f \in \mathcal{E}(\mathbb{R}^d)$, the equation*

$$P(D_x)u = f$$

has a compactly supported distributional solution u if and only if the equation

$$\xi \mapsto \frac{\hat{f}(\xi)}{P(\xi)}$$

is entire in ξ. In this case the compactly supported solution is unique, and the convex hull of the support of u is equal to the convex hull of the support of f.

Proof. If a solution u exists, then we obtain that

$$P(\xi)\hat{u}(\xi) = \hat{f}(\xi),$$

where $\hat{u}$ and $\hat{f}$ are entire functions of ξ. It follows that $\xi \mapsto \hat{f}(\xi)/P(\xi)$ is entire. Conversely, suppose that

$$h(\xi) = \frac{\hat{f}(\xi)}{P(\xi)}$$

is entire. Let P be a polynomial of degree m, and, possibly by changing our coordinates, we may assume without loss of generality that the coefficient a of ξ_1^m in P is nonzero. The function

$$z \mapsto P(\xi + ze_1)$$

has leading coefficient a. A maximal-principle argument (see Hormander, Lemma 7.3.3) shows that

$$|ah(\xi)| \leqslant \sup_{|z|=1} |P(\xi + ze_1)h(\xi + ze_1)| = \sup_{|z|=1} |\hat{f}(\xi + ze_1)|.$$

If f is a distribution of order N, compactly supported on some compact set K with supporting function H, then the Paley-Wiener theorem implies we have an estimate of the form

$$|\hat{f}(\xi)| \lesssim (1 + |\xi|)^N e^{H(\mathrm{Im}\,\xi)}.$$

It follows that we also have estimates of the form

$$|h(\xi)| \lesssim (1 + |\xi|)^N e^{H(\text{Im}\,\xi)}.$$

It thus follows that if u is the inverse Fourier transform of h, then u is a distribution of order N supported on K. But then $P(D_x)u = f$. $\qquad\qquad\qquad\square$

Remark. Modifying these calculations somewhat (see Theorem 7.3.8 of Hormander), we can show that the convex hull of the *singular* support of u is equal to the convex hull of the *singular* support of f. In particular, if u is a compactly supported distribution, and $f = P(D_x)u$, then

$$\frac{\widehat{f}(\xi)}{P(\xi)} = \widehat{u}(\xi)$$

is always entire. Thus if u is a compactly supported distribution, and $P(D_x)u$ is smooth, then u is smooth. This is certainly not true for differential operators L with non-constant coefficients. An example is the equation $Lu = xu'$ on the real line; if u is a compactly supported distribution with a 'jump at the origin', i.e. which agrees with $\mathbf{I}(x \geqslant 0)$ near the origin, and is smooth away from the origin, then Lu is smooth, but u is certainly not smooth. TODO: Is this something to do with the microlocal behaviour of constant coefficients operators, i.e. from an analytic perspective I'm not too comfortable with right now?*

One consequence of this result is that one can *separate variables* when solving constant coefficient partial differential equations on convex domains.

Theorem 15.7. *If X is convex, then the span of all solutions to the equation*

$$P(D_x)\phi_{Q,\xi} = 0$$

where $\phi_{Q,\xi}$ is of the form
$$\phi_{Q,\xi}(x) = Q(x)e^{2\pi i \xi \cdot x}$$
for some polynomial Q, and some $\xi \in \mathbb{R}^d$, is dense in the space of all $\phi \in \mathcal{E}(X)$ such that $P(D_x)\phi = 0$.

Proof. Fix a compactly supported distribution $u \in \mathcal{E}(X)^*$, and suppose that

$$\int u(x)\phi_{Q,\xi}(x)\,dx = 0$$

for all $\phi_{Q,\xi}$ such that $P(D_x)\phi_{Q,\xi} = 0$. By the Hahn-Banach theorem, it suffices to show that for any $\phi \in \mathcal{E}(X)$ satisfying $P(D_x)\phi = 0$,

$$\int u(x)\phi(x)\, dx = 0.$$

This will follow if we can show that $h(\xi) = \hat{u}(\xi)/P(-\xi)$ is an entire function of ξ, for then $u = P(-D_x)v$ for some compactly supported function v on X, and the result then follows by an integration by parts. TODO: Lemma 7.3.7 of Hormander.

$\square$

15.1   Fundamental Solutions

Let us now discuss the idea of a *fundamental solution* to a partial differential equation, a technique very useful to the study of the existence and uniqueness of such equations. For a differential operator $L = \sum c_\alpha D^\alpha$ with constant coefficients on $\mathbb{R}^n$, a *fundamental solution* for L is a distribution $\Phi \in \mathcal{D}(\mathbb{R}^n)^*$ such that $L\Phi = \delta$ is the Dirac delta function at the origin. The reason that fundamental solutions are so useful to the study of constant coefficient partial differential equations is that for any compactly supported distribution $v \in \mathcal{D}(\mathbb{R}^d)^*$, the distribution $u = \Phi * v$ is a solution to the equation $Lu = v$, since $Lu = (L\Phi) * v = \delta * v = v$. Thus fundamental solutions give rise to a natural right inverse to a differential operator. It is a general result that *all* constant coefficient differential equations have fundamental solutions, though here we only consider particular examples.

Example. *Consider the differential operator Δ, defining Poisson's equation $\Delta u = v$. To guess a fundamental solution for Δ, we first note that since Δ is invariant under rotations, as is the Dirac delta, the equation $\Delta u = \delta$ is radially symmetric. Thus we might expect to find a radially symmetric fundamental solution. Since Δ is elliptic, we expect fundamental solutions to be smooth away from the origin. Thus to make a guess on the fundamental solution, it will be a useful calculation to determine all smooth, radially symmetric functions $u : \mathbb{R}^n - \{0\} \to \mathbb{C}$ such that $\Delta u = 0$. If $u(x) = f(|x|)$ for $x \neq 0$ and $f \in C^\infty((0, \infty))$, then for $r > 0$,*

$$f''(r) + \frac{n-1}{r}f'(r) = 0.$$

For $n > 2$, this implies that $f(r) = a_1 r^{2-n} + a_2$, and for $n = 2$, $f(r) = a_1 \log(r) + a_2$. We might expect that some choice of constants gives the fundamental solution if we extend these functions to distributions on $\mathbb{R}^n$. We may without

loss of generality pick $a_2 = 0$, since constants do not factor into the output of the operator Δ. The correct choice of the constant a_1 gives the Poisson kernel

$$\Phi(x) = \begin{cases} -\frac{1}{2\pi} \log |x| & : if\ n = 2 \\ -\frac{1}{c_n} \frac{1}{n-2} \frac{1}{|x|^{n-2}} & : if\ n > 2. \end{cases}$$

Here c_n is the surface area of the unit sphere in $\mathbb{R}^n$. These distributions are all locally integrable near the origin and thus extend uniquely to distributions on $\mathbb{R}^n$. To prove that Φ is a fundamental solution, it suffices to show for any $\phi \in \mathcal{D}(\mathbb{R}^d)$,

$$\phi(0) = \int \Phi(x) \Delta\phi(x)\ dx.$$

Since Φ is locally integrable, Gauss' formula implies that as $\varepsilon \to 0$,

$$\int_{|x| \geqslant \varepsilon} \Phi(x) \Delta\phi(x)\ dx = \int_{|x| = \varepsilon} \Phi(x)[\nabla\phi(x) \cdot n(x)]dS - \int \nabla\Phi(x) \cdot \nabla\phi(x)\ dx.$$

Now a simple estimate gives that as $\varepsilon \to 0$,

$$\int_{|x| = \varepsilon} \Phi(x)[\nabla\phi(x) \cdot n(x)]dS \to 0.$$

Thus we conclude that

$$\int \Phi(x) \Delta\phi(x)\ dx = - \int \nabla\Phi(x) \cdot \nabla\phi(x)\ dx.$$

Again, we approximate away from the origin by a ε ball, integrate by parts, and calculate that

$$\int_{|x| \geqslant \varepsilon} \nabla\Phi(x) \cdot \nabla\phi(x)\ dx = \int_{|x| = \varepsilon} \phi(x)\ [\nabla\Phi(x) \cdot n(x)]dS - \int_{|x| \geqslant \varepsilon} \Delta\Phi(x)\phi(x)\ dx$$

$$= \int_{|x| = \varepsilon} \phi(x)\ [\nabla\Phi(x) \cdot n(x)]dS.$$

For $n = 2$, we have $\nabla\Phi(x) = (2\pi)^{-1}(x/|x|^2)$, hence

$$\int_{|x| = \varepsilon} \phi(x)[\nabla\Phi(x) \cdot n(x)]dS = \frac{1}{2\pi\varepsilon} \int_{|x| = \varepsilon} \phi(x)\ dx$$

$$= \phi(0) + O(\varepsilon).$$

261

For $n > 2$, $\nabla\Phi(x) = c_d^{-1}(x/|x|^d)$, so

$$\int_{|x|=\varepsilon} \phi(x)[\nabla\Phi(x) \cdot n(x)]dS = \frac{1}{c_d \varepsilon^{d-1}} \int_{|x|=\varepsilon} \phi(x) \, dS = \phi(0) + O(\varepsilon).$$

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

An alternate approach to obtaining this fundamental solution is to take Fourier transforms, assuming that we can find a tempered fundamental solution Φ. If $\Psi = \widehat{\Phi}$, then we conclude that Φ is a fundamental solution if and only if Ψ is tempered and

$$-4\pi^2|\xi|^2 \cdot \Psi(\xi) = 1.$$

For $n > 2$, we can interpret the formula $\Psi(\xi) = (-1/4\pi^2)|\xi|^{-2}$ as defining a distribution by integration against a locally integrable function. For $n = 2$ a version of this equation remains true provided that we interpret the distribution $1/|\xi|^2$ at the origin in the right way. Thus we have the alternate expression for the fundamental solution above.

Example. *Next, consider the operator $L = \partial_t - \Delta$ on $\mathbb{R}^{n+1}$, which gives rise to the heat equation. Set*

$$\Phi_+(x,t) = \frac{1}{(4\pi t)^{n/2}} \exp\left(-\frac{|x|^2}{4t}\right)$$

for $t > 0$, and $\Phi(x,t) = 0$ for $t \leqslant 0$. Then Φ is locally integrable, and a fundamental solution for L. Indeed, Φ is tempered in the x-variable, and the Fourier transform of Φ in the x variable is verified to be

$$U(\xi,t) = e^{-4\pi^2 t|\xi|^2}$$

which satisfies $\partial_t U = -4\pi^2|\xi|^2 U$, which is equivalent to the equation $L\Phi = \delta$. We see immediately that $\Phi(0+) = \lim_{t\to 0}\Phi(t)$ is the Dirac delta function at the origin.

More generally, similar calculations enable us to find fundamental solutions to any operator of the form $L = \partial_t - S$, where $S = \sum A_{ij}\partial_i\partial_j$. If we assume that there exists a fundamental solution Φ_+ to L, tempered in the x-variable and supported on $t \geqslant 0$, then taking Fourier transforms of the equation $L\Phi = \delta$ in the x-variable, and setting Ψ_+ to be this Fourier transform, we are lead to conclude that $\partial_t\Psi + 4\pi^2(\xi^T A\xi)\Psi = \delta(t)$. Let us assume that Φ is supported on $t \geqslant 0$. Thus we can write

$$\Psi(\xi,t) = H(t)A(\xi)e^{-4\pi^2(\xi^T A\xi)t}$$

for some distribution $A \in \mathcal{D}(\mathbb{R}^n)^*$. But we then calculate that

$$L\Psi = \delta(t)A(\xi)$$

and so we must have $A(\xi) = 1$, i.e.

$$\Psi(\xi, t) = H(t)e^{-4\pi^2(\xi^T A\xi)t}.$$

But this means that, in order to get a tempered fundamental solution we must assumed $Re(A)$ is positive semidefinite. If we, in addition, assume that A is invertible, then we have the Fourier transform for Ψ, namely, we conclude that

$$\Phi(x, t) = \frac{H(t)}{(4\pi t)^{n/2}(\det(A))^{1/2}}e^{-x^T A^{-1}x/4t}.$$

If the real part of A is positive definite, then the same is true for A^{-1}, and so Φ can be interpreted as locally integrable on $\mathbb{R} \times \mathbb{R}^n$. On the other hand, if A is positive semidefinite this is not the case, for instance, the Schrödinger equation $L = \partial_t - i\Delta_x$ has fundamental solution

$$\Phi(x, t) = \frac{H(t)}{(4\pi it)^{n/2}}e^{i|x|^2/4t},$$

which is not locally integrable, but we must treat this solution as a principal value in the t-variable, using the principle of stationary phase to show the limit is well defined, which allows us to conclude Φ is a distribution of order $n + 1$.

Before we consider other quadratic partial differential operators, let us perform a few calculations. Take an operator of the form

$$L_A = \sum_{i,j=1}^{n} A_{ij}D_x^{ij}$$

for some symmetric $n \times n$ matrix A. If T is an invertible linear transformation, then the chain rule implies that

$$D^i\{f \circ T\} = \sum_{k=1}^{n} T_{ki} \cdot (D^k\{f\} \circ T),$$

i.e. $\nabla\{f \circ T\} = T^t\{(\nabla f) \circ T\}$. Iterating the chain rule, we find that

$$D^{ij}\{f \circ T\} = \sum_{k=1}^{n}\sum_{l=1}^{n} T_{lj}T_{ki} \cdot (D^{kl}\{f\} \circ T).$$

263

Thus the Hessian is $H\{f \circ T\} = T^t \cdot \{Hf \circ T\} \cdot T$. This means that if T^*L_A is the operator given by the equation $T^*L_A\{f\} = L_A\{f \circ T\} \circ T^{-1}$, then

$$T^*L_A f = (L_{TAT^t} f) \circ T.$$

The use of these pullbacks is that if $f \in C^\infty(\mathbb{R}^n)$, and if $T^*f = f \circ T^{-1}$, then

$$T^*\{L_A f\} = L_A f \circ T^{-1} = L_A\{T^*f \circ T\} \circ T^{-1} = (T^*L_A)\{T^*f\}.$$

We can take several important points away from this result:

- $T^*L = L$ precisely when $TAT^t = A$, which is equivalent to the condition that if $B_A(x,y) = x^t Ay$ is the symmetric bilinear form specified by A, then $B_A(Tx, Ty) = B(x,y)$. The family of such linear maps from a *generalized orthogonal group*.

- Sylvester's law of inertia implies we can write any symmetric $n \times n$ matrix A as $T^{-1}BT^{-t}$, where B is a diagonal matrix consisting of zeroes, ones, and negative ones. But this means that

$$T^*L = \sum B_{ii}(D^{ii}\{f\} \circ T).$$

Thus the theory of a real-coefficient order two differential equation with constant coefficients is equivalent to one of the form

$$Lf = \sum_{1 \leqslant i < n_1} \frac{\partial^2 f}{\partial x_i^2} - \sum_{n_1 \leqslant i < n_2} \frac{\partial^2 f}{\partial x_i^2},$$

where $n_1 + n_2 \leqslant n$. This simplifies our calculations considerably, e.g. the theory of an operator L given by a positive definite matrix A, i.e. one which has signature $(n, 0)$, is equivalent to the theory of the Laplacian Δ. The theory of L given by a $(n+1) \times (n+1)$ matrix with signature $(1, n)$ is equivalent to the theory of the wave operator $\square = \partial_t^2 - \Delta_x$, and so on and so forth.

Inspired by these calculations, we now find a fundamental solution to any quadratic partial differential operator $L = \sum B_{ij}\partial^{ij}$, such that B is invertible. Abusing notation, we identify a matrix A with the function given by it's associated quadratic form, i.e. the function $x \mapsto x^T Ax$.

Lemma 15.8. *Let A be a real, invertible matrix with signature* (n_+, n_-). *If c_n is the surface area of the unit sphere in $\mathbb{R}^n$, then for $n > 2$, if we let $B = A^{-1}$, and*

$$L_B = \sum B_{ij} D^{ij},$$

then $L_B\{(A \pm i0)^{1-n/2}\} = -(n-2)c_n |\det(A)|^{-1/2} e^{\mp i\pi n_-/2} \delta_0$.

Proof. By conjugation, it suffices to show that

$$L_B\{(A + i0)^{1-n/2}\} = -(n-2)c_n |\det A|^{-1/2} e^{-i\pi n_-/2} \delta_0.$$

We will begin by showing that if A is a symmetric matrix with complex coefficients such that $\mathrm{Re}(A)$ is positive definite, and if B is the inverse of A, then

$$L_B\{A^{1-n/2}\} = -(n-2)c_n (\det A)^{-1/2} \delta$$

where the determinant power is defined to be the unique analytic branch such that $(\det A)^{-1/2} > 0$ if A is positive definite. Analytic continuation means we only need to prove this formula if A, and thus B, is positive definite. But then we can find a matrix S such $S B S^t = I$, and thus $S^{-t} A S^{-1} = I$. But then

$$
\begin{aligned}
S^* \{L_B\{A^{1-n/2}\}\} &= (S^* L_B)\{S^* A^{1-n/2}\} \\
&= L_I\{|x|^{2-n}\} \\
&= \Delta\{|x|^{2-n}\} \\
&= -(n-2)c_n \delta_0.
\end{aligned}
$$

The result then follows because

$$(S^{-1})^* \delta_0 = \det(S)^{-1} \delta_0 = \det(A)^{-1/2}.$$

To obtain the general result, we now note that if

$$A_\varepsilon = \varepsilon - iA \quad \text{and} \quad B_\varepsilon = (\varepsilon - iA)^{-1}.$$

Then

$$\det(A_\varepsilon)^{-1/2} = |\det(A)|^{-1/2} e^{i\pi \mathrm{sgn}(A)/4},$$

and $B_\varepsilon \to iB$ and $A_\varepsilon^{1-n/2} \to i^{-1} e^{i\pi n/4}(A + i0)^{1-n/2}$, so that this gives the result in the limit. $\qquad \square$

We would hope to define a fundamental solution to the equation L_B as the pullback

$$A^* \chi_+^{1-n/2}.$$

This is not quite possible using the spectral analysis of singularities since A is singular at the origin. But provided A is nonsingular, we can define the pullback viewing A as a map from $\mathbb{R}^n - \{0\} \to \mathbb{R}$, thus giving us a distribution away from the origin. Since $A^* \chi_+^{1-n/2}$ is then a homogeneous distribution of degree $2 - n$ on $\mathbb{R}^n - \{0\}$, this distribution then uniquely extends to a distribution on $\mathbb{R}^n$, and we claim this is a constant multiple of a fundamental solution.

Corollary 15.9. *Using the setup to the previous lemma,*

$$L_B\{A^* \chi_\pm^{1-n/2}\} = \pm 4\pi^{n/2-1} \sin(\pi n_\pm/2) |\det A|^{-1/2} \delta_0.$$

Proof. The result follows from the previous result given that

$$\chi_+^a(s) = \frac{i}{2\pi} \Gamma(-a)((s - i0)^a e^{i\pi a} - (s + i0)^a e^{-i\pi a})$$

for all a that is not a positive integer. To obtain this, we use the fact that $\Gamma(a + 1)\Gamma(-a) = -\pi/\sin(\pi a)$, which means that

$$\begin{aligned}
\chi_+^a(s) &= \frac{1}{\Gamma(a + 1)} s_+^a \\
&= -\frac{1}{\pi} \Gamma(-a) \sin(\pi a) \cdot s_+^a \\
&= \frac{i}{2\pi} \Gamma(-a)(e^{i\pi a} - e^{-i\pi a}) s_+^a.
\end{aligned}$$

For $\operatorname{Re}(a) > 0$, s_+^a is locally integrable, and

$$(s \pm i0)^a = \begin{cases} s^a & : s \geqslant 0, \\ |s| e^{\pm i\pi a} & : s < 0. \end{cases}$$

Thus we conclude that

$$(e^{i\pi a} - e^{-i\pi a}) s_+^a = e^{i\pi a}(s - i0)^a - e^{-i\pi a}(s + i0)^+.$$

The general case follows by analytic continuation. $\qquad\square$

Example. *Consider the d'Alembertian, or wave operator* $\Box f = \partial_t^2 f - \Delta_x f$. *The fundamental solutions for this equation are significantly richer than for the Laplace equation. There is no single preferred fundamental solution in this setting. Physical intuition tells us that waves travel at a* finite *speed of propagation, in this case, at a unit velocity. Thus we expect to find a fundamental solution supported on the interior of the* light cone

$$\Sigma = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : Q(x,t) = 0\},$$

i.e. supported on $\Sigma_+ = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : Q(x,t) \geqslant 0\}$, *where* $Q(x,t) = t^2 - |x|^2$. *In fact, for* $n \geqslant 2$, *we have already done the calculations to find such a fundamental solution, namely we have a fundamental solution of the form*

$$\Phi = \frac{1}{4\pi^{\frac{n-1}{2}}} Q^* \left(\chi_+^{-\frac{n-1}{2}} \right).$$

For $n = 1$, *this formula actually continues to hold, namely, we have*

$$\Phi(x,t) = (1/2)\mathbf{I}((x,t) \in \Sigma_+) = (1/2)H(t-x)H(t+x),$$

though we have to make do with more rudimentary calculations. These choices give fundamental solutions to the wave equation. To verify the equation gives a fundamental solution to the wave operator for $n = 1$, *we calculate that*

$$\begin{aligned}
\partial_t \Phi(x,t) &= (1/2)\delta(t-x)H(t+x) + (1/2)\delta(t+x)H(t-x) \\
&= (1/2)\left(\delta(t-x) + \delta(t+x)\right)H(t),
\end{aligned}$$

and so

$$\begin{aligned}
\partial_t^2 \Phi(x,t) &= (1/2)\left(\delta(t-x) + \delta(t+x)\right)\delta(t) + (1/2)\left(\delta'(t-x) + \delta'(t+x)\right)H(t) \\
&= \delta(x,t) + (1/2)\left(\delta'(t-x) + \delta'(t+x)\right)H(t).
\end{aligned}$$

Next, we calculate that

$$\begin{aligned}
\partial_x \Phi(x,t) &= (1/2)H(t-x)\delta(t+x) - (1/2)H(t+x)\delta(t-x) \\
&= (1/2)\left(\delta(t+x) - \delta(t-x)\right)H(t),
\end{aligned}$$

and thus

$$\begin{aligned}
\partial_x^2 \Phi(x,t) &= (1/2)\left(\delta(t+x) - \delta(t-x)\right)\delta(t) + (1/2)\left(\delta'(t+x) + \delta'(t-x)\right)H(t) \\
&= (1/2)\left(\delta'(t+x) + \delta'(t-x)\right)H(t).
\end{aligned}$$

267

Thus $\Box\Phi(x, t) = \delta(t, x)$. In particular, we note that when n is odd, then the support of

$$\chi_+^{-\frac{n-1}{2}}$$

is equal to $\{0\}$. This means that when $n \geq 3$ is odd, then Φ is actually supported *on Σ, which hints at* Huygen's principle, *i.e. that in odd dimensions, if u is a solution to the Cauchy problem $\Box u = 0$ with initial conditions u_0, then the behaviour of u at a point (t, x) is determined by the values of u_0 on a sphere of radius t around x.*

There are two important alternate fundamental solutions, the forward *and* backward, *or* advanced *and* retarded *fundamental solution*

$$\Phi_+(x, t) = 2H(t)\Phi(x, t) \quad and \quad \Phi_-(x, t) = 2H(-t)\Phi(x, t).$$

That these two distributions are also fundamental solution follows from the symmetry of all objects involved under time reflection. These are actually the unique *fundamental solutions supported on the forward and backward light cone and their interior. Indeed, if $u \in \mathcal{D}^*(\mathbb{R}^n \times \mathbb{R})$ is supported on the forward light cone and $\Box u = 0$, then*

$$u = (\Box\Phi_+) * u = \Phi_+ * \Box u = 0,$$

where the convolution is well defined because the projection

$$((x, t), (y, s)) \mapsto (x + y, t + s)$$

is then proper on $supp(u) \times supp(\Phi_+)$.

The wave equation plays nicely with the Fourier transform. If Φ is any fundamental solution, and we assume that it is tempered in the x-variable (which is the case for the distributions we consider above), supported on Σ, then applying the Fourier transform in this variable, i.e. setting $\Psi = \mathcal{F}_x\Phi$, then

$$\partial_t^2\Psi(\xi, t) + 4\pi^2|\xi|^2\Psi(\xi, t) = \delta(t)$$

This implies that, for $t > 0$, we have

$$\Psi(\xi, t) = A_+(\xi)e^{2\pi i|\xi|t} + B_+(\xi)e^{-2\pi i|\xi|t},$$

and a similar formula, allowing different functions, for $t < 0$. Because Ψ is smooth away from the boundary of the light cone, this implies that A_+ and B_+ are both smooth functions. We calculate that

$$\partial_t\Psi(\xi, 0) = 2\pi i|\xi|[A_+(\xi) - B_+(\xi)]$$

If we assume that Ψ is supported on $t \geqslant 0$, and we denote it by Ψ_+, then for any $\phi \in \mathcal{D}(\mathbb{R}^n \times \mathbb{R})$, an integration by parts shows that

$$\int_0^\infty \int_{\mathbb{R}^n} (\partial_t^2 \phi(\xi, t) + 4\pi^2 |\xi|^2 \phi(\xi, t)) \Psi_+(\xi, t)$$
$$= -[A_+(\xi) + B_+(\xi)] \partial_t \phi(\xi, 0) + 2\pi i |\xi| [A_+(\xi) - B_+(\xi)] \phi(\xi, 0).$$

Thus to get a fundamental solution, we should have $A_+(\xi) + B_+(\xi) = 0$, and $A_+(\xi) - B_+(\xi) = 1/2\pi i |\xi|$. Thus we conclude that

$$\Psi_+(\xi, t) = \frac{\sin(2\pi |\xi| t)}{4\pi |\xi|}.$$

This implies that there is a unique *fundamental solution supported on $t > 0$ and tempered in the x-variable, i.e. the fundamental solution Φ_+ we define above, and moreover, we have the Fourier representation*

$$\Phi_+(x, t) = \frac{H(t)}{2} \int \frac{\sin(2\pi |\xi| t)}{2\pi |\xi|} e^{2\pi i \xi \cdot x} \, d\xi.$$

A similar formula holds for the other fundamental solutions we constructed above, namely

$$\Phi_-(x, t) = \frac{H(-t)}{2} \int \frac{\sin(-2\pi |\xi| t)}{2\pi |\xi|} e^{2\pi i \xi \cdot x} \, d\xi,$$

and

$$\Phi(x, t) = \frac{1}{4} \int \frac{\sin(2\pi |\xi t|)}{2\pi |\xi|} e^{2\pi i \xi \cdot x} \, d\xi.$$

However, the Fourier calculation above yields some other interesting fundamental solutions, such as the Feynman fundamental solution

$$\Phi_F(x, t) = \frac{1}{4i} \int \frac{1}{2\pi |\xi|} e^{2\pi i (\xi \cdot x + |\xi t|)} \, d\xi$$

One quirk of this fundamental solution is that it is not *contained in the light cone, despite the finite speed of propogation of the d'Alembertian. In fact, $\operatorname{supp}(\Phi_F) = \mathbb{R} \times \mathbb{R}^n$. To see this, let $u(t) = \Phi_F(0, t)$ and $v(x) = \Phi_F(x, 0)$. That $\operatorname{supp}(\Phi_F) = \mathbb{R} \times \mathbb{R}^n$ will follow from the symmetry of Φ_F under the Lorentz transformation, and that $\operatorname{supp}(u) = \mathbb{R}$, $\operatorname{supp}(v) = \mathbb{R}^n$. To obtain this, we can explicitly calculate that*

for $n = 1$, $v(x)$ is a multiple of $\text{sgn}(x)$, and for $n > 1$, $v(x)$ is a multiple of $|x|^{1-n}$. In both cases, $\text{supp}(v) = \mathbb{R}^n$. Next, we calculate that

$$u(t) = \frac{1}{4i} \int \frac{e^{2\pi i |\xi| t}}{2\pi |\xi|} \, d\xi = \frac{A_{n-1}}{4i} \int_0^\infty \tau^{n-2} e^{2\pi i \tau t} \, d\tau.$$

Thus u is the Fourier transform of a homogeneous function of order $n - 2$ in the τ variable, and is thus a homogeneous function of order $1 - n$ in the t variable. Applying time reversal symmetry, and the fact that $u(t)$ is not supported at the origin (since it's Fourier transform is not a polynomial) imply that $\text{supp}(u) = \mathbb{R}$.

TODO: Move to Cauchy Problem For the study of the Cauchy problem for the wave operator, it is useful to study the behaviour of Φ_+ for $t > 0$. Since $\partial_t^2 \Phi - \Delta_x \Phi = 0$, it follows from the general theory of distributional solutions to ODEs that $\Phi \in C^2([0, \infty), \mathcal{D}(\mathbb{R}^n)^*)$. In particular, if we consider the distributions $\Phi(t)$ on $\mathcal{D}(\mathbb{R}^n)^*$, then both $\Phi(0+) = \lim_{t \to 0^+} \Phi(t)$ and $\partial_t \Phi(0+) = \lim_{t \to 0^+} \partial_t \Phi(t)$ are well defined. Moreover, for $\phi \in \mathcal{D}(\mathbb{R}^{n+1})$ we have the representation formula

$$\langle \Phi_+, \phi \rangle = \int_0^\infty \langle \Phi_+(t), \phi(t) \rangle \, dt,$$

since both sides of the equation define homogeneous distributions of order $1 - n$ which agree for $\phi \in \mathcal{D}(\mathbb{R}^{n+1} - \{0\})$, and thus agree for all $\phi \in \mathcal{D}(\mathbb{R}^{n+1})$. But integration by parts implies that

$$\begin{aligned}
\phi(0) &= \langle \Box \Phi_+, \phi \rangle \\
&= \int_0^\infty \langle \Phi_+(t), \Box \phi(t) \rangle \, dt \\
&= \int_0^\infty \langle \Phi_+(t), \partial_t^2 \phi(t) \rangle - \langle \Phi_+(t), \Delta \phi(t) \rangle \, dt \\
&= \langle \partial_t \Phi_+(0+), \phi(0) \rangle - \langle \Phi_+(0+), \partial_t \phi(0) \rangle.
\end{aligned}$$

Since this is true for arbitrary $\phi \in \mathcal{D}(\mathbb{R}^{n+1})$, we conclude that $\Phi_+(0+) = 0$ and $\partial_t \Phi_+(0+) = \delta$. For $t > 0$, we can compute a formula for the distribution Φ_+. The map $(t, x) \mapsto (t^2 - |x|^2, x)$ then has an inverse

$$H(s, x) = ((s + |x|^2)^{1/2}, x).$$

Thus $|\det(DH)| = (1/2)(s + |x|^2)^{-1/2}$. If $\phi \in \mathcal{D}(\mathbb{R}^n \times \mathbb{R})$ is supported on $t > 0$, and we set

$$\langle \Phi_+, \phi \rangle = \frac{1}{4\pi^{\frac{n-1}{2}}} \langle \chi_+^{-\frac{n-1}{2}}, \psi \rangle,$$

270

where

$$\psi(s) = \int \phi(H(s, x))(s + |x|^2)^{-1/2} \, dx.$$

If we set

$$\tilde{\phi}(t, r) = \frac{1}{r} \int_{|x|=r} \phi(t, x) \, dx,$$

then

$$\psi(s) = \int_{s^{1/2}}^{\infty} \tilde{\phi}(t, (t^2 - s)^{1/2}) \, dt$$

If we consider the compactly supported distribution $\Phi_+(t)$ on $\mathbb{R}^n$, then for $\eta \in C^\infty(\mathbb{R}^n)$,

$$\langle \Phi_+(t), \eta \rangle = \frac{1}{4\pi^{\frac{n-1}{2}}} \int_0^{t^2} \chi_+^{-\frac{n-1}{2}}(s) \tilde{\eta}((t^2 - s)^{1/2}) \, ds,$$

where

$$\tilde{\eta}(r) = \frac{1}{r} \int_{|x|=r} \eta(x) \, dx.$$

For $n = 1$, we have $\chi_+^0(s) = H(s)$, and

$$\Phi_+(t) = (1/2)\mathbf{I}_{[-t,t]}.$$

For $n = 2$, we have $\chi_+^{-1/2} = x_+^{-1/2} \pi^{-1/2}$, and so

$$\Phi_+(t) = (1/2\pi) \cdot \max(t^2 - |x|^2, 0)^{-1/2}.$$

For odd values $n = 2m + 1$, we have $\chi_+^{-\frac{n-1}{2}} = \chi_+^{-m} = D^{m-1}\delta_0$, and this leads to the conclusion that

$$\langle \Phi_+(t), \eta \rangle = \frac{1}{4\pi^m} \frac{d^{m-1}}{ds^{m-1}} \left(\tilde{\eta}(s^{1/2}) \right) \Big|_{s=t^2}.$$

Thus this value depends on the values of η, and their normal derivatives up to order $m - 1$, on a sphere of radius t. In particular, for $n = 3$, we have

$$\langle \Phi_+(t), \eta \rangle = \frac{1}{4\pi} \tilde{\eta}(t) = \frac{1}{4\pi t} \int_{|x|=t} \eta(x) \, dx.$$

271

TODO: Move to Cauchy Problem The existence and uniqueness of solutions to the Cauchy problem for the wave equation follow immediately from this calculation. Namely, if $\phi_0, \phi_1 \in C^\infty(\mathbb{R}^n)$, and $f \in C^\infty(\mathbb{R}^n \times [0, \infty))$, then there exists a unique $u \in C^\infty(\mathbb{R}^n \times [0, \infty))$ such that $\square u = f$ on $\mathbb{R}^n \times [0, \infty)$, and we have

$$u(t) = \Phi_+(t) * \phi_1 + (\partial_t \Phi_+)(t) * \phi_0 + \int_0^t (\Phi_+(t-s) * f(s)) \, ds.$$

Uniqueness follows the fact the fact that if u is a distribution supported on the forward light cone and $\square u = 0$, then $u = 0$. Conversely, the equation above certainly defines an element of $C^\infty([0, \infty) \times \mathbb{R}^n)$ and gives a solution since

$$\square(\Phi_+ * \phi_1) = \delta * \phi_1 = \phi_1$$

$$\square(\partial_t \Phi_+ * \phi_0) = \partial_t(\square \Phi_+ * \phi_0) = \partial_t \phi_0 = 0$$

and since $\Phi_+(0) = 0$, and $\partial_t \Phi_+(0) = \delta$, we find that

$$\partial_t \left\{ \int_0^t (\Phi_+(t-s) * f(s))) \right\} = \int_0^t (\partial_t \Phi_+(t-s) * f(s))$$

and so

$$\partial_t^2 \left\{ \int_0^t (\Phi_+(t-s) * f(s)) \, ds \right\} = f(t) + \int_0^t (\partial_t^2 \Phi_+(t-s) * f(s)).$$

Since $\square \Phi_+$ vanishes away from the origin,

$$\square \left\{ \int_0^t (\Phi_+(t-s) * f(s)) \, ds \right\} = f(t),$$

which completes the proof.

Example. *Consider the Cauchy-Riemann equations on $\mathbb{R}^2$, i.e. the operators $\partial_z = (1/2)(\partial_x - i\partial_y)$ and $\partial_{\bar{z}} = (1/2)(\partial_x + i\partial_y)$. The divergence theorem tells us that*

$$\int_\Omega \frac{\partial \phi}{\partial z} \, dx \, dy = \frac{1}{2} \int_{\partial \Omega} \phi \, d\bar{z}$$

and

$$\int_\Omega \frac{\partial \phi}{\partial \bar{z}} \, dx \, dy = \frac{1}{2} \int_{\partial \Omega} \phi \, dz$$

272

The variant of Green's formula then tells us that

$$\int_\Omega \left(u \frac{\partial v}{\partial z} + \frac{\partial u}{\partial z} v \right) \, dx \, dy = \frac{1}{2} \int_{\partial\Omega} uv \, d\bar{z}.$$

Thus the operators ∂_z and $\partial_{\bar{z}}$ are self adjoint. In particular, applying this formula gives that for any $\phi \in \mathcal{D}(\mathbb{R}^2)$,

$$\int_{|z| \geqslant \varepsilon} \frac{\partial \phi}{\partial \bar{z}} \left(\frac{1}{\pi i} \frac{1}{z} \right) \, dx \, dy = \frac{1}{2\pi i} \int_{|z|=\varepsilon} \frac{\phi}{z} \, dz$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \phi(\varepsilon e^{i\theta}) \, d\theta$$

$$= \phi(0) + O(\varepsilon).$$

Taking $\varepsilon \to 0$ allows us to conclude that

$$\Phi(x, y) = \frac{1}{i\pi} \frac{1}{x + iy}$$

is a fundamental solution to the operator $\partial_{\bar{z}}$. Similarily,

$$\Phi(x, y) = \frac{1}{i\pi} \frac{1}{x - iy}$$

is a fundamental solution to the operator ∂_z. The first fundamental solution is called the Cauchy Kernel.

A simple consequence of the fundamental solution technique is the intriguing statement that *any* distribution is a successive derivative of a family of continuous functions.

Theorem 15.10. *For any open set $U \subset \mathbb{R}^n$, and any distribution $u \in \mathcal{D}(U)^*$, there exists $f_\alpha \in C(U)$, such that the cover $\{\text{supp}(f_\alpha)\}$ is locally finite on U, and such that $u = \sum_\alpha D^\alpha f_\alpha$.*

Proof. Suppose first that $U = \mathbb{R}^n$. For any m, the function

$$\Phi(x) = \max(x_1, 0)^m \ldots \max(x_n, 0)^m / (m!)^n$$

is a fundamental solution to the operator $L = (\partial_1 \cdots \partial_n)^{m+1}$, and lies in $C^{m-1}(\mathbb{R}^d)$. If u is a distribution of order $m-1$, then it follows that $v = \Phi * u$ lies in $C(\mathbb{R}^d)$, and $Lv = u$, which completes the proof in this case. In general, we just localize. $\quad\square$

Using Fourier analytic techniques, we can construct fundamental solutions to all homogeneous elliptic constant coefficient partial differential equations, and *parametrices* for any elliptic constant coefficient partial differential equation $P(D_x)$, i.e. a distribution u such that $P(D_x)u = \delta + \phi$, where $\phi \in C^\infty(\mathbb{R}^d)$.

Theorem 15.11. *Every elliptic constant coefficient partial differential operator $P(D_x)$ has a parametrix u, and a fundamental solution if $P(D_x)$ is homogeneous.*

Proof. If P is elliptic and homogeneous of some degree m, then $P(\xi)$ is non-vanishing away from the origin. It thus follows that $1/P(\xi)$ is homogeneous of degree $-m$ away from the origin, and thus extends to some distribution v on $\mathbb{R}^d$, though we note that if $m \geqslant d$ it may fail to be homogeneous, though we can still make this choice in such a way that $P \cdot v = 1$. It will still be tempered, however, and if we let u denote it's inverse Fourier transform, then the Fourier transform of $P(D_x)u$ is equal to one, and thus $P(D_x)u = \delta$.

If P is no longer homogeneous, but it is elliptic, then there exists $\chi \in C^\infty(\mathbb{R}^d)$ supported away from the origin and such that $\chi(\xi) = 1$ if $|\xi|$ is sufficiently large, such that the function

$$v(\xi) = \frac{\chi(\xi)}{P(\xi)}$$

is well-defined and smooth. Then it is tempered, so we can take it's inverse Fourier transform u. The Fourier transform of $P(D_x)u - \delta$ is then supported on a neighborhood of the origin, which by the Paley-Wiener theorem, impies $P(D_x)u - \delta$ is a smooth function. $\qquad\square$

If P is elliptic and homogeneous, then the calculations above show that one choice of fundamental solution for the differential operator $P(D_x)$ can be written as

$$u(x) = \int \frac{e^{2\pi i \xi \cdot x}}{P(\xi)} \, d\xi.$$

By deforming the integration contour, one can actually show that u is *analytic* away from the origin. TODO: See Remark following Theorem 7.1.22 in Hormander.

The fundamental solutions to the heat equation and Cauchy-Riemann equation are smooth away from the origin. It will turn out that this is true for any constant coefficient partial differential equation which is *hypoelliptic*. Here are several consequences.

Lemma 15.12. *Suppose Φ is a fundamental solution to a constant coefficient partial differential operator L on $\mathbb{R}^n$, and suppose that sing-supp$(\Phi) = \{0\}$. Then for any open set $U \subset \mathbb{R}^n$, and any distribution u on U, sing-supp$(Lu) =$ sing-supp(u). Thus any distribution u on U with $Lu = 0$ lies in $C^\infty(U)$. If $\{u_n\}$ is a sequence of solutions to the equation $Lu = 0$, with $u_n \to u$ distributionally, then $u_n \to u$ in $C^\infty(U)$.*

Proof. It is simple to see that sing-supp$(Lu) \subset$ sing-supp(u). Conversely, fix a compact set $K \subset U$, and consider $\psi \in \mathcal{D}(U)$ equal to one on a neighbourhood of K. Then for any distribution u on U with $Lu = 0$,

$$
\begin{aligned}
K \cap \text{sing-supp}(Lu) &= K \cap \text{sing-supp}(L(\psi u)) \\
&= K \cap \text{sing-supp}(\Phi * (\psi u)) \\
&= K \cap \left[\text{sing-supp}(\Phi) + \text{sing-supp}(\psi u) \right] \\
&= K \cap \text{sing-supp}(u).
\end{aligned}
$$

The first part is therefore completed, since K was arbitrary. Now given the Fréchet space $X = \{u \in \mathcal{D}(U)^* : Lu = 0\}$, we have a natural inclusion map $i : X \to C^\infty(U)$. It's graph is obvious closed, so by the closed graph theorem, this map is continuous, and this implies the sequence property above. $\square$

The fundamental solutions above are actually better than just being smooth away from the origin, they are *analytic* away from the origin. This is true of any *elliptic* partial differential equation with constant coefficietns. The following lemma then applies.

Lemma 15.13. *Suppose Φ is a fundamental solution to a constant coefficient partial differential operator L on $\mathbb{R}^n$, and suppose that Φ is real analytic away from the origin. Then for any open set $U \subset \mathbb{R}^n$, there is an open set $V \subset \mathbb{C}^n$ such that if u is a distribution on U, and $Lu = 0$, then u is an analytic function extending to a complex analytic function on V. In particular, we conclude that if u vanishes in a neighborhood of a point $x_0 \in U$, then u vanishes on the connected component of U containing x_0. Moreover, if u_n is a sequence in $C^\infty(U)$ with $Lu_n = 0$, and with $u_n \to u$, then the extension of the functions u_n to analytic functions on V converge locally uniformly to the extension of u.*

Proof. Choose any open set $\mathbb{R}^n - \{0\} \subset V_0 \subset \mathbb{C}^n$ such that Φ extends to a complex analytic function on V_0. Since the question is local, consider a compact set K, and

$\psi \in \mathcal{D}(U)$ equal to one on a neighborhood of K. Then for $x \in K$, we have

$$u(x) = (\psi u)(x) = (\Phi * L(\psi u))(x) = \int \Phi(x - y)L(\psi u)(y) \, dy,$$

and this formula continues to define a complex analytic function if we replace x with a complex parameter z provided that for any $y \in \operatorname{supp}(\nabla(\psi u))$, $z - y \in V_0$. This is a neighborhood of K. We can then perform this extension for many different choices of ψ, and thus of K, to show that u extends to an analytic function on some set V depending only on U. To get the latter part of the Lemma, we just apply the closed graph theorem. $\qquad\square$

Let us now extend the approximation theorem of Runge to the case of general elliptic partial differential equations.

Lemma 15.14. *Suppose Φ is a fundamental solution to a constant coefficient partial differential operator L on $\mathbb{R}^n$, and suppose that Φ is real analytic away from the origin. Consider two open sets $V \subset U \subset \mathbb{R}^n$, such that $U - V$ does not have a compact connected component. Then every solution $v \in C^\infty(V)$ to the equation $Lv = 0$ is a limit in $C^\infty(V)$ of the restriction of solutions $u \in C^\infty(U)$ to the equation $Lu = 0$.*

Proof. By duality, it suffices to show that if w is a compactly-supported distribution in V, and

$$\int_V w(x)u(x) \, dx = 0$$

for any $u \in C^\infty(U)$ such that $Lu = 0$, then for any $v \in C^\infty(V)$ with $Lv = 0$,

$$\int_V w(x)v(x) \, dx = 0.$$

If L^* is the adjoint of the operator L, and we set $\Psi(x) = \Phi(-x)$, then Ψ is an analytic fundamental solution to L^*. If we can show that $\Psi * w$ is a compactly supported distribution in V, then for any $v \in C^\infty(V)$ with $Lv = 0$,

$$\int w(x)v(x) \, dx = \int L^*(\Psi * w)(x)v(x) \, dx = \int (\Psi * w)(x)Lv(x) \, dx = 0.$$

This would therefore complete the proof.

Let K denote the support of w. The distribution $\Psi * w$ is analytic function outside of K. Moreover, the support of $\Psi * w$ is certainly contained in U, because

276

if $x \notin U$, then the function $u(y) = \Psi(x - y)$ satisfies $Lu = 0$, and therefore by assumption on w, $(\Psi * w)(x) = 0$. Applying analyticity, it follows that $\Psi * w$ vanishes in every component of K^c containing a point in U^c. We claim this is sufficient to prove the result. If W is a bounded connected component of K^c contained in U, then the set $W \cap V^c$ is a compact connected component of $U - V$. This contradicts the hypothesis of the Lemma, and therefore W cannot exist. Because K^c has at most one unbounded connected component, we conclude that the theorem is proved, except in the case where U contains the unbounded connected component of K^c. We will address this case in two paragraphs. For now, note that this proves the theorem when U is a open ball.

Next, we address the case where $U = \mathbb{R}^d$, and V is a ball B, without loss of generality, centred at the origin. This is because if $v \in C^\infty(B)$ solves $Lv = 0$, then for each k, we can recursively find a sequence $u_{k,i} \in C^\infty(kB)$, such that $Lu_{k,i} = 0$, and such that $u_{1,i}$ converges to v in $C^\infty(B)$ as $i \to \infty$, such that $\|u_{k+1,i} - u_{k,i}\|_{L^\infty(kB)} \leqslant 1/2^{k+i}$ for any $|\alpha| \leqslant k + i$. It follows that we can define $u_i \in C^\infty(\mathbb{R}^d)$ as the pointwise limit of the functions $u_{k,i}$ (smooth because of analyticity). Then $Lu_i = 0$, and $u_i \to u$ in $C^\infty(B)$.

Now returning to the general case of the proof, if K is contained in a ball B of radius R, then for $|x| > R$, the function $u_x(y) = \Phi(y - x)$ solves the equation $Lu = 0$ in B, and thus is the limit in $C^\infty(B)$ of functions u_i solving $Lu_i = 0$ in $C^\infty(\mathbb{R}^n)$. But then

$$(\Phi * w)(x) = \int w(x)u(x) \, dx = \lim_{i \to \infty} \int w(x)u_i(x) \, dx = 0,$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark. This result is *only* possible given the topological hypothesis. Suppose the result is true, and K is a compact connected component of $U - V$. If $u_i \in C^\infty(U)$ is a sequence of solutions to the equation $Lu_i = 0$ such that u_i converges to a function $u \in C^\infty(V)$, then we claim that u_i converges to some $u \in C^\infty(V \cup K)$. In particular, we conclude every solution on $C^\infty(V)$ extends to a solution on $C^\infty(V \cup K)$. This causes a contradiction, since if $x_0 \in K$, then the function $u(y) = \Phi(y - x_0)$ solves the equation on V, but cannot be extended to a solution on $V \cup K$ (e.g. because any such solution would be analytic).

Let us now prove that u can be extended under our assumptions. If we consider $\phi \in \mathcal{D}(V \cup K)$ equal to one on a neighborhood W of K, then $L(\phi u_i)$ vanishes on W, and hence is supported in V. Since u_i converges to u in V, so $L(\phi u_i)$ converges to Lu in $V \cup K$, and this means that $u_i = \Phi * L(\phi u_i)$ converges in $C^\infty(V \cup K)$ to some function u in $C^\infty(V \cup K)$, which completes the proof of our claim.

Using Runge's approximation theorem, we can now finally prove our first major existence theorem for partial differential equations on proper subsets U of $\mathbb{R}^n$, assuming the existence of a fundamental solution (i.e. the existence of solutions to the equation $Lu = f$ for compactly supported distributions f on $\mathbb{R}^n$. To obtain the idea of the proof of the existence of solutions to the equation $Lu = f$, for a distribution f on U (not necessarily compactly supported), imagine that such a solution u existed. For any $\phi \in \mathcal{D}(U)$ equal to one on some open set W, we can find a distribution u_W such that $L(u_W) = \phi f$. Then $L(u - u_W) = 0$ on W, which implies that $u - u_W$ is analytic on W. If $u_V - u_W = 0$ on $V \cap W$ for any $V, W \subset U$, these solutions would combine to give us a choice of u on the whole domain. Instead, we only have $u_V - u_W$ *analytic* on $V \cap W$. The Runge approximation theorem will enable us to adjust these functions so that $u_V - u_W$ is *almost* equal to zero, and then we can take limits.

Theorem 15.15. *Suppose Φ is a fundamental solution to a constant coefficient partial differential operator L on $\mathbb{R}^n$, and suppose that Φ is real analytic away from the origin. Then for any open set U in $\mathbb{R}^n$, and any distribution $f \in \mathcal{D}(U)^*$, then there exists $u \in \mathcal{D}(U)^*$ such that $Lu = f$.*

Proof. Let $U_R = \{x \in U : |x| < R \text{ and } d(x, U^c) > 1/R\}$. Then we can apply Runge's approximation theorem for $U_R \subset U$. We leave this proof of this to the end for it is fairly technical and unenlightening. Now if $\phi_i \in \mathcal{D}(U)$ equals one on U_i, we set $u_i = \Phi * (\phi_i f)$. Then $Lu_i = \phi_i f$, in particular, $Lu_i = f$ on U_i. It suffices to modify u_i so that it converges distributionally to some function u, since it will then follow that $Lu = f$ on U. we note that $L(u_{i+1} - u_i) = 0$ in U_i. Thus Runge's theorem implies that there exists an analytic function w_i defined on U such that $\|(u_{i+1} - u_i) - w_i\|_{L^\infty(U_i)} \leqslant 1/2^i$. But then if we define $v_i = u_i - w_1 - \cdots - w_{i-1}$, then $Lv_i = \phi_i f$ on U, and

$$\|v_{i+1} - v_i\|_{L^\infty(U_i)} = \|(u_{i+1} - u_i) - w_i\|_{L^\infty(U_i)} \leqslant 1/2^i,$$

which implies the required convergence property.

To show that $U - U_R$ has no compact connected components, consider a potential compact connected component K of $U - U_R$. If B_R is the ball of radius R centred at the origin, $K \cap B_R^c$ is compact. It is also open in B_R^c, because we can write $K = V \cap U \cap U_R^c$ for some open set V, and then $K \cap B_R^c = V \cap U \cap U_R^c \cap B_R^c = V \cap U \cap B_R^c$ since $U_R \subset B_R$, and U and V are both open. But this means that $K \cap B_R^c = \varnothing$, i.e. $K \subset B_R$. Next, we claim $d(x, y) > 1/R$ for any $y \in U^c$. Otherwise, take y with $d(x, y) \leqslant 1/R$, and consider the line segment L between x and

y. Since U^c is closed, we may assume without loss of generality (by picking the closest point in U^c to *x* on the line and replacing *y* with it), that $L \cap U^c = \{y\}$. Since $x \notin U_R$, and $d(x_1, y) < 1/R$ for any other point x_1 on the line, $L \cap U_R = \emptyset$. Thus $L - \{y\} \subset U - U_R$. But $L - \{y\}$ is connected and contains $x \in K$, so $L - \{y\} \subset K$. But this is impossible for it implies that ∂K contains *y*, which is not an element of *K*, so *K* cannot be closed. Put together, this means that for any $x \in K$, $d(x, U^c) = 1/R$. But then it is impossible for *K* to be an open subset of $U - U_R$, so *K* cannot exist. $\qquad \square$

Example. *Suppose* $L = \sum_{|\alpha|} c_\alpha D^\alpha = P(D)$ *is a homogeneous elliptic partial differential equation of degree m on* $\mathbb{R}^n$. *Then* $1/P(\xi)$ *is a homogeneous function defined on* $\mathbb{R}^n - \{0\}$, *of degree* $-m$, *analytic in* $\mathbb{R}^n - \{0\}$. *Thus we have a fundamental solution* Φ *defined such that* $\hat{\Phi} = 1/P$.

- *If* $n > m$, *then L has a fundamental solution* $\Phi = \Phi_0$ *is homogeneous of degree* $m - n$, *and* C^∞ *in* $\mathbb{R}^n - \{0\}$.

- *If* $n \leqslant m$, *then L has a fundamental solution* $\Phi = \Phi_0 - Q(x) \log |x|$, *where* Φ_0 *is homogeneous of degree* $m - n$ *and*

$$Q(x) = \frac{1}{(2\pi)^m (m-n)!} \int_{|\xi|=1} \frac{(2\pi i x \cdot \xi)^{m-n}}{P(\xi)} \, d\xi.$$

TODO: Technical FT Calculation in Hörmander Volume 1, Theorem 7.1.20.

In either case, Φ_0 *is analytic in* $\mathbb{R}^n - \{0\}$.

15.2 The Cauchy Problem

Recall the Cauchy problem: Given a linear differential operator $L = \sum c_\alpha D^\alpha$ of order *m*, defined on an open set $\Omega_0 \subset \mathbb{R}^{n+1}$, with coefficients in $C^\infty(\Omega_0)$. We suppose that Ω_0 is divided into two parts by a smooth hypersurface *S*. Thus Ω_0 is divided into the union of three disjoint set $\Omega_0^- \cup S \cup \Omega_0^+$, where Ω_0^- and Ω_0^+ are connected, open sets.

Classically, the goal of the Cauchy problem, given some data *f* on Ω_0, and $\phi_0, \ldots, \phi_{m-1}$ on *S*, is to find a solution $Lu = f$, such that the first $m - 1$ derivatives of *u* in the direction tangent to *S* agrees with the functions $\{\phi_i\}$. For smooth data this is often a fine way to pose the problem; but for more general data. Instead, we consider $f \in \mathcal{D}(\Omega_0)$, and

Consider a linear differential operator $L = \sum c_\alpha D^\alpha$ of order m, defined on an open set Finally, we consider some $f \in \mathcal{D}(\Omega_0)^*$.

For rough data, the best way to formulate the Cauchy problem is as follows: We consider some initial data $\phi \in \mathcal{D}$

Cauchy problem for this data is, given some set $X \subset \mathcal{D}(\Omega_0)^*$, to find $u \in X$ such that $Lu = f$.

15.2.1 First Order Operators

TODO

15.2.2 Calderon's Uniqueness Theorem

Let us discuss the phenomenon of *uniqueness* for the Cauchy problem. Thus our goal is to determine, for a given set $X \subset \mathcal{D}(\Omega_0)^*$, whether $L : X \to \mathcal{D}(\Omega_0)^*$ is injective. If X is a linear subspace of $\mathcal{D}(\Omega_0)^*$, it suffices to show that if $u \in X$, and $Lu = 0$, then $u = 0$. This simplifies the notation in the problem. To make the problem possible to study microlocally, we work locally. Thus we focus on the case where X solely contains functions which vanish on Ω_0^-, and will try and find conditions which guarantee that if $Lu = 0$, then u vanishes on a neighborhood of S.

Let $P(x, \xi)$ denote the principal symbol of L. We begin by assuming that S is *non characteristic* at any of it's points, i.e. that if $x_0 \in S$, and $\eta_0 \in T_{x_0}^* \mathbb{R}^{n+1} - \{0\}$ has $T_{x_0}S$ as it's kernel (conormal to S), then $P(x_0, \eta_0) \neq 0$. We actually assume S is *strongly non characteristic*, i.e. that for any $\eta_0 \in T_{x_0}^* \mathbb{R}^{n+1}$ conormal to S, and any $\eta_1 \in T_{x_0}^* \mathbb{R}^{n+1}$ not conormal to S, the polynomial $P(x_0, \eta_0 + z\eta_1)$ has m simple roots in the z-variable. For $m = 1$, this is equivalent to being noncharacteristic. For $m > 1$, the surface S can be noncharacteristic but not satisfy this property, for instance, if S is the x-axis, and $L = (D_x + iD_y)^2 = D_x^2 + 2iD_xD_y - D_y^2$.

Being strongly noncharacteristic is a *local* conditions on L and of S; if we replace S with any other surface tangent to S at x_0, the condition will continue to hold in a neighborhood of x_0. Thus, working locally, we may switch to a coordinate system in which it suffices to study a more familiar version of the Cauchy problem, namely, we consider a set $\Omega_0 = (-T, T) \times \Omega$, where Ω is a connected, open subset of $\mathbb{R}^n$, $T > 0$, where $S = \{0\} \times \Omega$, and where

$$\Omega^- = \{(t, x) \in (-T, T) \times \Omega : t < 0\} \quad \text{and} \quad \Omega^+ = \{(t, x) \in (-T, T) \times \Omega : t > 0\}.$$

The principal symbol will then be written in the (t, x) variables, and their dual variables (τ, ξ) variables, i.e. as the polynomial $P(t, x, \tau, \xi)$. The assumption slightly stronger than being noncharacteristic then locally amounts to the fact that for any $\xi \in \mathbb{R}^n - \{0\}$, and any $z \in \mathbb{C}$, if $P(t, x, z, \xi) = 0$, then $\partial_z P(t, x, z, \xi) \neq 0$. It follows that locally on Ω we can write the solutions to $P(t, x, \tau, \xi) = 0$ in the τ variable for $1 \leqslant i \leqslant m$ as the smooth functions $\tau_1, \ldots, \tau_m : (-T, T) \times \Omega_x \times (\mathbb{R}^n_\xi - \{0\}) \to \mathbb{C}$. Given that P is homogeneous in the ξ variable of degree m, $\tau_1, \ldots, \tau_m$ will be homogeneous of degree one.

Example. *Let $L = \partial_t^2 + \Delta_x$. Then L has principal symbol $-4\pi^2(\tau^2 + |\xi|^2)$, and has roots $\pm i|\xi|$. The principal symbol factors as $-4\pi^2(\tau - i|\xi|)(\tau + i|\xi|)$.*

Since P is a homogeneous polynomial of degree m, the coefficient of τ^n in the expansion of P must be a non-vanishing, C^∞ function of P. If it vanished at a point (t_0, x_0), then we would have $P(t_0, x_0, \tau, 0) = 0$ for all $\tau \in \mathbb{R}$, which contradicts the fact that L is strongly noncharacteristic since then $\partial_\tau P(t_0, x_0, \tau, 0) = 0$. If $a \in C^\infty((-T, T) \times \Omega)$ is the coefficient of τ^n, it is therefore clear that the solutions of the operator $Lu = 0$ are the same as solutions of the operator $a^{-1} \cdot Lu = 0$. Thus in the sequel, we will assume without loss of generality that $a(t, x) = 1$ for all $t \in (-T, T)$ and all $x \in \Omega$.

Lemma 15.16. *Suppose that L is a differential operator of order m on $(-T, T) \times \Omega$, such that $S = \{(t, x) : t = 0\}$ is strongly noncharacteristic with respect to L. Then there exists m classical pseudodifferential operators $Z_1, \ldots, Z_m$, each of order one, with principal symbols $z_1, \ldots, z_m$ such that*

$$L = (D_t - Z_1) \ldots (D_t - Z_m) + R,$$

where R is a smoothing operator.

Proof. The composition calculus shows that $L - (D_t - z_1(t, x, D_x)) \ldots (D_t - z_m(t, x, D_x))$ is a pseudodifferential operator R_1 of order $m - 1$. As with many regularizing calculi. We proceed asymptotically, like in the construction of parametrices. Given a choice of operators $\{Z_i^N\}$ such that $R_N = L - (D_t - Z_1^N) \ldots (D_t - Z_m^N)$ is order $m - N$, we must consider operators $W_1, \ldots, W_m$ of order $1 - N$ such that $R_{N+1} = L - (D_t - Z_1^N - W_1) \ldots (D_t - Z_m^N - W_m)$ is an operator of order $m - N - 1$. If $w_1, \ldots, w_n$ are the symbols of $W_1, \ldots, W_m$, then it clearly suffices to choose these symbols such that if r_N is the principal symbol of R_N, then

$$r_N(t, x, \tau, \xi) = \sum_{i=1}^n w_i \left[\prod_{j \neq i} (\tau - z_j) \right].$$

If we pick
$$w_i = \frac{r_N}{\partial_z P(t, x, z_i, \xi)} = \frac{r_N}{\prod_{j \neq i}(z_i - z_j)},$$

then w_k is homogeneous of degree $m - N - (m-1) = 1 - N$, and since the rational function
$$f(\tau) = \sum_{i=1}^{n} \prod_{j \neq i} \frac{\tau - z_j}{z_i - z_j}$$

is a degree $m - 1$ polynomial with $f(z_1) = \cdots = f(z_m) = 1$, and so $f(\tau) = 1$ for all $\tau \in \mathbb{R}$, which implies the required formula for r_N. Now we just take $Z_i = \lim Z_i^N$, and we get the required result. $\qquad\square$

The next remark shows that we can assume that there exists a compact set K' such that $\mathrm{supp}_x(u) \subset K'$, such that the coefficients of L vanish outside of K', and the (x, y) support of the kernels of $Z_1, \ldots, Z_m$, and of R vanish outside of K', uniformly in t.

Remark. Switch to a coordinate system in which $S = \{(t, x) \in (-T, T) \times \Omega : t = |x|^2\}$, after possibly shrinking Ω. In this coordinate system, if u vanishes on Ω_0^-, then $u(t, x)$ vanishes for $t < |x|^2$. Thus, after shrinking T, we may assume that Ω contains a ball of radius T^2, and then $\mathrm{supp}_x(u)$ is contained in that ball of radius T^2. Switching back to our original coordinate system, we may assume without loss of generality that there exists a compact set $K \subset \Omega$ such that $\mathrm{supp}_x(u) \subset K$. If $g \in C_c^\infty(\Omega)$ is equal to one on a neighborhood of K, then $0 = Lu = g \cdot Lu = (D_t - gZ_1) \ldots (D_t - gZ_m) + R^\# u$, which is a similar expansion to the original except that we can now assume that the (x, y) supports of the kernels of all operators involved are contained in $K' \times K'$ for some compact set K'.

Define $u_1 = u$, and for $1 \leqslant i \leqslant m - 1$, iteratively define
$$u_{i+1} = (D_t - Z_i)u_i.$$

Then since $L = (D_1 - Z_1) \ldots (D_n - Z_m)u + R$, it follows that
$$(D_t - Z_m)u_m = Lu - Ru = -Ru.$$

We can then consider the diagonal matrix valued pseudodifferential operator
$$A = i \cdot \mathrm{diag}(Z_1, \ldots, Z_m).$$

Let N be the standard nilpotent matrix, with ones directly above the diagonal. Let $\tilde{R}(t)$ be the $m \times m$ equal to $R(t)$ in the lower left corner. Then if $U = (u_1, \ldots, u_m)$, then

$$\mathcal{L}U = \frac{\partial U}{\partial t} - A(t)U - i[N - \tilde{R}(t)]U = 0.$$

Let us now assume that $u \in C^i((-T, T), H_c^{m-j}(\Omega))$. Then $u_k \in C^j((-T, T), H_c^{m-k+1-j})$. Thus for $j = 0$ and $j = 1$, $U \in C^j((-T, T), H_c^{1-j}(\Omega) \otimes \mathbb{C}^m)$.

The uniqueness will now follow from an *inequality of Carleman type*. Namely, if $V(t, x) = \chi(t)U(t, x)$, where $\chi(t) = 1$ for $t < 8T/10$, and vanishes for $t > 9T/10$, then for all suitably large positive values $\rho > 0$,

$$\rho \int_0^T \int_\Omega e^{\rho(T-t)^2} |V(t, x)|^2 \, dx \, dt \lesssim \int e^{\rho(T-t)^2} |\mathcal{L}V(t, x)|^2 \, dx \, dt.$$

We then claim that U vanishes for $t < T/2$. Indeed, we can restrict the integration on the right hand side to the region $8T/10 \leqslant t \leqslant T$, because $\mathcal{L}U$ because it satisfies the ODE. But if we limit the integration on the left hand side to $0 \leqslant t \leqslant t/2$, we get that

$$\rho \cdot e^{\rho(24/25)T^2} \int_0^{t/2} \int_\Omega |V(t, x)|^2 \, dx \, dt \lesssim \int_{8T/10}^T \int_\Omega |\mathcal{L}V|^2 \, dx \, dt.$$

Unless the right hand side is equal to zero, it will become arbitrarily large as $\rho \to \infty$, which eventually gives a contradiction. Thus V must vanish on $0 \leqslant t \leqslant T/2$.

Remark. Note that we did not use ρ in the proof. It's utility is that if ρ is suitably large, then we can replace $\mathcal{L}$ by the diagonal part $\mathcal{L}_0 = I_m(\partial/\partial t) - A(t)$, and thus discard the other terms as extraneous.

Calderon proved this result (Uniqueness in the Cauchy problem of partial differential equations, 1958) under two assumptions:

- The roots $z_1, \ldots, z_m$ are all valued in $\mathbb{C} - \mathbb{R}$, i.e. so that P is elliptic.

- The roots $z_1, \ldots, z_m$ are all valued in $\mathbb{R}$. Under this condition, we say that $z_1, \ldots, z_m$ are *strongly hyperbolic*.

TODO: Rest of argument (Treves, Vol 1. Starting on page 112).

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, given two function spaces V and W, showing some operator $T : V \to W$ is bounded, i.e. so that we can 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 for some universal constant $C > 0$, we have $B(Tf) \leqslant CAf$ for all $f \in V$.

A core technique to the real-variable theory we employ is the method of *decomposition*. Given some function f, we find functions $\{f_k\}$ such that $f = \sum_k f_k$, where the functions f_k have particular properties, perhaps being localized in space or frequency. 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 in an appropriate manner. One core property required to ensure A and B are somewhat stable under addition is to assume 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 $g \in W$, since then the triangle inequality applies. 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 $\|f_k\|_V \lesssim 2^{-\varepsilon k}\|f\|_V$, which is often obtained by an appropriate *dyadic decomposition*. For such decompositions, it is also possible to generalize our techniques not only to norms, but also to *quasinorms*, i.e. homogeneous maps $\|\cdot\|$ satisfying a *quasi-triangle inequality* $\|v+w\| \lesssim \|v\|+\|w\|$.

Lemma 15.17. *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

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

A slight modification of the proof below even gives this claim provided T is *quasi sublinear*, in the sense that for all $f_1, f_2 \in V$, $\|T(f_1+f_2)\|_W \lesssim \|Tf_1\|_W + \|Tf_2\|_W$ for all $f_1, f_2 \in V$. However, such operators occur so rarely in practice that we leave the result as an exercise to the reader.

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

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

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

We can obtain somewhat similar results also in the presence of suitably large polynomial decay.

Lemma 15.18. *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 $\|f_k\|_V \leq C \cdot k^{-s}$, for $s > t$, then $\|f\|_V \lesssim_s C$.*

Proof. As in the previous lemma, pick $A > 0$ such that $\|f_1 + f_2\|_V \leq A(\|f_1\|_V + \|f_2\|_V)$ 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 $2^m \leq k \leq 2^{m+1} - 1$ via a binary tree with depth at most $O(m)$, we can ensure that

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

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 16

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, instead only caring about the 'height' and 'width' of a function. Height and width are of course, vague properties; but roughly speaking, the canonical example of a function with height H and width W is a function $f = H \cdot \mathbb{I}_E$, where $|E| = W$. More general functions can be thought of as being composed from a combination of functions with various heights and widths, and since a unique way to quantify the behaviour of these combinations is not obvious, we will obtain various different ways to quantify how various heights and widths come together.

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

- The *indicator functions* $\mathbb{I}_E(x) = \mathbb{I}(x \in E)$, for a measurable set E.

- The *simple functions* $f = \sum_{i=1}^{n} a_i \, \mathbb{I}_{E_i}$, 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.

16.1 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

$$\|f\|_{L^\infty(X)} = \min \left\{ t \geqslant 0 : |f(x)| \leqslant t \text{ almost surely} \right\},$$

a quantity often called the *essential supremum* of the function. 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^2 bounds is aided by the phenomena of orthogonality. Obtaining L^p bounds of an operator for $1 < p < 2$ and $2 < p < \infty$ therefore 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 homogeneous, since the width of f and αf are the same for each $\alpha \neq 0$.

Example. *If $f(x) = |x|^{-s}$ for $x \in \mathbb{R}^d$ 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 $B \in \mathbb{R}^d$. 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 $L^p(\mathbb{R}^d)$, 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 16.1. *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

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

Taking $\varepsilon \to 0$ gives the result. $\qquad\qquad\square$

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

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

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

291

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 $\lambda \geqslant 0$,*

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

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 16.2 (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

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

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 16.3. *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

$$\int |f(x) + g(x)|^p \leqslant \int |f(x)||f(x) + g(x)|^{p-1} + |g(x)||f(x) + g(x)|^{p-1}$$

$$\leqslant \|f\|_p \|(f + g)^{p-1}\|_q + \|g\|_p \|(f + g)^{p-1}\|_q = \|f + g\|_p^{p-1}$$

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 16.3 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 16.4. *If $p \geqslant 1$, then for any functions $f_1, \ldots, f_N \geqslant 0$,*

$$(\|f_1\|_p^p + \cdots + \|f_N\|_p^p)^{1/p} \leqslant \|f_1 + \cdots + f_N\|_p \leqslant \|f_1\|_p + \cdots + \|f_N\|_p. \qquad (16.1)$$

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

$$\|f_1\|_p + \cdots + \|f_N\|_p \leqslant \|f_1 + \cdots + f_N\|_p \leqslant (\|f_1\|_p^p + \cdots + \|f_N\|_p^p)^{1/p} \qquad (16.2)$$

Proof. The upper bound in (16.1) is just obtained by applying the triangle inequality iteratively. To obtain the lower bound, we note that for $A_1, \ldots, A_N \geqslant 0$,

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

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

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

Now assume $0 < p < 1$. We begin by proving the lower bound in 16.2. 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 $(A_1 + \cdots + A_N)^p \leqslant A_1^p + \cdots + A_N^p$, which holds for $A_1, \ldots, A_N \geqslant 0$, can be applied with $f_k = A_k$ and integrated to yield

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

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 16.5. *For $f_1, \ldots, f_N \in L^p(X)$, with $0 < p < 1$,*

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

Proof. By Hölder's inequality applied to sums,

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

295

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 $f_i = \mathbb{I}_{E_i}$, 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 $H\,\mathbb{I}_E$ 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

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

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 $0 < p \leqslant q \leqslant \infty$, $L^q(X) \subset L^p(X)$, because 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

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

If $\varepsilon = 1$, which occurs if $X = \mathbb{Z}$, 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 $(\mathbb{Z}/\varepsilon)^d$, 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 $X = \{1, \ldots, N\}$, 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. $f(n) = \mathbb{I}(n = 1)$. 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 16.6. *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 16.7. *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 $s_1 \leqslant s_2 \leqslant \cdots \to |f|$. 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

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

this completes the proof in this case.

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

16.2   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 16.8. *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 16.9. *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.

16.3 Weak Norms

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

Theorem 16.10. *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

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

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 $f \notin L^p(\mathbb{R}^d)$. 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 $f = f_1 + \cdots + f_N$, and $\alpha_1, \ldots, \alpha_N \in [0,1]$ satisfy $\alpha_1 + \cdots + \alpha_N = 1$, then

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

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

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

If $p = 1$, then

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

If $0 < p < 1$, then

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

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

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

303

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

$$(4/t) \int |j(x)| \, dx \leqslant (4/t) \sum_{k=1}^{N} \int |j_k(x)|$$

$$= (4/t) \sum_{k=1}^{N} \int_{\|f_k\|_{1,\infty}(t/2)}^{t/2} \frac{\|j_k\|_{1,\infty}}{s} \, ds$$

$$= (4/t) \sum_{k=1}^{N} \|f_k\|_{1,\infty} \log(1/\|f_k\|_{1,\infty}).$$

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

$$\frac{4}{t} \int |j(x)| \, dx \leqslant \frac{4}{t} \sum_{k=1}^{N} \int_{\|f_k\|_{p,\infty}^p (t/2)}^{t/2} \frac{\|f_k\|_{p,\infty}^p}{s^p} \, ds$$

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

$$= \frac{2^{p+1}}{t^p} \frac{1}{1-p} \max \|f_k\|_{p,\infty}^{p(1-p)} \lesssim_p \frac{1}{t^p},$$

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

Remark. For $p = 1$, compare this *logarithmic* failure to be a norm with the *polynomial* failure to be a norm found in the norms $\| \cdot \|_p$, when $p < 1$, in Theorem 16.5.

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

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

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

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

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

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

The weak L^p norms provide another monotone translation invariant norm, and it oftens comes up when finer tuning is needed in certain interpolation arguments, especially when dealing with maximal functions.

Example. *If $f = H \mathbb{I}_E$, with $|E| = W$, then*

$$F(t) = W \cdot \mathbb{I}_{[0,H)} .$$

Thus

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

305

If $f = H_1 \, \mathbb{I}_{E_1} + H_2 \, \mathbb{I}_{E_2}$, with $|E_1| = W_1$ and $|E_2| = W_2$, with $H_1 \leqslant H_2$, then

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

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 $L^p(\mathbb{R}^d)$, 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 16.12. *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 $f_k = \mathbb{I}_{F_k} f$, 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

$$\int_E |f(x)| \, dx \leqslant \sum_{k=-\infty}^{n-1} \int_E |f_k(x)| \, dx + \sum_{k=n}^{\infty} \int_E |f_k(x)| \, dx$$

$$\leqslant |E| 2^{n-1} + \|f\|_{p,\infty}^p \sum_{k=n}^{\infty} 2^{k(1-p)}$$

$$\lesssim_p |E| 2^n + \|f\|_{p,\infty}^p 2^{-k(1-p)}.$$

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

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

so

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

which gives $\|f\|_{p,\infty} \leq A$. $\qquad\square$

For $p \leq 1$, 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 16.13. *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 $|E'| \geq \alpha|E|$ such that*
$$\int_{E'} |f(x)| \, dx \lesssim_{\alpha,p} A|E'|^{1-1/p}.$$

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

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

then

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

307

and

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

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

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

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

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

On the other hand,

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

Rearranging this equation gives $|E_k'| \leq 2^{-pk}$, 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 $\|f\|_{p,\infty} \lesssim_{\alpha,p} 1$. □

16.4   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 $\mathbb{R}^+$. But if we change the integrality condition to the condition that $F(t)t^p \in L^q(\mathbb{R}^+)$ for some $0 < q \leq \infty$, 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

$$\|f\|_{p,q} = p^{1/q} \|t F^{1/p}\|_{L^q(\mathbb{R}^+)}$$

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

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

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

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. □

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 16.15. *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

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

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

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

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

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

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

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 $k \in \mathbb{Z}$,

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

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

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

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

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

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

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

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

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

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

311

$f = \sum_{k\in\mathbb{Z}} f_k$, then f_k is supported on a set with measure $W_k = F(2^{k-1}) - F(2^k) \leq F(2^{k-1})$, and so

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

Thus $A_2 \lesssim_q A_1$. Next, we bound A_4 in terms of A_1. Perform a horizontal layer cake decomposition, writing $f = \sum f_k$, where f_k is supported on a set with measure $W_k \leq 2^k$, and $H_{k+1} \leq |f_k(x)| \leq H_k$. Then a telescoping sum shows

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

Thus

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

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 $|f(x)| \leq \sum 2^k \mathbb{I}_{E_k}$, where $|E_k| \leq W_k$. We then find

$$F(2^k) \leq \sum_{m=1}^\infty W_{k+m}.$$

Thus

$$\int_{2^{k-1}}^{2^k} [tF(t)^{1/p}]^q \frac{dt}{t} \lesssim \left[2^k \left(\sum_{m=0}^\infty W_k \right)^{1/p} \right]^q.$$

312

Thus if $q \leqslant p$,

$$
\begin{aligned}
\|f\|_{p,q} &\lesssim_q \left(\sum_{k \in \mathbb{Z}} \left[2^k \left(\sum_{m=0}^{\infty} W_{k+m} \right)^{1/p} \right]^q \right)^{1/q} \\
&\leqslant \left(\sum_{k \in \mathbb{Z}} \sum_{m=0}^{\infty} \left[2^k W_{k+m}^{1/p} \right]^q \right)^{1/q} \\
&\leqslant \left(\sum_{m=0}^{\infty} 2^{-qm} \sum_{k \in \mathbb{Z}} \left[2^{k+m} W_{k+m}^{1/p} \right]^q \right)^{1/q} \\
&\leqslant \left(A_3^q \sum_{m=0}^{\infty} 2^{-mq} \right)^{1/q} \lesssim_q A_3.
\end{aligned}
$$

If $q \geqslant p$, we can employ the triangle inequality for $l^{q/p}$ to write

$$
\begin{aligned}
\|f\|_{p,q} &\lesssim_q \left(\sum_{k \in \mathbb{Z}} \left[2^k \left(\sum_{m=0}^{\infty} W_{k+m} \right)^{1/p} \right]^q \right)^{1/q} \\
&\leqslant \left(\sum_{m=0}^{\infty} \left(\sum_{k \in \mathbb{Z}} 2^{kq} W_{k+m}^{q/p} \right)^{p/q} \right)^{1/p} \\
&\leqslant \left(A_3^p \sum_{m=0}^{\infty} 2^{-mq} \right)^{1/p} \lesssim_{p,q} A_3.
\end{aligned}
$$

The bound of A_1 in terms of A_4 involves the same 'shifting' technique, and is left to the reader. $\qquad\square$

Remark. Heuristically, the theorem above says that if $f = \sum_{k \in \mathbb{Z}} f_k$, 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

$$
\|f\|_{p,q} \sim_{p,q} \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p} \right]^q \right)^{1/q}.
$$

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

313

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

$$\left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p} \right]^{q_1} \right)^{1/q_1} \leqslant N^{1/q_1 - 1/q_2} \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p} \right]^{q_2} \right)^{1/q_2}$$

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 $p_1 < p_2$,

$$\left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_1} \right]^{q} \right)^{1/q} \leqslant \max(W_k)^{1/p_1 - 1/p_2} \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_2} \right]^{q} \right)^{1/q},$$

$$\left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_2} \right]^{q} \right)^{1/q} \leqslant \left(\frac{1}{\min(W_k)} \right)^{1/p_1 - 1/p_2} \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_2} \right]^{q} \right)^{1/q}.$$

which gives

$$\min(W_k)^{1/p_1 - 1/p_2} \|f\|_{p_2,q} \lesssim_{p_1,p_2,q} \|f\|_{p_1,q} \lesssim_{p_1,p_2,q} \max(W_k)^{1/p_1 - 1/p_2} \|f\|_{p_2,q}.$$

Both of these inequalities can be tight. Because of the dyadic decomposition of f, we find $\max(W_k) \geqslant 2^N \min(W_k)$, so these two norms can differ by at least $2^{N(1/p_1 - 1/p_2)}$, and at *most* if the f_k occur over consecutive dyadic values, which is *exponential* in N. Conversely, if the heights change dyadically, we find that

$$\left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_2} \right]^{q} \right)^{1/q} \leqslant \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_2} \right]^{qp_2/p_1} \right)^{(p_1/p_2)/q}$$

$$\leqslant \max(H_k)^{1 - p_1/p_2} \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_1} \right]^{q} \right)^{(p_1/p_2)/q}$$

$$\left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_1} \right]^{q} \right)^{1/q} \gtrsim \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_1} \right]^{qp_1/p_2} \right)^{(p_2/p_1)/q}$$

$$\leqslant \left(\frac{1}{\min(H_k)} \right)^{p_2/p_1 - 1} \left(\sum_{k \in \mathbb{Z}} \left[H_k W_k^{1/p_2} \right]^{q} \right)^{(p_2/p_1)/q}$$

314

where $\lesssim\atop\sim$ denotes a factor ignoring polynomial powers of N occuring from the estimate. Thus

$$\min(H_k)^{p_2-p_1}\|f\|_{p_1,q}^{p_1} \underset{p_1,p_2,q}{\lesssim\atop\sim} \|f\|_{p_2,q}^{p_2} \lesssim_{p_1,p_2,q} \max(H_k)^{p_2-p_1}\|f\|_{p_1,q}^{p_1}$$

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 $f_k = f\,\mathbb{I}_{E_k}$. 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 $L^{p,\infty}(\mathbb{R}^d)$.

A simple consequence of the layer cake decomposition is Hölder's inequality for Lorentz spaces.

Theorem 16.18. *If $0 < p_1, p_2, p < \infty$ and $0 < q_1, q_2, q < \infty$ with*

$$1/p = 1/p_1 + 1/p_2 \quad and \quad 1/q \geqslant 1/q_1 + 1/q_2,$$

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 $|f| \leqslant \sum_{k\in\mathbb{Z}} H_k \mathbb{I}_{E_k}$ and $|g| \leqslant \sum_{k\in\mathbb{Z}} H_k' \mathbb{I}_{F_k}$, where $|E_k|, |F_k| \leqslant 2^k$. Then

$$|fg| \leqslant \sum_{k,k'\in\mathbb{Z}} H_k H_k' \mathbb{I}_{E_k \cap F_{k'}}$$

315

For each fixed k, $|E_{k+m} \cap F_m| \leqslant 2^m$, and so

$$\left\| \sum_{m \in \mathbb{Z}} H_{k+m} H'_m \, \mathbb{I}_{E_{k+m} \cap F_m} \right\|_{p,q} \lesssim_{p,q} \left(\sum_{m \in \mathbb{Z}} [H_{k+m} H'_m 2^{m/p}]^q \right)^{1/q}$$

$$= \left(\sum_{m \in \mathbb{Z}} \left[(H_{k+m} 2^{m/p_1})(H_m 2^{m/p_2}) \right]^q \right)^{1/q}$$

$$\leqslant \left(\sum_{m \in \mathbb{Z}} [H_{k+m} 2^{m/p_1}]^{q_1} \right)^{1/q_1} \left(\sum_{m \in \mathbb{Z}} [H'_m 2^{m/p_2}]^{q_2} \right)^{1/q_2}$$

$$\lesssim_{p,q,p_1,q_1,p_2,q_2} 2^{-k/p_1}.$$

Summing over $k > 0$ gives that

$$\left\| \sum_{k \geqslant 0} \sum_{m \in \mathbb{Z}} H_{k+m} H'_m \, \mathbb{I}_{E_{k+m} \cap F_m} \right\| \lesssim_{p,q,p_1,q_1,p_2,q_2} 1$$

By the quasitriangle inequality, it now suffices to obtain a bound

$$\left\| \sum_{k < 0} \sum_{m \in \mathbb{Z}} H_{k+m} H'_m \, \mathbb{I}_{E_{k+m} \cap F_m} \right\|_{p,q}.$$

This is done similarily, but using the bound $|E_{k+m} \cap F_m| \leqslant 2^{k+m}$ instead of the other bound. $\square$

Corollary 16.19. *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

$$\|f\|_{L^1(E)} = \| \mathbb{I}_E f \|_{L^1(X)} \lesssim_{p,q} |E|^{1-1/p} \|f\|_{p,q} < \infty. \qquad \square$$

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.

316

Theorem 16.20. *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 $f = \sum_{k \in \mathbb{Z}} f_k$, 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 $g = \sum_{k \in \mathbb{Z}} g_k$, where $g_k(x) = a_k^{q-p} \mathrm{sgn}(f_k(x))|f_k(x)|^{p-1}$. Then

$$\int f(x)g(x) = \sum_{k \in \mathbb{Z}} \int f_k(x)g_k(x) = \sum_{k \in \mathbb{Z}} a_k^{q-p} \int |f_k(x)|^p$$

$$\gtrsim_p \sum_{k \in \mathbb{Z}} a_k^{q-p} W_k 2^{kp} = \sum_{k \in \mathbb{Z}} a_k^q \gtrsim_{p,q} 1.$$

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

$$\|g\|_{p',q'} \lesssim_{m,p,q} \left(\sum [H_k W_k^{1/p'}]^{q'} \right)^{1/q'}$$

$$\lesssim \left(\sum_{k \in \mathbb{Z}} \left[\left(\sup_{l \geqslant 0} a_{k-l}^{q-p} 2^{kp} 2^{-lp/2} \right) (a_k 2^{-k})^{p-1} \right]^{q'} \right)^{1/q'}$$

$$\lesssim_p \left(\sum_{k \in \mathbb{Z}} \left[a_k^{p-1} \sum_{l=0}^{\infty} a_{k-l}^{q-p} 2^{-lp/2} \right]^{q'} \right)^{1/q'}$$

$$\lesssim \sum_{l=0}^{\infty} 2^{-lp/2} \left(\sum_{k \in \mathbb{Z}} \left[a_k^{p-1} a_{k-l}^{q-p} \right]^{q'} \right)^{1/q'}.$$

317

Applying's Hölder's inequality shows

$$\left(\sum_{k\in\mathbb{Z}}\left[a_k^{p-1}a_{k-l}^{q-p}\right]^{q'}\right)^{1/q'} \leqslant \left(\sum_{k\in\mathbb{Z}}a_k^q\right)^{(p-1)/q}\left(\sum_{k\in\mathbb{Z}}a_{k-l}^q\right)^{(q-p)/q}$$

$$\lesssim_{p,q}\|f\|_{p,q}^{q-1}\lesssim_{p,q}1. \qquad\qquad \square$$

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 $f^1,\ldots,f^m$, 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

$$\|f\|_{p,q} \lesssim_m \sum_{i=1}^m \|f^i\|_{p,q}$$

$$\lesssim_{p,q} \sum_{i=1}^m \left(\sum\left[H_{km+i}W_{km+i}^{1/p}\right]^q\right)^{1/q}$$

$$\lesssim_m \left(\sum\left[H_kW_k^{1/p}\right]^q\right)^{1/q}$$

On the other hand,

$$\|f\|_{p,q} \gtrsim \max_{1\leqslant i\leqslant m}\|f^i\|_{p,q}$$

$$\sim \max_{1\leqslant i\leqslant m}\left(\sum\left[H_{km+i}W_{km+i}^{1/p}\right]^q\right)^{1/q}$$

$$\gtrsim_m \left(\sum\left[H_kW_k^{1/p}\right]^q\right)^{1/q}.$$

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.

318

Theorem 16.21. *For each* $1 < q < \infty$, *there is* $\alpha(q) > 0$ *such that for any functions* $f_1, \ldots, f_N$,

$$\|f_1 + \cdots + f_N\|_{1,q} \lesssim (\log N)^{\alpha(q)} \left(\|f_1\|_{1,q} + \cdots + \|f_N\|_{1,q}\right).$$

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 $f_1, \ldots, f_N$, 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

$$E = \{x : 2^{k-1} < |f_1(x) + \cdots + f_N(x)| \leq 2^k\}$$

The measure of the set E is upper bounded by the measure of the set

$$E' = \left\{ x : 2^{k-2} < \left| \sum_{j=k-\lg(N)}^{k} f_{1j}(x) + \cdots + f_{Nj}(x) \right| \leq 2^{k+1} \right\}$$

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

$$\|f_1 + \cdots + f_N\|_{1,q} \lessapprox_q \left(\sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q}.$$

319

Applying Minkowski's inequality, we conclude

$$\left(\sum_{k=-\infty}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q} \lesssim \sum_{i=1}^{N} \left(\sum_{k=-\infty}^{\infty} \left(\sum_{j=k-\lg N}^{k} W_{ij} 2^j \right)^q \right)^{1/q}$$

$$\lesssim \sum_{i=1}^{N} \left(\sum_{k=-\infty}^{\infty} \sum_{j=k-\lg N}^{k} W_{ij}^q 2^{qj} \right)^{1/q}$$

$$\lesssim \sum_{i=1}^{N} \left(\sum_{j=-\infty}^{\infty} W_{ij}^q 2^{qj} \right)^{1/q} \lesssim 1. \qquad \square$$

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 16.22. *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 $\| f_n \|_{p_i, \infty} \leqslant C 2^n W_n^{1/p_i}$ 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 $f_{nm} = f_n \cdot \mathbb{I}(2^{n+m} \leqslant |f_n| < 2^{n+m+1})$. 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

$$\| \tilde{f}_m \|_{L^p(X)} \lesssim_p \left\| \left(\sum_n |f_{nm}|^p \right)^{1/p} \right\|_{L^p(X)} = \left(\sum_n \| f_{nm} \|_{L^p(X)}^p \right)^{1/p}.$$

320

Chebyshev's inequality implies that

$$\|f_{nm}\|_{L^p(X)}^p \leqslant 2^{(n+m)p} \cdot \min_i \{2^{-(n+m)p_i} \|f_n\|_{L^{p_i,\infty}(X)}^{p_i}\}$$
$$\lesssim \min_i \{2^{(n+m)(p-p_i)} \|f_n\|_{L^{p_i,\infty}(X)}^{p_i}\}.$$

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. □

16.5 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

$$\|f\|_{L^p(X)L^q(Y)} = \left(\int \left(\int |f(x,y)|^q \, dy\right)^{p/q} dx\right)^{1/p}.$$

Similarly, we can define $\|f\|_{L^q(Y)L^p(X)}$. The notation is justified by the fact that the first case is the L^p norm of the function f, viewed as a map from X to $L^q(Y)$, and the latter the L^q norm of f, viewed as a map from Y to $L^p(X)$. We have a duality theory here; for each $1 \leqslant p, q \leqslant \infty$ and any f with $\|f\|_{L^p(X)L^q(Y)} < \infty$, the standard L^p and L^q duality gives

$$\|f\|_{L^p(X)L^q(Y)} = \sup\left\{\int_{X \times Y} f(x,y)h(x,y) \, dx \, dy : \|h\|_{L^{p*}(X)L^{q*}(Y)} \leqslant 1\right\}.$$

321

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, i.e. in the notation we use, the exponents are monotonically increasing from left to right.

Theorem 16.23. *If $p \geqslant q$, $\|f\|_{L^p(X)L^q(Y)} \leqslant \|f\|_{L^q(Y)L^p(X)}$.*

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 $q = 1$, then this result is precisely the Minkowski inequality. But not complex interpolation justifies the result in general. More precisely, a variation of the proof of Riesz-Thorin using the duality established above gives the result. $\square$

Let us consider two special cases. Firstly, bounds for the pointwise maxima of functions dominate the maximum L^p norms

$$\sup_n \|f_n\|_{L^p(X)} \leqslant \left\|\sup_n f_n\right\|_{L^p(X)}.$$

Another special case is the triangle inequality

$$\left\|\sum_n f_n\right\|_{L^p(X)} \leqslant \sum_n \|f_n\|_{L^p(X)}$$

for $p \geqslant 1$.

The next result shows that if $p < q$ 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 we think the functions we are working with are close to tensor products.

Theorem 16.24. *Suppose $p > q$, 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

$$\left(\int_Y \left(\int_X |f(x,y)|^p \, dx\right)^{q/p} dy\right)^{1/q} = \left(\int_X \left(\int_Y |f(x,y)|^q \, dy\right)^{p/q} dx\right)^{1/p}.$$

322

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

$$\int_Y \left(\int_X |f(x, y)| \, dx \right)^p dy$$

$$= \int_X \left[\int_Y |f(x_1, y)| \left(\int_X |f(x_2, y)| \, dx_2 \right)^{p-1} dy \right] dx_1$$

$$\leqslant \int_X \left[\left(\int_Y |f(x_1, y)|^p \, dy \right)^{1/p} \left(\int_Y \left(\int_X |f(x_2, y)| \, dx_2 \right)^{(p-1)p^*} dy \right)^{1/p^*} \right] dx_1$$

$$= \left[\int_X \left(\int_Y |f(x_1, y)|^p \, dy \right)^{1/p} dx_1 \right] \left[\int_Y \left(\int_X |f(x_2, y)| \, dx_2 \right)^p \right]^{1/p^*}.$$

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. $\quad\square$

16.6   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 $\|f\|_\Phi \leq 1$, then

$$\int \Phi\left(|f(x)|\right) \, dx = 1.$$

Since a norm would be homogenous, this would imply that if $\|f\|_\Phi \leq A$, then

$$\int \Phi\left(\frac{|f(x)|}{A}\right) \, dx \leq 1.$$

If we want these norms to be monotone, we might ask that if $A < B$, then

$$\int \Phi\left(\frac{|f(x)|}{B}\right) \, dx \leq \int \Phi\left(\frac{|f(x)|}{A}\right),$$

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 $\{f : \|f\|_\Phi \leq 1\}$ 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

$$\|f\|_\Phi = \inf\left\{A > 0 : \int \Phi\left(\frac{|f(x)|}{A}\right) \, dx \leq 1\right\}.$$

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

$$\int \Phi\left(\frac{|f(x)|}{A}\right) \leq 1 \quad \text{and} \quad \int \Phi\left(\frac{|f(x)|}{B}\right) \leq 1,$$

then applying convexity gives

$$\int \Phi\left(\frac{|f(x) + g(x)|}{A + B}\right) \leq \int \Phi\left(\frac{|f(x)| + |g(x)|}{A + B}\right)$$
$$\leq \int \left(\frac{A}{A + B}\right) \Phi\left(\frac{|f(x)|}{A}\right) + \left(\frac{B}{A + B}\right) \Phi\left(\frac{|g(x)|}{B}\right) \leq 1.$$

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 16.25. *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

$$\int \Phi\left(\frac{|f(x)|}{A}\right) \, dx \leqslant \Phi(\varepsilon)|X| + C_\varepsilon \int \Psi\left(\frac{|f(x)|}{A}\right)$$
$$\lesssim \Phi(\varepsilon)|X| + \frac{C_\varepsilon}{A}.$$

325

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 $\|f\|_{\Phi(L)} \lesssim 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The Orlicz spaces satisfy an interesting duality relation. Given a function Φ, which we assume is *superlinear*, in the sense that $\Phi(x)/x \to \infty$ as $x \to \infty$, define it's *Young dual*, for each $y \in [0, \infty)$, by

$$\Psi(y) = \sup\{xy - \Phi(x) : x \in [0, \infty)\}.$$

Then Ψ is the smallest function such that $\Phi(x) + \Psi(y) \geqslant xy$ for each x, y. This quantity is finite for each y because Φ is superlinear; for each $y \geqslant 0$, there exists $x(y)$ such that $\Phi(x(y)) \geqslant xy$, and thus the maximum of $xy - \Phi(x)$ is attained for $x \leqslant x(y)$. In particular, since Φ is continuous, the supremum is actually attained. Conversely, for each $x_0 \in [0, \infty)$, convexity implies there exists a largest y such that the line $y(x - x_0) + f(x_0) \leqslant f(x)$ for all $x \in [0, \infty)$. This means that $\Psi(y) = x_0 y - x_0$.

We note also that $\Psi(0) = 0$, and Ψ is increasing. Most importantly, the function is convex. Given any $y, z \in [0, \infty)$, and any $x \in [0, \infty)$,

$$x(\alpha y + (1 - \alpha)z) - \Phi(x) \leqslant \alpha(xy - \Phi(x)) + (1 - \alpha)(xz - \Phi(x))$$
$$\leqslant \alpha\Psi(y) + (1 - \alpha)\Psi(z).$$

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 16.26. *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 $\Phi(x) = yx - \Psi(y) \leqslant \Pi(x)$. $\qquad\qquad\square$

Given the Orlicz space $\Phi(L)$ for superlinear Φ, we can consider the Orlicz space $\Psi(L)$, where Ψ is the Young dual of Φ. The inequality $xy \leqslant \Phi(x) + \Psi(y)$, then

$$|f(x)g(x)| \leqslant \Phi(|f(x)|) + \Psi(|g(x)|),$$

so if $\|f\|_{\Phi(L)}, \|g\|_{\Psi(L)} \leqslant 1$, then

$$\left| \int f(x)g(x) \right| \leqslant \int |f(x)||g(x)| \leqslant \int \Phi(|f(x)|) + \int \Psi(|g(x)|) \leqslant 2.$$

Thus in general, we have

$$\left| \int f(x)g(x) \right| \leqslant 2\|f\|_{\Phi(L)}\|g\|_{\Psi(L)},$$

a form of Hölder's inequality. The duality between convex functions extends to a duality between the Orlicz spaces.

Theorem 16.27. *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. $\square$

Let us now consider some examples of duality.

Example. *If $\Phi(x) = x^p$, for $p \geqslant 1$, and $1 = 1/p + 1/q$, then it's Young dual Ψ satisfies*

$$\Psi(y) = \sup_{x \geqslant 0} xy - x^p = y^{1+q/p}/p^{q/p} - y^q/p^q = y^q[p^{-q/p} - p^{-q}].$$

Thus the Young dual corresponds, up to a constant, to the conjugate dual in the L^p spaces.

Example. *Suppose X has finite measure. If $\Phi(t) = e^t - 1$, then it's dual satisfies, for large y,*

$$\Psi(y) = \sup_{x \geqslant 0} xy - (e^x - 1)$$
$$= y \log y - (y - 1) \sim y \log y.$$

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 17

Sobolev Spaces

TODO: Treves, Chapter 31 for cool discussion of Sobolev spaces of negative index.

Previously, we discussed the rearrangement invariant spaces, which quantify properties of a function $f : X \to Y$ in terms of is 'width' and 'height'. Here we discuss the 'regularity' of functions, i.e. how gradually they change, and the related topic (from a Fourier analytic perspective) of how fast the function 'oscillates' (a quantity inversely proportional to the 'wavelength' of a function. Such properties are certainly not quantified in terms of rearrangement invariant properties, since change and oscillation necessarily imply some topological features of the spaces involved. Here are some canonical, intuitive examples. To introduce them, we fix some function $\phi \in C_c^\infty(\mathbb{R})$, which we intuitively think of as having some height H and width W:

- If $\phi \in C_c^\infty(\mathbb{R})$, then the function $f(t) = \phi(t)e^{iNt}$ is the canonical example of a function oscillating at a frequency N. It has height H and width W. However, it's derivative $f'(t) = \phi'(t)e^{iNt} + iN\phi(t)e^{iNt}$ for large N has height roughly proportional to $N \cdot H$ and width roughly proportional to W, and as we successively differentiate, the height grows even faster in N. This means that the function fails to be 'regular' for large N.

- Let $f(t) = N^{-s}\phi(t)e^{iNt}$, or let $f(t) = N^{-s}\phi(Nt)$. Both of these functions has height $N^{-s}H$ and width W, and oscillates at a frequency N. But unlike the last examples, the height (e.g. the L^∞ norm) of the function $f^{(k)}(t)$ is uniformly bounded in t provided that $k \leqslant s$. We therefore think of f as having 's degrees of regularity'.

- Let us consider the more complex function $f(x) = \phi(x)|x|^s \mathbb{I}_{x>0}$. One can break this function up into a sum $\sum_{n=0}^{\infty} \phi_n(x)|x|^s$, where $\phi_n(x) = \phi(2^n x) - \phi(2^{n+1}x)$. The function $f_n(x) = \phi_n(x)|x|^s$ behaves very similarily to the function $\phi_n(x)2^{-ns}$, an example considered in the last bullet with $N = 2^n$. In particular, the function oscillates at a frequency 2^n, and has s degrees of regularity. Thus we would imaging f itself has s degrees of regularity, in a certain sense, but is composed of many different frequency scales.

There are various norms that quantify height, width, frequncy scale, and regularity. The most common is the Sobolev norm, though there are various refinements, like the Hölder norms, Besov norms, Triebel-Lizorkin norms.

17.1 Hölder Spaces

Hölder Spaces are a simplified version of Sobolev spaces which account for the height, regularity, and oscillation of a function, but not the width. For $f : \mathbb{R}^d \to \mathbb{C}$, and an integer $s \geqslant 0$, we define

$$\|f\|_{C^s(\mathbb{R}^d)} = \max_{1 \leqslant i \leqslant s} \|D^i f\|_{L^\infty(\mathbb{R}^d)}.$$

Note that a C^s function might not necessarily lie have finite $C^s(\mathbb{R}^d)$ norm if it's derivatives are unbounded. We let C^s_{loc} denote the space of all C^s functions. For instance, $e^x \in C^\infty_{\mathrm{loc}}(\mathbb{R})$, but $e^x \notin C^\infty(\mathbb{R})$. Each of the space $C^s(\mathbb{R}^d)$, where s is finite, forms a Banach space. The space $C^\infty(\mathbb{R}^d)$ is a Fréchet space.

It is a useful heuristic that only the smallest and largest derivatives really matter when measuring regularity in a space, a fact also tied up to interpolation. For instance, we have the following result in Hölder spaces.

Theorem 17.1. *For any function $f : \mathbb{R}^d \to \mathbb{C}$,*

$$\|f\|_{C^s(\mathbb{R}^d)} \sim_{s,d} \|f\|_{L^\infty(\mathbb{R}^d)} + \|D^s f\|_{L^\infty(\mathbb{R}^d)}.$$

Proof. Certainly the right hand side is upper bounded by the left hand side. On the other hand, we perform a Taylor expansion, writing

$$\left| f(x+y) - f(x) - \sum_{j=1}^{s-1}(D^j f)(x)(y) \right| \lesssim \|D^s f\|_{L^\infty(\mathbb{R}^d)} \cdot |y|^s.$$

We have $|f(x + y) - f(x)| \leqslant 2\|f\|_{L^\infty(\mathbb{R}^d)}$. Thus

$$\left| \sum_{j=1}^{s-1} (D^j f)(x)(y) \right| \lesssim \|f\|_{L^\infty(\mathbb{R}^d)} + \|D^s f\|_{L^\infty(\mathbb{R}^d)} |y|^s.$$

Taking the maximum over $|y| = 1$, we have, by finite dimensionality of the space of polynomials of degree s,

$$\sup_{|y| \leqslant 1} \left| \sum_{j=1}^{s-1} (D^j f)(x)(y) \right| \sim_{d,s} \sum_{j=1}^{s-1} |(D^j f)(x)|,$$

Taking the supremum over $|y| \leqslant 1$ in the inequality above, applying this equivalence, and then taking the supremum over x, we conclude that

$$\sum_{j=1}^{s-1} \|D^j f\|_{L^\infty(\mathbb{R}^d)} \lesssim \|f\|_{L^\infty(\mathbb{R}^d)} + \|D^s f\|_{L^\infty(\mathbb{R}^d)}.$$

But this immediately verifies that

$$\|f\|_{C^s(\mathbb{R}^d)} \lesssim \|f\|_{L^\infty(\mathbb{R}^d)} + \|D^s f\|_{L^\infty(\mathbb{R}^d)}. \qquad \square$$

More generally, for any $0 \leqslant \alpha \leqslant 1$, we set $C^{s,\alpha}(\mathbb{R}^d)$ to be the set of all C^s functions such that $D^s f$ satisfies a Hölder condition of order α, i.e.

$$\|f\|_{C^{s,\alpha}(\mathbb{R}^d)} = \|f\|_{L^\infty(\mathbb{R}^d)} + \|D^s f\|_{L^\infty(\mathbb{R}^d)} + \sup_{x \neq y} \frac{D^s f(x) - D^s f(y)}{|x - y|^\alpha}.$$

For $\alpha > 1$, this norm is finite if and only if $D^s f$ is constant, which is why only the case $\alpha \leqslant 1$ is interesting. It is simple to see that $C^s(\mathbb{R}^d)$ is contained in $C^{s,\alpha}(\mathbb{R}^d)$ for each such α. We also have $C^{s,1}(\mathbb{R}^d) \subset C^{s+1}(\mathbb{R}^d)$, because of Taylor's formula.

Example. *To see the relation between the Hölder norm and the height, width, frequency scale, etc, consider a bump function $\phi \in C_c^\infty(\mathbb{R}^d)$, and let $f(t) = A\phi(t/R)e^{Nt}$. Provided that $R \geqslant 1/N$, we think of this function as having height A, width R, and frequency N (if $R < 1/N$, the frequency term does not have enough 'space' to oscillate). Then $\|f\|_{C^{s+\alpha}(\mathbb{R}^d)} \lesssim_{\phi,d,s} AN^{s+\alpha}$. Thus R is not measured at all.*

Lemma 17.2. *An element of $C^{s,1}(\mathbb{R}^d)$ is precisely an element $f \in C^s(\mathbb{R}^d)$ such that the distributional derivative $D^{s+1} f$ is a tensor with coefficients in $L^\infty(\mathbb{R}^d)$.*

Proof. It suffices to prove the case $s = 0$. If $f \in C^{0,1}(\mathbb{R}^d)$ then f is precisely a bounded, uniformly Lipschitz function. Such functions are absolutely continuous with uniformly bounded derivative defined almost everywhere, and thus the distributional derivative Df lies in L^∞. Conversely, if $f \in L^\infty(\mathbb{R}^d)$ and $f' \in L^\infty(\mathbb{R}^d)$, then we have a representation formula

$$f(x) = \int_0^x f'(y) \, dy,$$

which implies f is Lipschitz. □

17.2 Sobolev Spaces

We now consider the 'classical' Sobolev spaces $W^{s,p}(\mathbb{R}^d)$, also denoted $L^p_s(\mathbb{R}^d)$. A function f lies in $W^{s,p}(\mathbb{R}^d)$ if it's weak derivatives $D^\alpha f$, for $|\alpha| \leqslant s$, lie in $L^p(\mathbb{R}^d)$. We then define the norm

$$\|f\|_{W^{s,p}(\mathbb{R}^d)} = \sum_{j=0}^s \|D^j f\|_{L^p(\mathbb{R}^d)}.$$

Then $W^{s,p}(\mathbb{R}^d)$ is a Banach space, and $W^{s,\infty}(\mathbb{R}^d)$ is the same space as $C^{s-1,1}(\mathbb{R}^d)$. The functions $C_c^\infty(\mathbb{R}^d)$ are dense in $W^{s,p}(\mathbb{R}^d)$ for $p < \infty$, which means in practice when establishing estimates we need not worry about the technicalities of weak derivatives. For $p = \infty$, the closure of $C_c^\infty(\mathbb{R}^d)$ is the space $C_0^s(\mathbb{R}^d)$ of C^s functions whose derivatives of order at most s go to zero at ∞.

One useful property of Sobolev spaces is the *Sobolev embedding theorem*, which allows one to *trade regularity for integrability*.

Lemma 17.3. *Suppose* $1/q = 1/p - s/d$. *The following are continuous inclusions:*

- *For $p < d/s$ ($q < \infty$), we have $L^p_s(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$, and in particular,*

$$\|f\|_{L^q(\mathbb{R}^d)} \lesssim \|D^s f\|_{L^p(\mathbb{R}^d)}.$$

- *For $p = d/s$ ($q = \infty$), $L^p_s(\mathbb{R}^d) \to BMO(\mathbb{R}^d)$, and in particular*

$$\|f\|_{BMO(\mathbb{R}^d)} \lesssim \|D^s f\|_{L^p(\mathbb{R}^d)}.$$

- *For $p > d/s$, and $r + \alpha = s - d/p$, $L^p_s(\mathbb{R}^d) \to C^{r,\alpha}(\mathbb{R}^d)$, and in particular*

$$\|f\|_{C^{r,\alpha}(\mathbb{R}^d)} \lesssim \|D^s f\|_{L^p(\mathbb{R}^d)}.$$

Proof. To obtain this result, we must find a way to express f in terms of it's derivatives. We must recall some Fourier multiplier operators:

- The Riesz transforms

$$(R_j f)(x) = c_d \text{ p.v} \int \frac{y_j}{|y|^{d+1}} f(x - y) \, dy$$

  are Fourier multipliers with symbol $ix_j / |x|$. They are all singular integrals of Calderón-Zygmund type.

- The differentiation operators D_j are operators with symbol $2\pi i x_j$.

- For $0 < \alpha < d$, we have the Riesz potentials

$$(I_\alpha f)(x) = \frac{1}{c_\alpha} \int_{\mathbb{R}^d} |x - y|^{\alpha - d} f(y) \, dy,$$

  a Fourier multiplier with symbol $(2\pi |x|)^{-\alpha}$.

We then have, by the symbol calculus,

$$f = -I_1 \left\{ \sum_j R_j D_j f \right\}.$$

The Hardy-Littlewood-Sobolev inequality and the theory of singular integrals implies that provided $p > 1$ and $q < \infty$,

$$\|f\|_{L^q(\mathbb{R}^d)} = \left\| I_1 \left\{ \sum_j R_j D_j f \right\} \right\|_{L^q(\mathbb{R}^d)}$$

$$\lesssim \left\| \sum_j R_j D_j f \right\|_{L^p(\mathbb{R}^d)}$$

$$\lesssim \sum_j \|R_j D_j f\|_{L^p(\mathbb{R}^d)}$$

$$\lesssim \sum_j \|D_j f\|_{L^p(\mathbb{R}^d)}.$$

The case $q = \infty$ is proved similarily, using the fact that the operator I_α maps $L^{d/\alpha}(\mathbb{R}^d)$ into $\text{BMO}(\mathbb{R}^d)$ (one can prove this directly from a simple calculation).

We must prove the case $p = 1$ must be proved by other methods. Using the fundamental theorem of calculus, we write, for each j,

$$|f(x)| \leqslant \int_{\mathbb{R}} |(D_j f)(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_d)| \, dt$$
$$\leqslant |g_j(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d)|,$$

where

$$g_j(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d) = \int |(\nabla f)(x_1, \ldots, x_{j-1}, t, x_{j+1}, \ldots, x_d)| \, dt.$$

Now $\|g_j\|_{L^1(\mathbb{R}^{d-1})} = \|\nabla f\|_{L^1(\mathbb{R}^d)}$. By the Loomis-Whitney inequality,

$$\|f\|_{L^{\frac{d}{d-1}}(\mathbb{R}^d)} \leqslant \left\| g_1^{1/d} \cdots g_d^{1/d} \right\|_{L^{\frac{d}{d-1}}(\mathbb{R}^d)}$$
$$= \|g_1 \ldots g_d\|_{L^{\frac{1}{d-1}}(\mathbb{R}^d)}^{1/d}$$
$$\lesssim \prod_j \|g_j\|_{L^1(\mathbb{R}^{d-1})}^{1/d} = \|\nabla f\|_{L^1(\mathbb{R}^d)}.$$

For the case $p > d$, and $s = 1$, we have $r = 0$ and $\alpha = 1 - d/p$. Without loss of generality, it suffices to prove Hölder estimates at the origin. For a function g, we have

$$(I_1 g)(x) - (I_1 g)(0)$$
$$= \frac{1}{c_d} \int_{\mathbb{R}^d} \left(\frac{1}{|x - z|^{d-1}} - \frac{1}{|z|^{d-1}} \right) g(z) \, dz.$$

For $|z| \geqslant 2|x|$, we have a cancellation condition

$$\left| \frac{1}{|x - z|^{d-1}} - \frac{1}{|z|^{d-1}} \right| \lesssim_d \frac{|x|}{|z|^d},$$

which can be obtained by Taylor's theorem. This implies, together with Hölder's

334

inequality, that

$$\left| \int_{|z| \geqslant 2|x|} \left(\frac{1}{|x-z|^{d-1}} - \frac{1}{|z|^{d-1}} \right) g(z) \, dz \right|$$

$$\lesssim \int_{|z| \geqslant 2|x|} \frac{|x|}{|z|^d} |g(z)| \, dx$$

$$\lesssim \|g\|_{L^p(\mathbb{R}^d)} |x| \left(\int_{|z| \geqslant 2|x|} \left(\frac{1}{|z|^d} \right)^{p'} \right)^{1/p'}$$

$$= \|g\|_{L^p(\mathbb{R}^d)} |x| \left(\int_{2|x|}^{\infty} r^{d-1-dp'} \right)^{1/p'}$$

$$= \|g\|_{L^p(\mathbb{R}^d)} |x|^{1-d/p}.$$

On the other hand, we can also apply the triangle inequality, and Hölder's inequality, to show that, for $p > d$,

$$\left| \int_{|z| \leqslant 2|x|} \left(\frac{1}{|x-z|^{d-1}} - \frac{1}{|z|^{d-1}} \right) g(z) \, dz \right|$$

$$\lesssim \int_{|z| \leqslant 3|x|} \frac{1}{|z|^{d-1}} |g(z)| \, dx$$

$$\lesssim \|g\|_{L^p(\mathbb{R}^d)} \left(\int_{|z| \geqslant 2|x|} \left(\frac{1}{|z|^{(d-1)}} \right)^{p'} \right)^{1/p'}$$

$$= \|g\|_{L^p(\mathbb{R}^d)} \left(\int_0^{3|x|} r^{d-1-(d-1)p'} \right)^{1/p'}$$

$$= \|g\|_{L^p(\mathbb{R}^d)} |x|^{1-d/p}.$$

We can then put these estimates together to show that

$$|(I_1 g)(x) - (I_1 g)(0)| \lesssim |x|^{1-d/p} \|g\|_{L^p(\mathbb{R}^d)}.$$

Applying this estimate to the representation of f in terms of I_1 and the Riesz transforms $\{R_j\}$ yields the Hölder continuity bound

$$|f(x) - f(y)| \lesssim |x-y|^{1-d/p} \sum_j \|D_j f\|_{L^p(\mathbb{R}^d)}.$$

335

But now we can obtain the L^∞ bound for f easily, i.e. because

$$|f(0)| \lesssim \int_{|x| \leqslant 1} |f(0) - f(x)| + \int_{|x| \leqslant 1} |f(x)|$$

$$\lesssim \left(\sum_j \|D_j f\|_{L^p(\mathbb{R}^d)} \right) + \|f\|_{L^p(\mathbb{R}^d)}.$$

Now that we have proved the cases where $s = 1$, the cases for integers $s > 1$ follow by applying these results iteratively. For instance, if $p < d/2$, then for any $f \in L_2^p(\mathbb{R}^d)$, the case $s = 1$ implies $\nabla f \in L^r(\mathbb{R}^d)$ where $1/r = 1/p - 1/d$, and then the case $s = 1$ implies that $f \in L^q(\mathbb{R}^d)$. $\qquad \square$

In the case of $L_d^1(\mathbb{R}^d)$, we can actually obtain an $L^\infty(\mathbb{R}^d)$ inclusion rather than just an inclusion in $\mathrm{BMO}(\mathbb{R}^d)$.

Lemma 17.4. *We have a continuous inclusion $W^{d,1}(\mathbb{R}^d) \subset C_0(\mathbb{R}^d)$, and*

$$\|f\|_{L^\infty(\mathbb{R}^d)} \lesssim \|f\|_{W^{d,1}(\mathbb{R}^d)}.$$

Proof. If $f \in C_c^\infty(\mathbb{R}^d)$, we may apply the fundamental theorem of calculus to write

$$f(x) = \int D^1 f(x + te_1) \, dt = \cdots = \int_{t_1 \leqslant x_1} \cdots \int_{t_d \leqslant x_d} D^{1,\ldots,1} f(x + t_1 e_1 + \cdots + t_n e_n).$$

But this means

$$\|f\|_{L^\infty(\mathbb{R}^d)} \leqslant \|D^{1,\ldots,1} f\|_{L^1(\mathbb{R}^d)}.$$

For a given x, let $Q(x) = \{t : t_1 \leqslant x_1, \ldots, t_d \leqslant x_d\}$. If $\{x^j\}$ is a sequence in $\mathbb{R}^d$ such that $x_i^j \to -\infty$ for some i, then we can use Fatou's Lemma and the dominated convergence theorem to conclude that

$$\left| \lim_{j \to \infty} f(x^j) \right| \leqslant \lim_{j \to \infty} \int_{Q(x^j)} |D^{1,\ldots,1} f(u)| \, du \leqslant \int_{\lim_j Q(x^j)} |D^{1,\ldots,1} f(u)| = 0,$$

since $\lim_j Q(x^j) = \varnothing$. Using the opposite formula for the fundamental theorem of calculus proves the same result if $x_i^j \to \infty$ for some i. But from this, we conclude that for any sequence $\{x^j\}$ with $|x^j| \to \infty$, any subsequence of $\{f(x^j)\}$ possesses a further subsequence converging to zero. But this implies $\lim_j f(x^j) = 0$, and thus $f \in C_0^\infty(\mathbb{R}^d)$. $\qquad \square$

The fact that the Bessel potentials $\{I_\alpha\}$ are integrable at the origin makes them useful as smoothing operators, but their global behaviour runs into problems since they are not globally integrable. To fix this, in light of the fact that $I_\alpha = (-\Delta)^{\alpha/2}$, we define the *Bessel potentials* $J_\alpha = (I - \Delta)^{\alpha/2}$; on the frequency side of things, the Bessel potentials do not vanish near the origin, which leads to less singular behaviour at low frequencies.

We want to calculate the convolution kernel corresponding to the potentials $\{J_\alpha\}$. One useful trick is to apply the subordination principle (TODO: More exposition here) to write

$$(1 + 4\pi^2|\xi|^2) = \frac{(4\pi)^{\alpha/2}}{\Gamma(\alpha/2)} \int_0^\infty e^{-\frac{\delta}{4\pi}(1+4\pi^2|\xi|^2)} \delta^{\alpha/2} \frac{d\delta}{\delta}.$$

It is more simply to take the inverse Fourier transform of the right hand side, we thus define the *Bessel kernels*

$$G_\alpha(x) = \frac{1}{(4\pi)^{\alpha/2}\Gamma(\alpha/2)} \int_0^\infty e^{-\pi|x|^2/\delta} e^{-\delta/4\pi} \delta^{\frac{\alpha-d}{2}} \frac{d\delta}{\delta}.$$

Fubini's theorem implies

$$\|G_\alpha\|_{L^1(\mathbb{R}^d)} \lesssim \frac{1}{(4\pi)^{\alpha/2}\Gamma(\alpha/2)} \int_0^\infty \delta^{\alpha/2} e^{-\delta/4\pi} \frac{d\delta}{\delta} = 1.$$

We can also expand

$$\frac{|x|^{-n+\alpha}}{\gamma(\alpha)} = \frac{1}{(4\pi)^{\alpha/2}} \frac{1}{\Gamma(\alpha/2)} \int_0^\infty e^{-\pi|x|^2/\delta} \delta^{\frac{\alpha-d}{2}} \frac{d\delta}{\delta},$$

which tells us that, for $|x| \leqslant 1$, and as $x \to 0$,

$$G_\alpha(x) = \frac{|x|^{\alpha-d}}{\gamma(\alpha)} + o(|x|^{\alpha-d}).$$

On the other hand, for $|x| \geqslant 1$,

$$|G_\alpha(x)| \lesssim e^{-c|x|},$$

so we get very rapid decay away from the origin.

Lemma 17.5. *For $\alpha > 0$, there exists finite measures μ_α, ν_α, and λ_α such that*

$$\hat{\mu}_\alpha(\xi) = \frac{(2\pi|\xi|)^\alpha}{(1 + 4\pi^2|\xi|^2)^{\alpha/2}}$$

and

$$(1 + 4\pi^2|\xi|^2)^{\alpha/2} = \hat{\nu}_\alpha(\xi) + (2\pi|\xi|)^\alpha \hat{\lambda}_\alpha(\xi).$$

Thinking in terms of the characterization of Fourier multipliers bounded on $L^p(\mathbb{R}^d)$ for all $1 \leqslant p \leqslant \infty$, this result says that for all such p,

$$\|I_\alpha f\|_{L^p(\mathbb{R}^d)} \lesssim \|J_\alpha f\|_{L^p(\mathbb{R}^d)},$$

and

$$\|J_\alpha f\|_{L^p(\mathbb{R}^d)}$$

Then for $\alpha > 0$, $G_\alpha \in L^1(\mathbb{R}^d)$, There is also a Bessel function representation (TODO), hence the name.

The cases $L_s^p(\mathbb{R}^d) = W^{s,2}(\mathbb{R}^d)$ are also denoted by $H^s(\mathbb{R}^d)$, and are of special importance because they are *Hilbert spaces*, i.e. if we define

$$\langle f, g \rangle_{H^s(\mathbb{R}^d)} = \sum_{|\alpha| \leqslant s} \int (D^\alpha f)(x) \overline{(D^\alpha g)(x)} \, dx,$$

then this inner product induces the norm on $H^s(\mathbb{R}^d)$.

17.3   lipschitz

We can obtain some alternate results using *Lipschitz spaces*. Given a function f, define the *difference operator*

$$(\Delta_h f)(x) = f(x + h) - f(x).$$

Fix a norm space X, and define a normalized kth modulus of continuity $\omega_{k,\alpha}$ by setting

$$\omega_{k,\alpha}(h) = |h|^{-\alpha} \|\Delta_h^k f\|_X.$$

Then we define the *Lipschitz norm*

$$\|f\|_{\Lambda_{X,p}^\alpha(\mathbb{R}^d)} = \|\omega_{k,\alpha}\|_{L^p(\mathbb{R})}.$$

The value k is irrelevant; as long as we take $k > \alpha$ we obtain equivalent norms.

Lemma 17.6. *Fix $\alpha \geqslant 0$. If $k, k' > \alpha$ and X is a translation-invariant norm space, then for any function f, with associated kth and k'th moduli of continuity $\omega_{k,\alpha}$ and $\omega_{k',\alpha}$, then*

$$\|\omega_{k,\alpha}\|_{L_h^p(\mathbb{R})X_x(\mathbb{R}^d)} \sim_{k,k'} \|\omega_{k',\alpha}\|_{L_h^p(\mathbb{R})X_x(\mathbb{R}^d)}.$$

Proof. Without loss of generality, assume $k' > k$. We note that, using the triangle inequality and the translation invariance of X,

$$
\begin{aligned}
\|\Delta_h^{k+1} f(x)\|_X &= \|\Delta_h^k f(x+h) - \Delta_h^k f(x)\|_X \\
&\leqslant \|\mathrm{Trans}_h \Delta_h^k f\|_X + \|\Delta_h^k f\|_X \\
&= 2\|\Delta_h^k f\|_X.
\end{aligned}
$$

Iterating this argument shows that

$$\|\Delta_h^{k'} f\|_{X_x(\mathbb{R}^d)} \leqslant 2^{k'-k}\|\Delta_h^k f\|_{X_x(\mathbb{R}^d)}.$$

Thus

$$\|\omega_{k,\alpha}\|_{L^p(\mathbb{R})} \leqslant 2^{k'-k}\|\omega_{k',\alpha}\|_{L^p(\mathbb{R})}.$$

Now we need to prove the opposite inequality. Without loss of generality, we may assume $k' = k+1$. We rely on some polynomial identities. For each k, there exists constants $\{a_j\}$, depending on k, such that

$$(T+1)^k = (2 + (T-1))^k = 2^k + (T-1)\sum_{j=0}^{k-1} a_j T^j.$$

Now

$$(T^2 - 1)^k = (T-1)^k(T+1)^k = 2^k(T-1)^k + (T-1)(T+1)^k \sum_{j=0}^{k-1} a_j T^j.$$

Rearranging, we have

$$2^k(T-1)^k = (T^2-1)^k - (T-1)^{k+1}\sum_{j=0}^{k-1} a_j T^j.$$

Iterating this identity, we conclude that

$$2^{mk}(T-1)^k = (T^{2^m} - 1)^k - \sum_{j=0}^{m-1} 2^{(m-i)k}(T^{2^i} - 1)^{k+1}\sum_j a_j T^{2^i j}.$$

339

If we interpret T as the operator Trans_h, and apply both sides of the identity to a function f, then the identity reads

$$2^{mk}\Delta_h^k f = \Delta_{2^m h}^k f - \sum_{i=0}^{m-1}\sum_{j=0}^{k-1} a_j 2^{(m-i)k}(\Delta_{2^i h}^{k+1} f)(x + 2^i jh).$$

But this means that, using the translation invariance of X,

$$2^{mk}\omega_{k,\alpha}(h) \leqslant 2^{\alpha m}\omega_{k,\alpha}(2^m h) + c_k 2^{mk}\sum_{i=0}^{m-1} 2^{i(\alpha-k)}\omega_{k+1,\alpha}(2^i h),$$

where $c_k = \sum |a_j|$. But this means that

$$2^{mk}\|\omega_{k,\alpha}\|_{L^p(\mathbb{R})} \leqslant 2^{(\alpha-1/p)m}\|\omega_{k,\alpha}\|_{L^p(\mathbb{R})} + c_k 2^{mk}\left(\sum_{i=0}^{m-1} 2^{i(\alpha-k-1/p)}\right)\|\omega_{k+1,\alpha}\|_{L^p(\mathbb{R})}.$$

Dividing by 2^{mk} on both sides, and taking $m \to \infty$ yields that, for $k > \alpha - 1/p$,

$$\|\omega_{k,\alpha}\|_{L^p(\mathbb{R})} \leqslant c_k\left(\sum_{i=0}^{\infty} 2^{i(\alpha-k-1/p)}\right)\|\omega_{k+1,\alpha}\|_{L^p(\mathbb{R})}.$$

The infinite sum is finite for $k > \alpha - 1/p$. $\qquad\square$

340

Chapter 18

Sobolev Spaces

Let Ω be an open subset of $\mathbb{R}^d$. 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 $\phi \in C_c^\infty(\Omega)$,

$$\int_\Omega f_\alpha(x)\phi(x) \, dx = (-1)^{|\alpha|} \int_\Omega f(x)\phi_\alpha(x) \, dx.$$

341

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 $\mathbb{R}^d$, 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

$$\int_{\varepsilon < |x| \leqslant 1} \phi_i(x)u(x) = \int_{|x| = \varepsilon} \phi(x)u(x)v_i\, dS + \int_{\varepsilon < |x| \leqslant 1} \frac{s\phi(x)x_i}{|x|^{s+2}}\, dx,$$

where v_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 18.1. *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 $\phi \in C_c^\infty(\Omega)$,

$$(-1)^{|\alpha|} \int \phi_\alpha(x) v(x) \, dx = \lim_{n \to \infty} (-1)^{|\alpha|} \phi_\alpha u_n(x) \, dx$$

$$= \lim_{n \to \infty} \int \phi(x)(D^\alpha u_n)(x) \, dx$$

$$= \int \phi(x) v_\alpha(x) \, dx.$$

Thus $W^{k,p}(\Omega)$ is complete. $\square$

18.1 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 $\mathbb{R}^n$ 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 $u \, \mathbb{I}_\Omega$ on $\mathbb{R}^d$. Then u^ε is a smooth function on $\mathbb{R}^d$ 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 $D^\alpha(u \, \mathbb{I}_\Omega)$ does not behave like a locally integrable function. This is the only problem, however.

Theorem 18.2. *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

$$((D^\alpha u) * \eta_\varepsilon)(x) = (u_\alpha \, \mathbb{I}_\Omega * \eta_\varepsilon)(x),$$

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 $u_\alpha \mathbb{I}_\Omega * \eta_\varepsilon$ converges to $u_\alpha \mathbb{I}_\Omega$ in $L^p(\mathbb{R}^d)$ 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. $\qquad\square$

If we are a little more careful, then we can fully approximate elements of $W^{k,p}(\Omega)$ by smooth functions on U.

Theorem 18.3. $C_c^\infty(\Omega) \cap W^{k,p}(\Omega)$ *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 $W^{k,p}(\Omega)$. $\qquad\square$

Approximation by elements of $C^\infty(\overline{\Omega})$ requires some more care, and additional assumptions on the behaviour of $\partial\Omega$.

Chapter 19

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*. There are both complex methods of interpolation, and real methods of interpolation. These methods are interchangable, *most of the time*, except when dealing with the boundedness of operators at endpoints. Sometimes complex interpolation is more useful, and sometimes real interpolation works better. So it is useful to know both types of interpolation.

19.1   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 $L^{p_0,\infty}(X) \cap L^{p_1,\infty}(X) \subset L^{p_\theta}(X)$.

19.2   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 19.1 (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*

$$M(s) = \sup_{s \in \mathbb{R}} |f(s + it)|$$

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

$$|f(it)| \leqslant A_0 \quad \text{and} \quad |f(1 + it)| \leqslant A_1 \quad \text{for all } t \in \mathbb{R},$$

then for any $s \in [a, b]$ and $t \in \mathbb{R}$,

$$|f(s + it)| \leqslant A_0^{1-s} A_1^s.$$

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 $|\mathrm{Im}(z)| \to \infty$,

$$|f(z)u_\varepsilon(z)| \leqslant A e^{Be^{(\pi-\delta)|z|} - C\varepsilon^2 e^{(\pi-\varepsilon)|z|}} \to 0.$$

Applying the previous case to the function $|f(z)u_\varepsilon(z)|$, we conclude that for any $\varepsilon > 0$,

$$|f(z)| \leqslant \frac{1}{|u_\varepsilon(z)|}.$$

Thus

$$|f(z)| \leqslant \lim_{\varepsilon \to 0} \frac{1}{|u_\varepsilon(z)|} = 1,$$

which completes the proof. □

Remark. The function $e^{-ie^{\pi i s}}$ 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 $t \in \mathbb{R}$,*

$$|f(z)| \leqslant C_1(1 + |z|)^\alpha.$$

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

$$|f(z)| \lesssim |1 + z|^\alpha \lesssim (1 + |z|)^\alpha. \qquad \square$$

TODO: Is Lemma 20.1 of Treves Distribution Theory an example of interpolation?

19.3 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.

19.4 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 19.2 (Riesz-Thorin). *Let $p_0, p_1 \in (0, \infty]$ and $q_0, q_1 \in [1, \infty]$. Suppose that*

$$\|Tf\|_{L^{q_0}(Y)} \leqslant A_0 \|f\|_{L^{p_0}(X)} \quad and \|Tf\|_{L^{q_1}(Y)} \leqslant A_1 \|f\|_{L^{p_1}(X)}.$$

Then for any $\theta \in (0, 1)$, if

$$1/p_\theta = (1 - \theta)/p_0 + \theta/p_1 \quad and \quad 1/q_\theta = (1 - \theta)/q_0 + \theta/q_1,$$

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 $t \in \mathbb{R}$, we have

$$\|F_0^{1-it} F_1^{it} a\|_{L^{p_0}(X)} = \|F_0\|_{L^{p_0}(X)} = 1,$$

350

$$\|G_0^{1-it}G_1^{it}b\|_{L^{q_0}(Y)} = \|G_0\|_{L^{q_0}(X)} = 1.$$

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 $t \in \mathbb{R}$. 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 $f, g \in L^1(\mathbb{R}^d)$, we can define an integrable function $f * g \in L^1(\mathbb{R}^d)$ almost everywhere by the integral formula*

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

If $f \in L^1(\mathbb{R}^d)$ and $g \in L^p(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, for some $p \geqslant 1$, then Minkowski's integral inequality implies

$$\|f * g\|_p = \left(\int |(f * g)(x)|^p \ dx\right)^{1/p} \leqslant \int \left(\int |f(y)g(x - y)|^p dx\right)^{1/p} dy$$

$$= \int |f(y)|\|g\|_{L^p(\mathbb{R}^d)} = \|f\|_{L^1(\mathbb{R}^d)}\|g\|_{L^p(\mathbb{R}^d)}.$$

Hölder's inequality implies that if $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$, where p and q are conjugates of one another, then

$$\left|\int f(y)g(x - y) \ dy\right| \leqslant \int |f(y - x)||g(x)| \leqslant \|f\|_{L^p(\mathbb{R}^d)}\|g\|_{L^q(\mathbb{R}^d)}.$$

Thus we have the bound

$$\|f * g\|_{L^\infty(\mathbb{R}^d)} \leqslant \|f\|_{L^p(\mathbb{R}^d)}\|g\|_{L^q(\mathbb{R}^d)}.$$

Now that these mostly trivial results have been proved, we can apply convolution. For each $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$, we have a convolution operator $T : L^1(\mathbb{R}^d) \to$

351

$L^1(\mathbb{R}^d)$ *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* $L^r(\mathbb{R}^d)$ *to* $L^{p/\theta}(\mathbb{R}^d)$ *with operator norm one. Reparameterizing gives* Young's convolution inequality. *Note that we never really used anything about* $\mathbb{R}^d$ *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* $\mathbb{E}(\cdot|\Sigma) : L^2(X, \Sigma_0) \to L^2(X, \Sigma)$, *which we call the* conditional expectation operator. *The properties of the projection operator imply that for any* $f, g \in L^2(X, \Sigma_0)$,

$$\int \mathbb{E}(f|\Sigma)\overline{g} = \int f\overline{g} = \int \mathbb{E}(f|\Sigma)\overline{\mathbb{E}(g|\Sigma)}.$$

If $g \in L^2(X, \Sigma)$, *then*

$$\int \mathbb{E}(f|\Sigma)\overline{g} = \int f\overline{g}.$$

This gives a full description of $\mathbb{E}(f|\Sigma)$. *In particular, if* $u \in L^\infty(X, \Sigma_0)$, *then for each* $g \in L^2(X, \Sigma)$

$$\int \mathbb{E}(uf|\Sigma)\overline{g} = \int f[u\overline{g}] = \int u\,\mathbb{E}(f|\Sigma)\overline{g}.$$

Since this is true for all $g \in L^2(X, \Sigma)$, *we find* $\mathbb{E}(uf|\Sigma) = u\,\mathbb{E}(f|\Sigma)$. *Moreover, if* $0 \leqslant f \leqslant g$, *then* $\mathbb{E}(f|\Sigma) \leqslant \mathbb{E}(g|\Sigma)$. *This is easy to see because if* $f \geqslant 0$, *and* $F = \{x : \mathbb{E}(f|\Sigma) < 0\}$, *then if* $|F| \neq 0$,

$$0 > \int \mathbb{E}(f|\Sigma)\,\mathbb{I}_F = \int f\,\mathbb{I}_F \geqslant 0.$$

Thus $|F| = 0$, *and so* $\mathbb{E}(f|\Sigma) \geqslant 0$ *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 $\mathbb{I}_E \in L^2(X)$, so

$$|\mathbb{E}(f|\Sigma)|\,\mathbb{I}_E = |\mathbb{E}(\mathbb{I}_E\,f|\Sigma)| \leqslant \mathbb{E}(\mathbb{I}_E\,|f||\Sigma) \leqslant \|f\|_\infty \mathbf{E}(\mathbb{I}_E\,|\Sigma) = \|f\|_\infty\,\mathbb{I}_E\,.$$

Since Σ is a sigma finite sigma algebra, we can take $E \to \infty$ to conclude $\|\mathbb{E}(f|\Sigma)\|_\infty \leqslant \|f\|_\infty$. 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 $|E| < \infty$,

$$\int |\mathbb{E}(f|\Sigma)|\,\mathbb{I}_E \leqslant \int \mathbb{E}(|f||\Sigma)\,\mathbb{I}_E = \int_E |f|\,\mathbb{I}_E \leqslant \|f\|_1.$$

Since Σ is σ finite, we can take $E \to \infty$ to conclude $\|\mathbb{E}(f|\Sigma)\|_1 \leqslant \|f\|_1$. Thus the Riesz interpolation theorem implies that for each $1 \leqslant p \leqslant \infty$, $\|\mathbb{E}(f|\Sigma)\|_p \leqslant \|f\|_p$.

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 $g \in L^\infty(\Sigma)$, $\mathbb{E}(fg|\Sigma) = g\,\mathbb{E}(f|\Sigma)$. This is because for any $E \in \Sigma$ with $|E| < \infty$, $\mathbb{E}(f\,\mathbb{I}_E\,|\Sigma) = \mathbb{I}_E\,\mathbb{E}(f|\Sigma)$ is uniquely defined since $f\,\mathbb{I}_E \in L^2(\Sigma_0)$, 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 $\Sigma_1, \Sigma_2, \ldots$ 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 $f \in L^p(\Sigma_0)$, $\lim_{i \to \infty} \mathbb{E}(f|\Sigma_i) = \mathbb{E}(f|\Sigma_\infty)$. This is because the operators $\{\mathbb{E}(\cdot|\Sigma_i)\}$ 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 19.3 (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*

$$\|Tf\|_{L^{q_0}(X, v_0)} \leqslant A_0 \|f\|_{L^{p_0}(X, w_0)} \quad and \quad \|Tf\|_{L^{q_1}(X, v_1)} \leqslant A_1 \|f\|_{L^{p_1}(X, w_0)}.$$

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 $t \in \mathbb{R}$

$$\|T_{it} f\|_{L^{q_0}(Y)} = \|Tf\|_{L^{q_0}(Y, w_0)} \leqslant A_0 \|f v_0^{-1/p_0}\|_{L^{p_0}(X, v_0)} = A_0 \|f\|_{L^{p_0}(X)}.$$

354

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. □

19.5 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,

$$\|T(\mathbb{I}_E)\|_{L^q(Y)} \lesssim |E|^{1/p}.$$

The equivalence is proven by breaking any sub-step function f with height H and width W into a dyadic sum $\sum_{k=1}^{\infty} H\mathbb{I}_{E_k} 2^{-k}$, 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,

$$\|T(\mathbb{I}_E)\|_{L^{q,\infty}(Y)} \lesssim |E|^{1/p}.$$

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 16.13.

Lemma 19.4. *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*

$$\int_{F'} |T(\mathbb{I}_E)| \lesssim_\alpha |E|^{1/p}|F|^{1-1/q}.$$

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 19.5. *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

$$\int_{F_i} |T(\mathbb{I}_E)| \lesssim |E|^{1/p_i}|F_i|^{1-1/q_i}.$$

If we let $F_\theta = F_0 \cap F_1$, then $|F_\theta| \geqslant |F|/2$, and for each i,

$$\int_{F_\theta} |T(\mathbb{I}_E)| \lesssim |E|^{1/p_i}|F_\theta|^{1-1/q_i}.$$

Scalar interpolation implies

$$\int_{F_\theta} |T(\mathbb{I}_E)| \lesssim |E|^{1/p_\theta}|F_\theta|^{1-1/q_\theta},$$

and thus we have shown

$$\|T(\mathbb{I}_E)\|_{q_\theta,\infty} \lesssim |E|^{1/p_\theta}.$$

This is sufficient to show T is restricted weak type (p_θ, q_θ). $\qquad\square$

Theorem 19.6 (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

$$\|A\|_{l^r(\mathbb{Z})}, \|B\|_{l^{r'}(\mathbb{Z})} \lesssim 1.$$

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

$$\int |Tf||g| \leqslant \sum_{i,j} \int |Tf_i||g_j|$$

$$\lesssim \sum_{i,j} H_i H'_j \min_{k \in \{0,1\}} \left[2^{i/p_k + j(1-1/q_k)} \right]$$

$$\lesssim \sum_{i,j} A_i B_j \min_{k \in \{0,1\}} \left[2^{i(1/p_k - 1/p_\theta) + j(1/q_\theta - 1/q_k)} \right].$$

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

$$\sum_{i,j} A_i B_j \min_{k \in \{0,1\}} \left[2^{i(1/p_k - 1/p_\theta) + j(1/q_\theta - 1/q_k)} \right] \sim \sum_{k=-\infty}^{\infty} \min(2^{\varepsilon k}, 2^{-\varepsilon k}) \sum_i A_i B_{k - \lfloor i/\lambda \rfloor}.$$

357

Applying Hölder's inequality,

$$\sum_i A_i B_{k-\lfloor i/\lambda \rfloor} \leqslant \|A\|_{l^r(\mathbb{Z})} \left(\sum_i |B_{k-\lfloor i/\lambda \rfloor}|^{r'} \right)^{1/r'}$$

$$\lesssim \lambda^{1/r'} \|A\|_{l^r(\mathbb{Z})} \|B\|_{l^{r'}(\mathbb{Z})} \lesssim 1.$$

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 20

Basic Integral Operator Estimates

We now consider a very general class of operators, which can be seen as the infinite dimensional analogue of matrix multiplication, replacing summation over coordinates by integration. We fix two measure spaces X and Y, and consider a function (or, more generally, if X and Y are smooth manifolds, distributions) $K : X \times Y \to \mathbb{C}$, which we call a *kernel*. From this kernel, if K is suitably regular, 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 behaviour of the operator T_K.

Example. *Let $X = \{1, \ldots, N\}$ and $Y = \{1, \ldots, M\}$, 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 $v \in \mathbb{R}^N$ 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 in the standard basis, T_K is just matrix multiplication by A.

Example. *Let $X = Y = \mathbb{R}^d$. Then the operator with the kernel $K(x, y) = e^{2\pi i x \cdot y}$, i.e., the operator*

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

This is just the Fourier transform in disguise.

In the case of the Fourier transform (and what will be true for most kernel operators defined on infinite measure spaces), the integral formula can only be interpreted in the sense of a Lebesgue integral if the input function f lies in $L^1(\mathbb{R}^d)$. There are several methods to define T as an operator on more general families of functions, which require interpreting the formula in more general ways than is allowed by the Lebesgue integral:

- We can work with a dense subclass of functions, and then prove bounds which enable us to define the operator by taking limits of inputs. For instance, the classical Hausdorff-Young inequality shows that if $1 \leqslant p \leqslant 2$, then for any $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$,

$$\|Tf\|_{L^{p^*}(\mathbb{R}^d)} \leqslant \|f\|_{L^p(\mathbb{R}^d)}.$$

  The Hahn-Banach theorem thus implies that for each $1 \leqslant p \leqslant 2$, T extends uniquely to an operator from $L^p(\mathbb{R}^d)$ to $L^{p^*}(\mathbb{R}^d)$. To interpret Tf for a general input $f \in L^p(\mathbb{R}^d)$, we therefore consider a sequence $\{f_n\}$ in $L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$, and then
$$Tf = \lim_{n \to \infty} Tf_n,$$
  where the limit is taken in $L^{p^*}(\mathbb{R}^d)$.

- Another method is to replace the kernel K by a more regular kernel, and then to prove uniform estimates on this family. For instance, one may study the truncated Fourier integrals

$$T_R f(y) = \int_{|x| \leqslant R} e^{2\pi i x \cdot y} f(x) \, dx,$$

  which are well-defined in the Lebesgue sense for any $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. We will see in this chapter that, uniformly in R, for $1 \leqslant p \leqslant 2$,

$$\|T_R f\|_{L^{p^*}(\mathbb{R}^d)} \leqslant \|f\|_{L^p(\mathbb{R}^d)}.$$

  It then follows by the uniform boundedness theorem that for $f \in L^p(\mathbb{R}^d)$, $Tf = \lim_{R \to \infty} T_R f$, and the right hand side can be interpreted as a Lebesgue integral.

- We can apply a duality argument to obtain a 'distributional definition' of the operator T. The operator T is continuous on the class $\mathcal{S}(\mathbb{R}^d)$ of Schwartz

functions on $\mathbb{R}^d$, and so the adjoint operator T^* is continuous. A simple calculation shows that

$$(T^*f)(x) = (Tf)(-x).$$

Thus we can define T on the more general family of tempered distributions $\mathcal{S}(\mathbb{R}^d)^*$, such that for any tempered distribution u, Tu is the tempered distribution which maps $\phi \in \mathcal{S}(\mathbb{R}^d)$ to $\langle u, T^*\phi \rangle$.

20.1 Schur's Lemma

It is a useful heuristic that determining the boundedness of the operator T mapping *from $L^1(X)$*, or *into $L^\infty(Y)$* is almost always trivial. This is one motivation for introducing the intermediate L^p norms, since these norms enable us to extract more features out of the kernel operator K, in particular, the *oscillation*, or *cancellation*, that occurs in the kernel.

In this section, we introduce a method that completely characterizes mapping from $L^1(X)$, or into $L^\infty(Y)$, and also, if applied in a clever enough manner, is optimal for bounding the behaviour of any L^p mapping properties *if one is unwilling to take into account oscillation or cancellation*. In particular, we will reduce our analysis to non-negative kernels, i.e. considering the sublinear operator S_K defined by setting

$$(S_K f)(y) = \int_X |K(x,y)||f(x)| \, dx.$$

The advantage of working in this setting is that we do not need to care about the regularity of our inputs: the flexibility of the theory of non-negative Lebesgue integrals ensures $S_K f$ is defined for *any* measurable function f, though possibly takes infinite values at various points. Our first result gives a tight bound from $L^1(X)$.

Theorem 20.1. *Fix $q \geqslant 1$. Then*

$$\|S_K f\|_{L^q(Y)} \leqslant \|K\|_{L^\infty(X)L^q(Y)} \|f\|_{L^1(X)}.$$

Thus $T_K f(y)$ are well defined by a Lebesgue integral for almost every $y \in Y$, and

$$\|T_K f\|_{L^q(Y)} \leqslant \|K\|_{L^\infty(X)L^q(Y)} \|f\|_{L^1(X)}.$$

Proof. The proof is a simple consequence of Minkowski's inequality, i.e.

$$\|S_K f\|_{L^q(Y)} = \|Kf\|_{L^q(Y)L^1(X)} \leqslant \|Kf\|_{L^1(X)L^q(Y)} \leqslant \|K\|_{L^\infty(X)L^q(Y)} \|f\|_{L^1(X)}. \qquad \square$$

Remark. For many operators, this result gives tight estimates for $L^1(X) \to L^p(Y)$ bounds. For instance, suppose

$$K = \sum_{i=1}^{N} \sum_{j=1}^{M} a_{ij} \mathbb{I}_{E_i \times F_j}$$

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 $x \in E_i$,

$$\left(\int |K(x,y)|^q \, dy \right)^{1/q} = \left(\sum_{j=1}^{M} |a_{ij}|^q |F_j| \right)^{1/q} = \|K\|_{L^\infty(X)L^q(Y)}.$$

If $f = \mathbb{I}_{E_i}$, then $\|f\|_{L^1(X)} = |E_i|$, and $T_K f = \sum_{j=1}^{M} a_{ij} \mathbb{I}_{F_j}$, so

$$\|T_K f\|_{L^q(Y)} = \left(\sum_{j=1}^{M} |a_{ij}|^q |F_j| \right)^{1/q} = \|K\|_{L^\infty(X)L^q(Y)} \|f\|_{L^1(X)}.$$

Thus we conclude that for a certain 'dense' family of kernels K, the inequality above is tight. This 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. It is difficult to come up with a general statement of this form, because we require some regularity in the kernel to even formulate the operators under study. But by taking monotone limits we can at least justify that the norm of S_K from $L^1(X)$ to $L^q(Y)$ is given by $\|K\|_{L^\infty(X)L^q(Y)}$.

A dual statement trivializes the analysis of bounds from $L^p(X)$ to $L^\infty(Y)$.

Theorem 20.2. *Suppose* $1 \leqslant p \leqslant \infty$. *Then*

$$\|S_K f\|_{L^\infty(Y)} \leqslant \|K\|_{L^\infty(Y)L^{p^*}(X)} \|f\|_{L^p(X)}.$$

Thus if $\|K\|_{L^\infty(Y)L^{p^*}(X)} < \infty$, *then* $T_K f(y)$ *is well defined for almost every* $y \in Y$, *and*

$$\|T_K f\|_{L^\infty(Y)} \leqslant \|K\|_{L^\infty(Y)L^{p^*}(X)} \|f\|_{L^p(X)}.$$

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

$$\|S_K f\|_{L^\infty(Y)} = \|Kf\|_{L^\infty(Y)L^1(X)} \leqslant \|K\|_{L^\infty(Y)L^{p^*}(X)} \|f\|_{L^p(X)}. \qquad \square$$

Though trivial, the two kernel bounds can often be applied together with an interpolation argument to give more sophisticated bounds.

Theorem 20.3 (Schur's Test). *Fix $1 \leqslant r \leqslant \infty$, and suppose that*

$$\|K\|_{L^\infty(X)L^r(Y)} \leqslant A$$

and

$$\|K\|_{L^\infty(Y)L^r(X)} \leqslant B.$$

Then if $1 \leqslant p \leqslant q \leqslant \infty$ satisfy $1/p + 1/r = 1/q + 1$, then for $f \in L^p(X)$, $(T_K f)(y)$ is well defined by an absolutely convergent integral for almost every y, and

$$\|T_K f\|_{L^q(Y)} \leqslant A^{r/q} B^{1-r/q} \|f\|_{L^p(X)}.$$

Proof. The previous two results imply that S_K is bounded from $L^1(X)$ to $L^r(Y)$ and from $L^{r^*}(X)$ to $L^\infty(Y)$. Real interpolation (we cannot use complex interpolation since S_K is sublinear) shows that S_K is bounded from $L^p(X)$ to $L^q(Y)$ for the p and q satisfying the conditions above. This means $S_K f(x)$ is well defined for almost every $x \in X$, so 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

$$\|T_K f\|_{L^r(Y)} \leqslant A\|f\|_{L^1(X)}$$

and

$$\|T_K f\|_{L^\infty(Y)} \leqslant B\|f\|_{L^{r^*}(X)},$$

we obtain the required result. $\qquad \square$

Example. *Here is a version of Schur's Lemma that is more hands on and illustrative of the result. Let M be an $m \times n$ matrix, such that the sums of the absolute values of the entries of M on each row are upper bounded by A, and the sums of the absolute values of the entries of M on each column are upper bounded by B. Then for $x \in \mathbb{R}^n$, we have*

$$|Mx| \leqslant \sqrt{AB} \cdot |x|.$$

363

Example. *Young's Inequality can be justified by Schur's lemma, over any locally compact group G equipped with a Haar measure. Indeed, for a fixed g, the kernel of the operator $Cf = f * g$ is given by*

$$K(x, y) = g(x^{-1}y)$$

and then

$$\|K\|_{L_x^\infty(G)L_y^r(G)} = \|g\|_{L^r(G)}$$

and

$$\|K\|_{L_y^\infty(G)L_x^r(G)} = \|g\|_{L^r(G)}$$

so Schur's Lemma shows that for $1/p + 1/r = 1/q + 1$,

$$\|f * g\|_{L^q(G)} = \|Cf\|_{L^q(G)} \lesssim \|f\|_{L^p(G)}\|g\|_{L^r(G)}.$$

Example. *We cannot use Schur's Lemma to bound the fractional integration operators*

$$T_1 f(y) = \int \frac{f(x)}{|x - y|^{d-s}} \, dx$$

since the kernel $K_1(x, y) = 1/|x - y|^{d-s}$ does not lie in any of the spaces $L^p(X)$. On the other hand, we can use Schur's Lemma to bound the modified fractional integration operators

$$T_2 f(y) = \int \frac{f(x)}{\langle x - y \rangle^{d-s}} \, dx,$$

where $\langle t \rangle = (1 + |t|^2)^{1/2}$ is the Japanese bracket. The kernel

$$K_2(x, y) = 1/\langle x - y \rangle^{d-s}$$

lies in $L_x^\infty(\mathbb{R}^d)L_y^r(\mathbb{R}^d)$ and $L_y^\infty(\mathbb{R}^d)L_x^r(\mathbb{R}^d)$ provided that $r > d/(d - s)$. Thus Schur's Lemma implies that for $1/p - 1/q > s/d$,

$$\|T_2 f\|_{L^q(\mathbb{R}^d)} \lesssim_{d,s,p,q} \|f\|_{L^p(X)}.$$

We will obtain a weak-type version of Schur's test later on in this chapter, which will enable us to bound T_1.

For $1 < p < \infty$, we do not expect Schur's test to be sharp in general. But a good heuristic is that a variant of the inequality can be made to be sharp provided that two properties hold:

- For all $y \in Y$,
$$\int |K(x,y)| \, dx \approx A$$

and for all $x \in X$,
$$\int |K(x,y)| \, dy \approx B.$$

This kind of homogeneity will be present if $K(x,y) = k(x-y)$ is a convolution kernel.

- There is little oscillation in the kernel K.

Assuming for simplicity that X and Y have finite measure, from the first property we conclude that
$$A|Y| \approx B|X|.$$

Thus if we set $f = \mathbb{I}_X$, 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

$$\lambda = \int \lambda f(x)^p \, dx$$

$$= \int f(x) T_K^*((T_K f)^{p-1})(x) \, dx$$

$$= \int (T_K f)^p(x) = A^p.$$

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, carried out in particular for the next example, 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.

Example. *Let us use the Lagrange multiplier technique above to obtain a weighted estimate. Consider the operator*

$$Tf(y) = \int_0^\infty \frac{f(x)}{x + y} \, dx.$$

We will prove that T is bounded on L^2, a result which cannot be proven using Schur's Lemma. The kernel of this operator is

$$K(x, y) = \frac{1}{x + y}.$$

Since K is non-negative, and monotonically decreasing in y, maximizers for an L^p to L^q estimate must be monotonically decreasing, with a singularity at the origin. The worst case singularity we can have near the origin while remaining close to

366

being in L^2 is a singularity of the form $x^{-1/2}$. Thus we plug in $f(x) = x^{-1/2}$ into the operator T. Then we calculate that

$$Tf(y) = \int_0^\infty \frac{1}{x^{1/2}(x+y)}\, dx$$
$$\sim \int_0^y \frac{1}{x^{1/2}y} + \int_y^\infty \frac{1}{x^{3/2}}\, dx \sim y^{-1/2}.$$

Thus if we let $w(x) = x^{-1/2}$, and $v(y) = y^{-1/2}$, then we have already calculated that

$$\int K(x,y)[w(x)/v(y)]\, dx \lesssim 1.$$

Similarily,

$$\int K(x,y)[v(y)/w(x)]\, dy \lesssim 1.$$

Define

$$K_t(x,y) = K(x,y)w(x)^{2t-1}v(y)^{1-2t}.$$

Then $\{K_t\}$ is an analytic family of kernels giving us an analytic family of operators $\{T_t\}$, to which we can apply complex interpolation. For $t = 0$, we obtain that

$$\|T_0 f\|_{L^1(\mathbb{R}^+)} \lesssim \|f\|_{L^1(\mathbb{R}^+)}.$$

For $t = 1$, we obtain that

$$\|T_1 f\|_{L^\infty(\mathbb{R}^+)} \lesssim \|f\|_{L^\infty(\mathbb{R}^+)}.$$

For $t = 1/2$, we therefore obtain by interpolation that

$$\|T_{1/2} f\|_{L^2(\mathbb{R}^+)} \lesssim \|f\|_{L^\infty(\mathbb{R}^+)}.$$

But $T_{1/2} = T$, which means we have proved the required boundedness statement.

We can also get L^p bounds for all $1 < p < \infty$, though the operator is not bounded on $L^1(\mathbb{R}^d)$ or $L^\infty(\mathbb{R}^d)$, as can be seen by plugging in indicator functions of intervals. To do this, we obtain a weak type bound from $L^1(\mathbb{R}^+)$ to $L^{1,\infty}(\mathbb{R}^+)$ by proving that for $f \in L^1(\mathbb{R}^+)$,

$$Tf(y) \lesssim \frac{\|f\|_{L^1(\mathbb{R}^+)}}{y},$$

367

from which it follows that $\|Tf\|_{L^{1,\infty}(\mathbb{R}^+)} \lesssim \|f\|_{L^1(\mathbb{R}^+)}$. *To prove this result, we note that for* $f \geqslant 0$,

$$Tf(y) = \int \frac{f(x)}{x+y}\,dx \lesssim \int_0^y \frac{f(x)}{y}\,dx + \int_y^\infty \frac{f(x)}{x}\,dx \lesssim \frac{\|f\|_{L^1(\mathbb{R}^+)}}{y} + \frac{\|f\|_{L^1(\mathbb{R}^+)}}{y}.$$

Real interpolation gives bounds from $L^p(\mathbb{R}^+)$ *to* $L^p(\mathbb{R}^+)$ *for* $1 < p \leqslant 2$. *Applying the fact that* T *is self-adjoint, duality implies boundedness for* $2 \leqslant p < \infty$.

As another example of the trivial nature of the boundedness of kernel operators from $L^1(X)$, let us consider a version of Schur's Lemma that characterizes the boundedness from the $L^1(X)$ norm to a suitable family of Lorentz spaces, and from those Lorentz spaces to $L^\infty(Y)$.

Theorem 20.4. *Fix* $1 < q \leqslant \infty$ *and* $1 \leqslant s \leqslant \infty$. *Then*

$$\|S_K f\|_{L^{q,s}(Y)} \lesssim_{q,s} \|K\|_{L^\infty(Y)L^{q,s}(X)} \|f\|_{L^1(X)},$$

which means $T_K f$ *is well-defined as a Lebesgue integral for* $f \in L^1(X)$, *and satisfies equivalent bounds. If* $0 < p \leqslant \infty$, *and* $1 \leqslant s \leqslant \infty$, *then*

$$\|S_K f\|_{L^\infty(Y)} \lesssim_{p,s} \|K\|_{L^\infty(X)L^{p^*,s^*}(Y)} \|f\|_{L^{p,s}(X)}.$$

Thus $T_K f$ *is well defined as a Lebesgue integral for* $f \in L^{p,s}(X)$, *and satisfies equivalent bounds. In particular, if* $1 < r < \infty$, *and*

$$\|K\|_{L^\infty(Y)L^{r,\infty}(X)} \leqslant A,$$

and

$$\|K\|_{L^\infty(X)L^{r,\infty}(Y)} \leqslant B,$$

for some $1 < r < \infty$, *then for* $1 < p < q < \infty$ *with* $1/p + 1/r = 1/q + 1$, *and any* $0 < s \leqslant \infty$,

$$\|S_K f\|_{L^{q,s}(X)} \lesssim_{p,q,r,s} A^{r/q} B^{1-r/q} \|f\|_{L^{p,s}(Y)}.$$

Proof. The space $L^{q,s}(Y)$ is a norm space for $q > 1$, so we can employ Minkowski's inequality to conclude that

$$\|S_K f\|_{L^{q,s}(Y)} = \|S_K f\|_{L^{q,s}(Y)L^1(X)} \lesssim \|S_K f\|_{L^1(X)L^{q,s}(Y)}.$$

And then we conclude

$$\|S_K f\|_{L^1(X)L^{q,s}(Y)} \lesssim \|K\|_{L^\infty(X)L^{q,s}(Y)} \|f\|_{L^1(X)}$$

which proves the first claim. To obtain the second claim, we apply Hölder's inequality in the $L^{p,q}$ norms, which shows that

$$\|S_K f\|_{L^\infty(Y)} = \|Kf\|_{L^\infty(Y)L^1(X)} \lesssim_{p,s} \|K\|_{L^\infty(Y)L^{p^*,s^*}(X)} \|f\|_{L^{p,s}(X)}. \qquad \square$$

Remark. Establishing bounds from $L^1(X)$ to $L^{1,\infty}(Y)$ can be much more subtle than indicated by the other results here, e.g. for the Hardy-Littlewood Maximal function or for the Hilbert transform.

Example. *For $s > 0$, and $f \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$, we can define*

$$Tf(x) = \int \frac{f(y)}{|y - x|^{d-s}} \, dy,$$

because $z \mapsto 1/|z|^{d-s}$ is integrable for $|z| \leqslant 1$, and bounded for $|z| \geqslant 1$. Applying the result above, since $1/|z|^{d-s}$ lies in $L^r(\mathbb{R}^d)$ for $r = d/(d-s)$, we conclude that for $1 < p < q < \infty$ with $s = d(1/p - 1/q)$,

$$\|Tf\|_{L^q(\mathbb{R}^d)} \lesssim_{d,s,p,q} \|f\|_{L^p(\mathbb{R}^d)}.$$

This is the Hardy-Littlewood-Sobolev *inequality.*

   TODO: Discuss weighted variants (e.g. higher dimensional version of Q9 of Tao, Notes 2).

20.2   Hilbert-Schmidt Techniques

If T has a kernel $K : X \times Y \to \mathbb{C}$ which is *square integrable*, then we find via Hölder's inequality that

$$
\begin{aligned}
\|Tf\|_{L^2(Y)}^2 &= \int_Y \left| \int_X K(x, y) f(x) \, dx \right|^2 dy \\
&\leqslant \int_Y \|K\|_{L_x^2}^2 \|f\|_{L_x^2}^2 \, dy \\
&= \|K\|_{L_x^2 L_y^2}^2 \|f\|_{L_x^2}^2.
\end{aligned}
$$

Thus $\|T\|_{L^2 \to L^2} \leqslant \|K\|_{L^2(X \times Y)}$. The quantity $\|K\|_{L^2(X \times Y)}$ is the *Hilbert-Schmidt*, or *Frobenius* norm of the operator T. If this inequality is tight at all steps for some

input f, then tightness for Hölder's inequality implies that there exists $a(y)$ for almost every y such that $K(x, y) = a(y)\overline{f(x)}$, so that

$$Tg(y) = a(y)\langle g, f \rangle$$

is a rank one operator. Heuristically, this means that employing the Hilbert-Schmidt / Frobenius norm will only yield efficient bounds when T is a low rank operator. In particular, T must be qualitatively low-rank, because any Hilbert-Schmidt integral operator is *compact*.

20.3   Change of Variables

TODO

On the other hand, it is not necessarily possible to change all variables in $\mathbb{R}^n \times \mathbb{R}^m$ without affecting the mapping properties of the operator. For instance, if T is the circular means operator from $\mathbb{R}^d$ to $\mathbb{R}^d$, then T has kernel $K(x, y) = \delta(1 - |x - y|)$. If we change variables simultaneously, for instance, letting $x = z + w$ and $y = w$, then we obtain the kernel $K'(z, w) = K(z + w, w) = \delta(1 - |z|)$ is independent of w, which has completely different mapping properties, in particular, the output of the kernel does not actually depend on the input.

20.4   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. By doing this, we can isolate different features of a function to be controlled.

Many of the techniques of harmonic analysis are only possible once localized. In other words, rather than proving estimates of the form

$$\|Tf\|_{L^q(Y)} \lesssim \|f\|_{L^p(X)},$$

we only prove estimates of the form

$$\|T_{\phi_1, \phi_2} f\|_{L^q(Y)} \lesssim \|f\|_{L^p(X)}$$

where ϕ_1 and ϕ_2 are supported on finite measure subsets of X and Y, and $T_{\phi_1, \phi_2} f = \phi_1 T(\phi_2 f)$. Because of the localized nature of the problem, the difficulty of the

bound increases as q becomes larger, and as p becomes smaller. If T has a kernel $K \in L^\infty(\operatorname{supp}(\phi_1) \times \operatorname{supp}(\phi_2))$, then T_{ϕ_1,ϕ_2} maps $L^1(\mathbb{R}^d)$ to $L^\infty(\mathbb{R}^d)$, which makes the study of the qualitative behaviour of T_{ϕ_1,ϕ_2} with respect to the convex norms trivial. Working quantitatively, we should therefore expect studying localized estimates to depend on quantitative measures of the singular nature of K, e.g. most crudely, the L^∞ norm. Because of the sensitivity to the singular nature of K, choosing ϕ_1 and ϕ_2 to be non-singular often makes the problem more amenable to analysis.

Intuitively, the smaller we make the kernel, the more well behaved the resulting operator will be. If K is a non-negative kernel, then it is certainly true that the L^p to L^q boundedness of an operator T with kernel K implies the boundedness of an operator T_Ω with kernel $K \mathbb{I}_\Omega$ for any $\Omega \subset X \times Y$, since we will then have a pointwise bound $|T_\Omega(f)| \leqslant T(|f|)$. But if K is not necessarily non-negative, then this is not necessarily true anymore, since T may only be bounded because of more subtle cancellation properties.

Example. *Set $K(x,y) = e^{-2\pi i x \cdot y}$, and let $\Omega \subset \mathbb{R}^d \times \mathbb{R}^d$ be the set of all pairs (x,y) such that $\operatorname{Re}(e^{-2\pi i x \cdot y}) \geqslant 0$. Then the operator T with kernel K is unitary, and thus bounded from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$. The operator T_Ω with kernel $K \mathbb{I}_\Omega$, on the other hand, is not bounded from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$, since, if $f = \mathbb{I}(|x| \leqslant R)$, then we will have $|T_\Omega f(y)| \gtrsim_d R^d$ for all $y \in \mathbb{R}^d$, so that $T_\Omega f \notin L^2(\mathbb{R}^d)$.*

There are several cases where we can perform an arbitrary truncation. If $p = 1$, or $q = \infty$, then the operator norm of T with kernel K depends solely on mixed L^p norms of the kernel K, and thus behave well under truncation. Another case is if $\Omega = E \times F$, for then

$$\|T_\Omega f\|_{L^q(Y)} = \|\mathbb{I}_E\, T(\mathbb{I}_F\, f)\|_{L^q(Y)} \leqslant \|T(\mathbb{I}_F\, f)\|_{L^q(Y)} \leqslant \|\mathbb{I}_F\, f\|_{L^p(X)} \leqslant \|f\|_{L^p(X)}.$$

Thus $\|T_\Omega\|_{L^p \to L^q} \leqslant \|T\|_{L^p \to L^q}$. Similarily, if $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$, $\Omega_1 \subset \mathbb{R}^n$ and $\Omega_2 \subset \mathbb{R}^m$ are precompact open sets, $L_1 : \mathbb{R}^n \to \mathbb{R}^n$ and $L_2 : \mathbb{R}^m \to \mathbb{R}^m$ are invertible linear maps, and $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{C}$ is a bump function adapted to $L_1(\Omega_1) \times L_2(\Omega_2)$, then

$$\|T_{K\phi}\|_{L^p \to L^q} \lesssim_{\Omega_1,\Omega_2} \|T_K\|_{L^p \to L^q}.$$

The trick here is to replace ϕ with a tensor product using a density argument / Fourier series which reduces our study to block diagonal type truncations. Similar results can be obtained if X and Y are smooth manifolds.

More generally, if we have a family of disjoint sets $\{E_n\}$ and $\{F_n\}$, then we can truncate to the *block diagonal* region $\Omega = \bigcup (E_n \times F_n) = \bigcup \Omega_n$. The norm of such an operator can be calculated exactly from the behaviour on each subregion, as the next lemma shows.

Lemma 20.5. *Consider a family of Banach spaces $\{X_n\}$ and $\{Y_n\}$, and a family of bounded operators $T_n : X_n \to Y_n$. Then for $p \leqslant q$,*

$$\sup_{x_n \in X_n} \frac{\|T_n(x_n)\|_{l_n^q}}{\|x_n\|_{l_n^p}} = \sup_n \|T_n\|.$$

and if $p > q$, then

$$\sup_{x_n \in X_n} \frac{\|T_n(x_n)\|_{l_n^q}}{\|x_n\|_{l_n^p}} = \|T_n\|_{l_n^r},$$

where $1/p + 1/r = 1/q$.

Proof. We calculate that for $p \leqslant q$,

$$\|T_n(x_n)\|_{l_n^q} \leqslant \|T_n(x_n)\|_{l_n^p} \leqslant \|T_n\|_{l_n^\infty} \|x_n\|_{l_n^p}.$$

The converse bound for the supremum is simple. For $p > q$, Hölder's inequality implies that

$$\|T_n(x_n)\|_{l_n^q} \leqslant \|\|T_n\| \|x_n\|\|_{l_n^q} \leqslant \|T_n\|_{l_n^r} \|x_n\|_{l_n^p}.$$

Conversely, if we pick x_n with $\|x_n\| = 1$ for each n such that $\|T_n(x_n)\| \geqslant (1 - \varepsilon)\|T_n\|$, and if we pick $a_n = \|T_n\|^{r/q-1} = \|T_n\|^{r/p}$, then

$$\|T_n(a_n x_n)\|_{l_n^q} \geqslant (1 - \varepsilon)\|\|T_n\|^{r/q}\|_{l_n^q} \geqslant (1 - \varepsilon)\|T_n\|_{l_n^r}^{r/q},$$

whereas

$$\|a_n x_n\|_{l_n^p} = \|\|T_n\|^{r/p}\|_{l_n^p} = \|T_n\|_{l_n^r}^{r/p}.$$

Thus

$$\frac{\|T_n(a_n x_n)\|_{l_n^q}}{\|a_n x_n\|_{l_n^p}} \geqslant (1 - \varepsilon)\|T_n\|_{l_n^r}^{r/q - r/p} = (1 - \varepsilon)\|T_n\|_{l_n^r}.$$

Taking $\varepsilon \to 0$ gives the lower bound. $\qquad\square$

A similar result holds for restriction to *almost* block diagonal operators, i.e. operators of the form

$$T = \sum_{(n,m) \in G} T_{nm}$$

where $T_{nm} : X_n \to Y_m$, and G is a graph on the index set such that each node has degree at most $O(1)$. One can then break the graph into $O(1)$ many block diagonal operators $T_1, \ldots, T_K$, from which we obtain that e.g. for $p \leqslant q$,

$$\|T_{nm}x_n\|_{l_n^q} \lesssim \sup \|T_{nm}\|_{l_{nm}^\infty} \|x_n\|_{l_n^p}.$$

and for $p < q$,

$$\|T_{nm}x_n\|_{l_n^q} \lesssim \|T_{nm}\|_{l_{nm}^r} \|x_n\|_{l_n^p}.$$

On the other hand, we certainly have

$$\frac{\sup_x \|T_{nm}x_n\|_{l_n^q}}{\|x_n\|_{l_n^p}} \geqslant \|T_{nm}\|_{l_{nm}^\infty}.$$

TODO: Do we have a l^r lower bound as well, or not?

We can also truncate to 'upper triangular' diagonals rather than just block diagonals, provided that $q > p$. For $q = p$, we 'lose a logarithm'.

Theorem 20.6 (Christ-Kiselev). *Consider an operator T mapping functions on X to functions on Y, consider subsets $\{E_n\}$ of X and subsets $\{F_n\}$ of Y, and let*

$$\Omega = \bigcup_{n \leqslant m} E_n \times F_m.$$

Then if $1 \leqslant p < q \leqslant \infty$,

$$\|T_\Omega\|_{L^p \to L^q} \lesssim_{p,q} \|T\|_{L^p \to L^q}.$$

If $1 \leqslant p \leqslant \infty$, and n ranges over $\{1, \ldots, N\}$, then

$$\|T_\Omega\|_{L^p \to L^p} \lesssim \log(N) \cdot \|T\|_{L^p \to L^p}.$$

Proof. Let us first address the case where $q > p$. We assume the index set is finite, i.e. $1 \leqslant n \leqslant N$ for some large $N > 0$, and prove the result by induction, though our constant will be independent of N and thus give us results for infinite index sets by taking limits. Fix a large constant $A_{p,q}$. Then the base case $N = 1$ is automatically satisfied if $A_{p,q}$ is large enough. Without loss of generality, assume $\|T\|_{L^p \to L^q} = 1$, and consider $f \in L^p(X)$ with $\|f\|_{L^p(X)} = 1$. Our proof is complete if we can show $\|T_\Omega f\|_{L^q(Y)} \leqslant A_{p,q}$. The idea here is to divide and conquer in an intelligent way. We can find an index $0 \leqslant n_0 \leqslant N$ such that

$$\|f \mathbb{I}_{E_{\leqslant n_0}}\|_{L^p(X)}^p \leqslant 1/2 < \|f \mathbb{I}_{E_{\leqslant n_0+1}}\|_{L^p(X)}^p.$$

373

Write $f_0 = f \mathbb{I}_{E_{\leqslant n_0}}$, $f_1 = f \mathbb{I}_{E_{n_0+1}}$, and $f_2 = f \mathbb{I}_{E_{\geqslant n_0+2}}$. Then $\|f_0\|_{L^p(X)}, \|f_2\|_{L^p(X)} \leqslant 1/2^{1/p}$. Applying induction, we have

$$\| \mathbb{I}_{F_{\leqslant n_0}} T_\Omega f_0 \|_{L^q(Y)} \leqslant A_{p,q} 2^{-1/p}$$

and

$$\| T_\Omega f_2 \|_{L^q(Y)} \leqslant A_{p,q} 2^{-1/p}.$$

On the other hand, the other restrictions are block diagonals, which gives

$$\| \mathbb{I}_{F_{\geqslant n_0+1}} T_\Omega f_0 \|_{L^q(Y)} \leqslant 2^{-1/p},$$

and

$$\| T_\Omega f_1 \|_{L^q(Y)} \leqslant \|f_1\|_{L^p(X)} \leqslant \|f\|_{L^p(X)} \leqslant 1.$$

The triangle inequality implies for $q > p$ that if $A_{p,q}$ is chosen large enough, depending on p and q,

$$\| \mathbb{I}_{F_{\leqslant n_0}} T_\Omega f \|_{L^q(Y)} \leqslant (1 + A_{p,q}) 2^{-1/p} \leqslant 2^{(1/2)(1/p-1/q)} A_{p,q} 2^{-1/p},$$

$$\| \mathbb{I}_{F_{n_0+1}} T_\Omega f \|_{L^q(Y)} \leqslant 1 + 2^{1-1/p} \leqslant 3,$$

and

$$\| \mathbb{I}_{F_{\geqslant n_0+2}} T_\Omega f \|_{L^q(Y)} \leqslant 1 + (1 + A_{p,q}) 2^{-1/p} \leqslant 2^{(1/2)(1/p-1/q)} A_{p,q} 2^{-1/p}.$$

Putting these three estimates together, using the disjoint support, gives, again if $A_{p,q}$ is suitably large, depending on p and q, that

$$\| T_\Omega f \|_{L^q(Y)} \leqslant (2(2^{(1/2)(1/p-1/q)})^q A_{p,q}^q 2^{-q/p} + 3^q)^{1/q}$$
$$\leqslant (2^{-(1/2)(q/p-1)} A_{p,q}^q + 3^q)^{1/q} \leqslant A_{p,q}.$$

This completes the proof in the case $q > p$.

Now consider the case $q = p$, and continue an induction on N, assuming without loss of generality that $\|T\|_{L^p \to L^q} = 1$. Write $A_p(N)$ for the optimal constant such that

$$\| T_\Omega \|_{L^p \to L^p} \leqslant A_p(N).$$

over all upper diagonal block operators of size $\leqslant N$. Consider f with $\|f\|_{L^p(X)} = 1$, and break $f = f_0 + f_1$, where $f_0 = f \mathbb{I}_{E_{\leqslant n_0}}$ and $f_1 = f \mathbb{I}_{E_{> n_0}}$, where $n_0, N - n_0 \leqslant N/2$. Let $\|f_0\|_{L^p(X)} = r$ and $\|f_1\|_{L^p(X)} = s$, so $r^p + s^p = 1$. Applying induction, we have

$$\| \mathbb{I}_{F_{\leqslant n_0}} T_\Omega f_0 \|_{L^p(Y)} \leqslant A_p(N/2) r$$

374

and
$$\|T_\Omega f_1\|_{L^p(Y)} \leqslant A_p(N/2)s.$$

We also have
$$\|\mathbb{I}_{F_{\geqslant n_0}} T_\Omega f_0\|_{L^p(Y)} \leqslant 1.$$

Putting this together yields

$$
\begin{aligned}
\|T_\Omega f\|_{L^p(Y)} &\leqslant (A_p(N/2)^p r^p + (1 + A_p(N/2)s)^p)^{1/p} \\
&\leqslant (A_p(N/2)^p(r^p + s^p) + O_p(A_p(N/2)^{p-1} s^{p-1}))^{1/p} \\
&= (A_p(N/2)^p + O_p(A_p(N/2)^{p-1} s^{p-1}))^{1/p} \\
&\leqslant (A_p(N/2)^p + O_p(A_p(N/2)^{p-1}))^{1/p} \\
&\leqslant A_p(N/2) + O_p(1).
\end{aligned}
$$

Thus we find $A_p(N) \leqslant A_p(N/2) + O_p(1)$, and this implies that $A_p(N) \lesssim_p \log(N)$, completing the proof. $\qquad\square$

The decomposition is arbitrary, which implies (via linearization) a variant by maximal functions.

Theorem 20.7 (Christ-Kiselev). *Consider $E_1 \subset E_2 \subset \cdots \subset X$, let $T : L^p(X) \to L^q(Y)$, and consider the maximal operator*

$$Mf = \sup_n |T(\mathbb{I}_{E_n} f)|.$$

For $1 \leqslant p < q \leqslant \infty$,
$$\|M\|_{L^p \to L^q} \lesssim_{p,q} \|T\|_{L^p \to L^q}$$
and for $1 \leqslant p \leqslant \infty$, and $E_1 \subset E_2 \subset \cdots \subset E_N \subset X$ are given,

$$\|M\|_{L^p \to L^p} \lesssim \log(N)\|T\|_{L^p \to L^p}.$$

Proof. By monotone convergence, we may assume without loss of generality that the index set is finite. For $f \in L^p(X)$, and $y \in Y$, let $n(y)$ be an index such that $Mf(y) = T(\mathbb{I}_{E_{n(y)}} f)(y)$. Let $X_n = E_n - E_{n-1}$, and let $Y_n = \{y \in Y : n(y) = n\}$. Then for each $y \in Y$, if $\Omega = \bigcup_{n \leqslant m} X_n \times Y_m$,

$$Mf(y) = T(\mathbb{I}_{E_{n(y)}} f)(y) = T_\Omega f(y).$$

We can then apply the Christ-Kiselev Lemma above, since Ω is upper triangular.
$$\square$$

375

We can apply this result to study the *maximal Fourier transform*

$$\mathcal{F}^* f(\xi) = \sup_I \int_I f(x) e^{-2\pi i \xi \cdot x} \, dx = \sup_I \widehat{f \, \mathbb{I}_I},$$

where I ranges over all compact intervals.

Theorem 20.8 (Menshov-Paley-Zygmund). *For $1 \leqslant p < 2$,*

$$\left\| \mathcal{F}^* f \right\|_{L^{p^*}(\mathbb{R})} \lesssim_p \|f\|_{L^p(\mathbb{R})}.$$

Proof. Applying the triangle inequality, we may assume one endpoint of each interval in the supremum lies at the origin. By symmetry, we may assume that this is the left hand endpoint. A convergence argument allows us to assume that the other endpoint is a positive rational number, and we may restrict ourselves by monotone convergence to a finite subset $\{q_1 < q_2 < \cdots < q_N\}$. But then the result follows by the Christ-Kiselev Maximal Function Lemma above, and the Hausdorff-Young inequality. $\qquad\square$

A qualitative consequence is that if $f \in L^p(\mathbb{R})$ for some $1 \leqslant p < 2$, then

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

for almost every $\xi \in \mathbb{R}$.

Remark. The result is also true for $p = 2$, but is a *much harder theorem*, due to Carleson.

The Christ-Kiselev Lemma and its 'vector valued variants' becomes very useful in the study of nonlinear dispersive partal differential equations.

376

Chapter 21

Maximal Averages

This chapter is an introduction to the behaviour of basic averaging operators. A classical example, given a function $f \in L^1_{\text{loc}}(\mathbb{R})$, are the averaging operators

$$A_\delta f(x) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} f(y) \, dy.$$

If $f \in C(\mathbb{R})$, then for each $x \in \mathbb{R}$, $\lim_{\delta \to 0} A_\delta f(x) = f(x)$. 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 $x \in \mathbb{R}$,

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{1}{h} \int_x^{x+h} f(y) \, dy = f(x).$$

Our main goal will be study whether pointwise convergence of the averages $A_\delta f$ hold in higher dimensions, for a more general family of functions that are not necessarily continuous, and for a more general family of averaging operators, thus testing the extent to which the 'fundamental theorem of calculus' holds.

The classical family of averaging operators on $\mathbb{R}^d$ are defined for $\delta > 0$, $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, and $x \in \mathbb{R}^d$ by setting

$$A_\delta f(x) = \fint_{B(x,\delta)} f(y) \, dy,$$

where $B(x, \delta)$ is the ball of radius δ centred at x. A simple application of Schur's lemma shows that $\|A_\delta f\|_{L^p(\mathbb{R}^d)} \leqslant \|f\|_{L^p(\mathbb{R}^d)}$ 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 $f \in L^p(\mathbb{R}^d)$, for $1 \leqslant p < \infty$, $A_\delta f$ converges to f in L^p norm. This implies that for any $f \in L^p(\mathbb{R}^d)$, 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*. To do this, we introduce the fundamental tool of *maximal functions*.

21.1   The Method of Maximal Functions

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 a sublinear operator M by setting

$$M f = \sup_{\delta > 0} A_\delta |f|.$$

The next theorem indicates why obtaining bounds on the operator M gives pointwise convergence results.

Theorem 21.1. *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

$$\{f \in V : \lim_{t \to \infty} T_t f(y) = S f(y) \text{ for a.e } y\}$$

is closed in V.

Proof. Fix $\{u_n\}$ in V converging to $u \in V$, and suppose for each n,

$$\lim_{t \to \infty} (T_t u_n)(x) = S u_n(x)$$

holds for almost every $x \in X$. For each $\lambda > 0$, we find

$$|\{x \in X : \limsup_{t \to \infty} |T_t u(x) - S u(x)| > \lambda\}|$$

$$\leqslant |\{x \in X : \limsup_{t} |T_t(u - u_n)(x) - S(u - u_n)(x)| > \lambda\}|$$

$$\leqslant |\{x \in X : |T_*(u - u_n)(x)| > \lambda/2\}| + |\{x : |S(u - u_n)(x)| > \lambda/2\}|$$

$$\lesssim_{p,q} \frac{\|u - u_n\|_V^q}{\lambda^q} + \frac{\|u - u_n\|_V^p}{\lambda^p}.$$

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 $x \in X$. □

Since we know $A_\delta f$ converges to f pointwise for any $f \in C(\mathbb{R})^d$, we see from the result above that taking $T_t = A_\delta$ and S the identity map gives almost everywhere convergence if we can obtain bounds for the maximal operator M of the form

$$\left\| \sup_{\delta > 0} A_\delta f \right\|_{L^{q,\infty}(\mathbb{R}^d)} \lesssim \|f\|_V$$

for an appropriate norm $\| \cdot \|_V$ and $0 < q < \infty$ in which $C(\mathbb{R}^d)$ is dense. We have already obtained a bound

$$\sup_{\delta > 0} \|A_\delta f\|_{L^{q,\infty}(\mathbb{R}^d)} \leqslant \sup_{\delta > 0} \|A_\delta f\|_{L^q(\mathbb{R}^d)} \leqslant \|f\|_{L^q(\mathbb{R}^d)}$$

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 not only controls the height and width of functions, but also shows that the main contributions to the height and widths of the functions $A_\delta f$ are supported on common regions of space.

21.2   Covering Methods

The bound $\|Mf\|_{L^\infty(\mathbb{R}^d)} \leqslant \|f\|_{L^\infty(\mathbb{R}^d)}$ follows from a direct calculation. This it is trivial to control the height of the function Mf in terms of the heigt of the function

379

f. The difficult part is obtaining control of the width of Mf in terms of the width of f. Control on width can only be obtained up to a certain degree, because for any locally integrable $f \neq 0$, $\mathrm{supp}(Mf) = \mathbb{R}^d$, so the width of f 'explodes'. A slightly more technical calculation shows that we cannot even have a bound of the form

$$\|Mf\|_{L^1(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}.$$

In fact, Mf is not integrable for any nonzero $f \in L^1(\mathbb{R}^d)$.

Lemma 21.2. *If $f \in L^1(\mathbb{R}^d)$ is nonzero, then Mf is not integrable. Moreover, there exists $f \in L^1(\mathbb{R}^d)$ such that $Mf \notin L^1_{loc}(\mathbb{R}^d)$.*

Proof. Fix a nonzero $f \in L^1(\mathbb{R}^d)$. By rescaling, we may assume without loss of generality that $\|f\|_{L^1(\mathbb{R}^d)} = 2$. 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

$$\int_{\mathbb{R}^d} |Mf(x)| \gtrsim \int_{\mathbb{R}^d} \frac{1}{|x|^d} = \infty.$$

Thus $Mf \notin L^1(\mathbb{R}^d)$.

To find f such that $Mf \notin L^1_{\mathrm{loc}}(\mathbb{R}^d)$ take

$$f(x) = \frac{1}{|x|(\log |x|)^2}.$$

Then 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 near the origin. $\qquad\square$

The last lemma shows that if $f \in L^1(\mathbb{R}^d)$, then we have a pointwise bound $|Mf(x)| \gtrsim_f \langle x \rangle^{-d}$. Note, however, that $|x|^{-d}$ is only *barely* nonintegrable. We will also show that Mf is barely nonintegrable by obtaining a bound

$$\|Mf\|_{L^{1,\infty}(\mathbb{R}^d)} \lesssim_d \|f\|_{L^1(\mathbb{R}^d)}.$$

Interpolation thus shows that $\|Mf\|_{L^p(\mathbb{R}^d)} \lesssim_{d,p} \|f\|_{L^p(\mathbb{R}^d)}$ for all $1 < p \leqslant \infty$.

The standard real-variable technique of obtaining the $L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R}^d)$ bound of the maximal function is geometric, applying a covering argument. To obtain the weak-type bound, we must show that the set

$$E_\lambda = \{x \in \mathbb{R}^d : |Mf(x)| > \lambda\}$$

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

$$\|f\|_{L^1(\mathbb{R}^d)} \geqslant \sum_{i=1}^N \int_{B_i} |f(y)| \, dy > \lambda \sum_{i=1}^N |B_i| \gtrsim_d \lambda |E_\lambda|$$

which would show $|E_\lambda| \lesssim_d \|f\|_{L^1(\mathbb{R}^d)}/\lambda$, i.e. that $\|Mf\|_{L^{1,\infty}(\mathbb{R}^d)} \lesssim_d \|f\|_{L^1(\mathbb{R}^d)}$. 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 in which an averaging operator is involved. All that is really required for the basic theory is a basic 'covering type argument' that holds in a great many situations. To prove the Vitali Covering Lemma in this greater level of generality, we fix a locally compact space X equipped 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. We then assume the following property holds:

- There is $c > 0$ such that for any $x \in X$ and $\delta > 0$, if

$$B^*(x, \delta) = \bigcup \{B(x', \delta) : B(x, \delta) \cap B(x', \delta) \neq \varnothing\},$$

then $|B^*(x, \delta)| \leqslant c|B(x, \delta)|$.

If X is a metric space, and $B(x, \delta)$ is the ball of radius δ, then we have $B^*(x, \delta) \subset B(x, 3\delta)$. If the radon measure satisfies a *doubling condition*

$$|B(x, 3\delta)| \leqslant c|B(x, \delta)|,$$

then the assumption will hold for the same c. For instance, the Lebesgue measure $X = \mathbb{R}^d$ satisfies the doubling condition

$$|B(x, 3\delta)| \leqslant 3^d |B(x, \delta)|.$$

So in this case we can set $c = 3^d$.

Lemma 21.3 (Vitali Covering Lemma). *If $B_1, \ldots, B_N$ is a finite collection of balls in X, then there is a disjoint subcollection $B_{i_1}, \ldots, B_{i_M}$ such that*

$$\bigcup_i B_i \subset \bigcup_j B_{i_j}^*.$$

Thus given the property above, we have

$$\left| \bigcup_{i=1}^N B_i \right| \leqslant c \sum_{j=1}^M |B_{i_j}|.$$

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 $B_{i_1}, \ldots, B_{i_k}$, 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

$$\left| \bigcup_{i=1}^N B_i \right| \leqslant \sum_{j=1}^M |B_{i_j}^*| \leqslant c \sum_{j=1}^M |B_{i_j}|. \qquad \square$$

Let us now prove the bonuds for the maximal average using the Vitali covering lemma. The following technical assumptions are not always required, but make certain parts of the argument cleaner, and hold in most applications:

382

- 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 most particular instance, but in some situations they can be worked around to obtain similar results that we will obtain here. 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 21.4. *If $f \in L_1^{loc}(X)$, then $A_\delta f$ is a measurable function.*

Proof. If $f = a_1 \mathbb{I}_{U_1} + \cdots + a_N \mathbb{I}_{U_N}$ is a simple function, where $U_1, \ldots, U_N$ are open sets, then

$$A_\delta f(x) = a_1 \frac{|B(x, \delta) \cap U_1|}{|B(x, \delta)|} + \cdots + a_N \frac{|B(x, \delta) \cap U_N|}{|B(x, \delta)|}$$

is a continuous function by our technical assumptions. Next, if $f \geq 0$ 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 $x \in X$, and any open neighborhood U of x, there exists δ_0 such that for $\delta \leq \delta_0$, $\overline{B(x, \delta)} \subset U$. It follows that for any $f \in C(X)$ and $x \in X$,

$$\lim_{\delta \to 0} A_\delta f(x) = f(x).$$

383

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, we have seen that this implies 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, this pointwise convergence also holds for any $f \in L^1_{\text{loc}}(X)$. In particular, it also holds for any $f \in L^p(X)$ for $1 \leqslant p \leqslant \infty$.

Theorem 21.5. *For any $f \in L^1(X)$,*

$$\|Mf\|_{L^{1,\infty}(X)} \leqslant c \cdot \|f\|_{L^1(X)}.$$

Proof. Set

$$E_\lambda = \{x \in X : Mf(x) > \lambda\}.$$

Since we are working with a Radon measure, and E_λ is open, we have

$$|E_\lambda| = \sup_{K \subset E_\lambda} |K|,$$

where K is compact, and thus has finite measure. Fix any such compact subset K. Then K is covered by finitely many balls $B_1, \ldots, B_N$ such that on each ball B_i,

$$\int_{B_i} |f(y)| \, dy > \lambda |B_i|.$$

Using the Vitali lemma, extract a disjoint subfamily $B_{i_1}, \ldots, B_{i_M}$ 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

$$|K| \leqslant \frac{c \|f\|_{L^1(X)}}{\lambda}.$$

384

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. $\qquad\qquad\qquad\qquad \square$

Remark. A similar covering argument can be used to show that 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, also satisfies a bound

$$\|M'f\|_{L^{1,\infty}(X)} \leqslant c\|f\|_{L^1(X)}.$$

Given the weak assumption above, the uncentered and centered maximal functions can actually behave quite differently. We have the pointwise inequality $Mf \leqslant M'f$, but in general $M'f$ can be significantly bigger than Mf. However, we have $M'f \lesssim Mf$ under the following slightly stronger assumption on the family of balls we are working with: that there exists two constants $c_1, c_2 > 0$ such that the following two properties hold:

- (The Engulfing Condition) If $B(x, \delta) \cap B(x', \delta) \neq \varnothing$, then $B(x', \delta) \subset B(x, c_1\delta)$.

- (The Doubling Condition) $|B(x, c_1\delta)| \leqslant c_2|B(x, \delta)|$.

If $B = B(x', \delta)$ contains x, then B is contained in $B(x, c_1\delta)$, and conversely, $B(x, \delta)$ is contained in $B(x', c_1\delta)$. Thus

$$|B| = |B(x', \delta)| \geqslant c_2^{-1}B(x', c_1\delta) \geqslant c_2^{-1}|B(x, \delta)| \geqslant c_2^{-2}|B(x, c_1\delta)|.$$

Thus

$$\frac{1}{|B|}\int_B |f(y)|\, dy \leqslant \frac{c_2^2}{|B(x, c_1\delta)|}\int_{B(x,c_1\delta)} |f(y)|\, dy \leqslant c_2^2 Mf(x).$$

Taking suprema gives $M'f(x) \leqslant c_2^2 Mf(x)$. Thus under these stronger conditions, we have a pointwie bound $Mf \approx M'f$ with implicit constants independent of f. The engulfing and doubling conditions arise in more sophisticated techniques in this theory, e.g. when we analyze Calderon-Zygmund decompositions that come up in the real-variable analysis of singular integrals.

385

With the result proved, we obtain the Lebesgue differentiation theorem: If $f \in L^{1,\mathrm{loc}}(\mathbb{R}^d)$, then for almost every $x_0 \in \mathbb{R}^d$,

$$f(x_0) = \lim_{\delta \to 0} \int_{|x - x_0| \leqslant \delta} f(x)\, dx.$$

For $d = 1$, we can directly connect this to the fundamental theorem of calculus; if $f \in L^{1,\mathrm{loc}}(\mathbb{R}^d)$, then the function

$$g(x) = \int_0^x f(y)\, dy$$

is locally Lipschitz, and conversely we can find such f for any g by the Radon-Nikodym theorem. The Lebesgue differentiation theorem then says precisely that if g is locally Lipschitz, then g is differentiable almost everywhere, and that

$$g(x) = g(0) + \int_0^x g'(x)\, dx.$$

This result is called the *Rademacher differentiation theorem*. We develop similar ideas, and in higher dimensions, in the next chapter.

Under various other assumptions, one can obtain different covering lemmas that improve upon constants in the Vitali covering lemma. For instance, one 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 $I_1, \ldots, I_N$ such that any point in $\mathbb{R}$ is contained in at most two of the intervals. A modification of the argument above shows this gives the slightly better bound

$$\|Mf\|_{L^{1,\infty}(\mathbb{R})} \leqslant 2\|f\|_{L^1(\mathbb{R})},$$

rather than the bound

$$\|Mf\|_{L^{1,\infty}(\mathbb{R})} \leqslant 3\|f\|_{L^1(\mathbb{R})}.$$

One can also modify the proof to involve an infinite cover rather than a finite cover, a result called the *Wiener Covering Lemma*, which requires a slightly more technical algorithm.

Lemma 21.6. *Let X be a space satisfying the engulfing and doubling property, and also the following additional property:*

- *For all $\varepsilon > 0$, there exists $\delta > 0$ such that if $x \in X$ and $r \leqslant \delta$, $|B(x, r)| \leqslant \varepsilon$.*

386

Let $\mathcal{B}$ be a family of balls covering a set E, such that

$$\sup\{r_i : B(x_i, r_i) \in \mathcal{B}\} < \infty.$$

Then there exists a subfamily $\mathcal{B}' \subset \mathcal{B}$ of disjoint balls such that

$$|E| \lesssim \bigcup_{B \in \mathcal{B}'} B^*.$$

Proof. Let

$$C_1 = \sup\{r_i : B(x_i, r_i) \in \mathcal{B}\}.$$

As in the Vitali covering lemma, we greedily construct a disjoint family of balls. We start by picking a ball B_0 with diameter at least $C/2$ in the set. Given that we have chosen $\{B_1, \ldots, B_N\}$, we set

$$C_N = \sup\{r : B(x, r) \in \mathcal{B} \text{ and } B(x, r) \cap B_i = \varnothing \text{ for all } 1 \leqslant i \leqslant N\}.$$

If the right hand side is nonempty, we can choose a ball $B_{N+1} = B(x_{N+1}, r_{N+1})$ with $r_{N+1} \geqslant C_N/2$ which is disjoint from $B_1, \ldots, B_N$, and we continue the algorithm. Otherwise, the algorithm terminates. In either case, we let $\mathcal{B}'$ be the set of balls chosen. If

$$\sum_{B \in \mathcal{B}'} |B| = \infty,$$

then both sides of the inequality are infinity and there is nothing to be proven. Otherwise, we have $\sum_{B \in \mathcal{B}'} |B| < \infty$, and so $|B_i| \to 0$ as $i \to \infty$, so that $r_i \to 0$. We claim that

$$E \subset \bigcup_i B(x_i, 2cr_i).$$

If $B = B(x, r) \in \mathcal{B}$, then there therefore must exist i such that $r > 2r_i$. Let i_0 be the first such integer for which this is true. We must have $i_0 > 0$ since B_0 is chosen so that this inequality is false for all $B \in \mathcal{B}$. Thus for $0 \leqslant i < i_0$ we must have $r \leqslant 2r_i$. It must be true that $B \cap B_i \neq \varnothing$ for some $0 \leqslant i < i_0$ since otherwise we would have

$$r \leqslant C_N \leqslant 2r_{i_0}.$$

Thus the engulfing property implies that $B \subset B(x_i, 2cr_i)$. Thus we find

$$|E| \leqslant \sum_i |B(x_i, 2cr_i)| \lesssim \sum_i |B(x_i, r_i)|. \qquad \square$$

387

Under these assumptions, if we study the sets

$$E_\alpha = \{x \in X : Mf(x) > \alpha\},$$

associated with the maximal average of some $f \in L^1(X)$, then for each $x \in E_\alpha$, one may find a ball $B = B(x, r)$ such that

$$\int_B |f(x)|\, dx \geq \alpha |B|.$$

Since we also have

$$\int_B |f(x)|\, dx \leq \|f\|_{L^1(X)},$$

we have $|B| \leq \|f\|_{L^1(X)}/\alpha$. Our assumption thus implies that if α is suitably large, then r is bounded from above. Thus we may apply the Wiener covering lemma to find a disjoint subcover $\{B_i\}$ of these balls with

$$|E_\alpha| \lesssim \sum |B_i| \lesssim \frac{\|f\|_{L^1(X)}}{\alpha}.$$

Thus we can still obtain estimates on these sets without having to study compact subsets of E_α.

Finally, we consider a covering lemma which allows one to consider much more general families of measures than the ones satisfying the doubling and engulfing type properties considered above. We shall restrict ourselves to normal Euclidean balls in $\mathbb{R}^d$.

Lemma 21.7. *(Besicovitch Covering Lemma) Let A be a bounded subset of $\mathbb{R}^d$, and let $\mathcal{B}$ be a family of closed balls with centres in A. Then there is a subcollection of balls $\mathcal{B}' \subset \mathcal{B}$ which covers A, and with the property that each point in $\mathbb{R}^d$ is contained in at most $O_d(1)$ of the balls in $\mathcal{B}'$. Moreover, we can find $O(d)$ subfamilies $\mathcal{B}_1, \ldots, \mathcal{B}_N \subset \mathcal{B}$, each disjoint from one another, and each individually containing pairwise disjoint balls.*

Proof. TODO: See Matilla, Theorem 2.7. □

The Besicovitch Covering Lemma gives a Vitali like covering argument for arbitrary Radon measures on $\mathbb{R}^d$.

Theorem 21.8. *Let μ be a Radon measure on $\mathbb{R}^d$. if $A \subset \mathbb{R}^d$, and $\mathcal{B}$ is a family of closed balls such that each point in A is the centre of arbirarily small balls of $\mathcal{B}$, then we can find a disjoint family of balls $\mathcal{B}'$ such that*

$$\mu\left(A - \bigcup_{B' \in \mathcal{B}'} B'\right) = 0.$$

Proof. For simplicity, assume first that A is bounded (the general proof is given in Matilla, Theorem 2.8). Applying outer regularity, for any $\varepsilon > 0$, we can find an open set U containing A such that

$$\mu(U) \leqslant (1 + \varepsilon)\mu(A).$$

Then consider the family

$$\mathcal{B}' = \{B \in \mathcal{B} : B \subset U\}.$$

The assumption implies $\mathcal{B}'$ still covers A, so we can choose $N_d = O_d(1)$ families of balls $\mathcal{B}'_1, \ldots, \mathcal{B}'_N \subset \mathcal{B}'$ as in Besicovitch's covering theorem. But this means that

$$\mu(A) \leqslant \sum_{i=1}^N \sum_{B \in \mathcal{B}'_i} \mu(B) \leqslant .$$

There there exists some i_0 such that

$$\frac{\mu(A)}{N} \leqslant \sum_{B \in \mathcal{B}'_{i_0}} \mu(B).$$

Moreover, we can find a family subcolllection of balls $B_1, \ldots, B_K$ from $\mathcal{B}'_{i_0}$ such that

$$\frac{\mu(A)}{N} \leqslant 2(\mu(B_1) + \cdots + \mu(B_K)).$$

Let $A' = A - B_1 - \cdots - B_K$. Then

$$\mu(A') \leqslant \mu(U - B_1 - \cdots - B_K) = \mu(U) - \mu(B_1) - \cdots - \mu(B_K) \leqslant (1+\varepsilon)\mu(A) - (1/2N_d)\mu(A).$$

Choosing $\varepsilon = 1/4N_d$ gives $\mu(A') \leqslant (1 - 1/4N_d)\mu(A)$. Thus we may repeat this process infinitely to obtain a disjoint family of balls which give the required condition. $\square$

We obtain a maximal average theorem for general Radon measures using this covering argument.

Theorem 21.9. *Let μ be a Radon measure on $\mathbb{R}^d$, and define*

$$M_\mu f(x) = \sup_{r>0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f|\,d\mu.$$

Then

$$\|M_\mu f\|_{L^{1,\infty}(\mu)} \lesssim \|f\|_{L^1(\mu)}.$$

Proof. Fix $f \in L^1(\mu)$, and let

$$E_\alpha = \{x : M_\mu(f) > \alpha\}.$$

For each $x \in E_\alpha$, we can find a ball B such that

$$\int_B |f|\,d\mu \geqslant \alpha|B|.$$

Apply the Besicovitch Covering Lemma, finding a finite subcollection of these balls $\mathcal{B}$ covering E_α, and covering any other point in $\mathbb{R}^d$ at most $O_d(1)$ times. Thus

$$\mu(E_\alpha) \leqslant \sum_i \mu(B_i) \leqslant \alpha^{-1} \sum_i \int_{B_i} |f|\,d\mu \lesssim \frac{\|f\|_{L^1(\mu)}}{\alpha}.$$

$\square$

Interpolating between the trivial $L^\infty \to L^\infty$ bound, we obtain that $\|M_\mu f\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)}$ for all $1 < p \leqslant \infty$. Moreover, if for a finite Radon measure ν, we set

$$M_\mu \nu(x) = \sup_{r>0} \frac{\nu(B(x,r))}{\mu(B(x,r))},$$

then the same argument above implies that $\|M_\mu \nu\|_{L^{1,\infty}(\mu)} \lesssim \nu(\mathbb{R}^d)$.

Corollary 21.10. *If $f \in L^1(\mu)$, then*

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f(y)\,d\mu(y) = f(x)$$

for μ almost every $x \in \mathbb{R}^d$.

Proof. If $f \in C_c(\mathbb{R}^d)$, then it is simple to see that

$$\lim_{r \to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f(y) \, d\mu(y) = f(x)$$

for all $x \in \mathbb{R}^d$. But $C_c(\mathbb{R}^d)$ is dense in $L^1(\mu)$, i.e. by the Riesz representation theorem, so a density argument using the maximal function implies the corollary.

$\square$

21.3   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

$$\{Q_{n,k} : n \in \mathbb{Z}, k \in 2^n \, \mathbb{Z}^d\}$$

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 $Q_1, \ldots, Q_N$ is any collection of dyadic cubes, there exists an almost disjoint subcollection $Q_{i_1}, \ldots, Q_{i_k}$ 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 independent 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 $\|M_\Delta f\|_{L^{1,\infty}(\mathbb{R}^d)} \leqslant \|f\|_{L^1(\mathbb{R}^d)}$, with no implicit constant depending on d. The bound $\|M_\Delta f\|_{L^\infty(\mathbb{R}^d)} \leqslant \|f\|_{L^\infty(\mathbb{R}^d)}$ is easy, so interpolation gives $\|M_\Delta f\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}$ for all $1 < p \leqslant \infty$, with a constant now *independent of dimension*.

Two families of sets $\{B(x,\delta) : x \in \mathbb{R}^d, \delta > 0\}$ and $\{B'(x,\delta) : x \in \mathbb{R}^d, \delta > 0\}$ 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

391

$1/2^k$, 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 $x \in \mathbb{R}^d$,

$$M_\Delta f(x) \lesssim_d Mf(x).$$

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 $\mathbb{R}$. 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 21.11. *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

$$a = 0.a_1 a_2 \ldots \quad \text{and} \quad b = b_1 b_2 \ldots.$$

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

$$Q_1 = \left[0.a_1 \ldots a_{n-1}, 0.a_1 \ldots a_{n-1} + 1/2^{n-1}\right]$$

which has length $1/2^{n-1}$. Find $0 \leqslant i < \infty$ such that

$$a = 0.a_1 \ldots a_{n-1} 01^i 0 \ldots$$

and $0 \leqslant j < \infty$ such that

$$b = 0.a_1 \ldots a_{n-1} 10^j 1.$$

If no such j exists, then $b = 0.a_1 \ldots a_{n-1} 1$, and so $[a,b]$ is contained in the rational interval

$$Q_2 = \left[0.a_1 \ldots a_{n-1} 01^i, 0.a_1 \ldots a_{n-1} 01^i + 1/2^{n+i}\right]$$

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

$$1/3 + Q_3 = \left[0.a_1 \ldots a_{n-1} 01^{\min(i,j)-r} 1010 \ldots, 0.a_1 \ldots a_{n-1} 01^{i-r} 1010 \cdots + 1/2^{n+\min(i,j)-r}\right]$$

and so $1/3 + Q_3$ contains $[a,b]$ and $|Q_3| = 1/2^{n+\min(i,j)-r} \leqslant 2^5(b-a)$. $\qquad \square$

It follows that for each $x \in \mathbb{R}^d$, and any function f,

$$Mf(x) \lesssim_d (M_\Delta f)(x) + (M_\Delta \text{Trans}_{1/3} f)(x).$$

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 $\mathbb{R}^d$ into an almost disjoint union of cubes. If we define the conditional expectation operators

$$E_n f(x) = \sum_{Q \in \mathcal{B}(n)} \left(\frac{1}{|Q|} \fint_Q f \right) \cdot \mathbb{I}_Q$$

then $M_\Delta f = \sup_{n \in \mathbb{Z}} E_n f$. In particular, it is easy to see from the bounds on M_Δ that for any $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, $\lim_{n \to \infty} E_n f(x) = f(x)$ holds for almost every $x \in \mathbb{R}^d$. It is simple to conclude from this result a very useful technique, known as the *Calderón-Zygmund decomposition*.

Theorem 21.12. *Given $f \in L^1(\mathbb{R}^d)$ and $\lambda > 0$, we can write $f = g + b$, where $\|g\|_{L^\infty(\mathbb{R}^d)} \lesssim_d \lambda$, and there is an almost disjoint family of dyadic cubes $\{Q_i\}$ such that g is supported on $\bigcup_i Q_i$,*

$$\sum_i |Q_i| \leq \frac{\|f\|_{L^1(\mathbb{R}^d)}}{\lambda},$$

and for each i,

$$\int_{Q_i} f(y) \, dy = 0.$$

We also have $\|g\|_{L^1(\mathbb{R}^d)}, \|b\|_{L^1(\mathbb{R}^d)} \lesssim_d \|f\|_{L^1(\mathbb{R}^d)}$.

Proof. Write $E = \{x : M_\Delta f(x) > \lambda\}$. By the dyadic Hardy-Littlewood maximal inequality,

$$|E| \leq \frac{\|f\|_{L^1(\mathbb{R}^d)}}{\lambda}.$$

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 $\|g\|_{L^\infty(\mathbb{R}^d)} \lesssim_d \lambda$. 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.

21.4 Linearization of Maximal Operators

It is often a very useful technique to study a maximal operator by reducing it's analysis to studying a family of linear operators. The technique is very general, and applies to virtually all the maximal functions, so we give the general discussion here. Fix a family of linear operators $\{T_t\}$, and consider the associated maximal operator $Mf = \sup_t |T_t f|$. By monotone convergence, we can obtain $L^p \to L^q$ operator norm bounds on the maximal operator M if we can prove uniform

$L^p \rightarrow L^q$ operator norm bounds on the maximal operators $Mf = \sup_{t \in T} |T_t f|$, where $T = \{t_1, \ldots, t_N\}$ is a finite subset of indices. For each j, the set

$$\{x : \sup_{t \in T} |T_t f| = |T_{t_j} f|\}$$

is measurable. Thus we can divide Y into N disjoint sets $Y_1 \cup \cdots \cup Y_N$, such that

$$Mf = \sum_{i=1}^{N} \mathbb{I}_{Y_i} |T_i f|.$$

If we write $\mathbb{I}_{Y_i} |T_i f| = \chi_i T_i f$, then we can further write

$$Mf = \sum_{i=1}^{N} \chi_i T_i f.$$

To bound the *sublinear* operator M, it therefore suffices to obtain uniform bounds on the behaviours of the family of *linear operators* of the form

$$\sum_{i=1}^{N} \chi_i T_i.$$

This is the method of linearization.

21.5 TT^* **Arguments**

The method of TT^* arguments enables us to obtain bounds on a linear operator T by exploiting cancellation to understand TT^*. The main calculation from which the method follows is that if H is a Hilbert space, and X is a Banach space, and $T : H \rightarrow X$ is a bounded operator, then

$$\|TT^*\|_{X^* \rightarrow X} = \|T\|_{H \rightarrow X}^2,$$

because certainly $\|T^*\|_{X^* \rightarrow H} = \|T\|_{H \rightarrow X}$. For any $f \in X^*$,

$$\|T^* f\|_H^2 = \langle T^* f, T^* f \rangle = \langle TT^* f, f \rangle \leqslant \|TT^*\|_{X^* \rightarrow X} \|f\|_{X^*}^2,$$

so $\|T\|^2 \leqslant \|TT^*\|$, and conversely,

$$\|TT^*\| \leqslant \|T\| \|T^*\| = \|T\|^2.$$

Thus we study the boundedness of an operator T via the boundedness of the associated operator TT^*.

Let us apply this method to prove the Hardy-Littlewood Maximal Inequality. By linearization, it suffices to obtain uniform bounds on the behaviour of the linear operators of the form

$$T_r f(y) = A_{r(y)} f(y),$$

where $r : Y \to (0, \infty)$ is measurable. Now

$$T_r^* g(x) = \int_{|x-y| \leqslant r(y)} \frac{g(y)}{|B(y, r(y))|} \, dy$$

Thus

$$T_r T_r^* g(y) = A_{r(y)} T^* g(y)$$
$$= \frac{1}{V_d^2} \int_{|x-y| \leqslant r(y)} \int_{|x-z| \leqslant r(z)} r(y)^{-d} r(z)^{-d} g(z) \, dx \, dz.$$

On the integrand, we must have $|z - y| \leqslant r(y) + r(z)$, and x is constrained to lie in a ball of radius $\min(r(y), r(z))$. Thus

$$|T_r T_r^* g(y)| \leqslant \frac{1}{V_d} \int_{|y-z| \leqslant r(y)+r(z)} \frac{g(z)}{\max(r(y), r(z))^d} \, dz.$$

We can split this integral into two parts, depending on whether $r(y) \geqslant r(z)$ or $r(y) \leqslant r(z)$, which yields

$$|T_r T_r^* g(y)| \leqslant \frac{1}{V_d} \int_{|y-z| \leqslant 2r(y)} \frac{g(z)}{r(y)^d} \, dy + \int_{|y-z| \leqslant 2r(z)} \frac{g(z)}{r(z)^d} \, dy$$
$$\leqslant 2^d (T_{2r} g(y) + T_{2r}^* g(y)).$$

Thus if $C_N = \sup_r \|T_r\|_{L^2 \to L^2}$, where the supremum is over functions taking over at most N values, then we conclude that

$$\|T_r T_r^* g(y)\|_{L^2} \leqslant 2^d (\|T_{2r} g\|_{L^2} + \|T_{2r}^* g\|_{L^2}) \leqslant 2^{d+1} C_N,$$

so that for any r, we find that $\|T_r\|_{L^2 \to L^2}^2 = \|T_r T_r^*\|_{L^2 \to L^2}^{1/2} \leqslant 2^{d+1} C_N$. Thus $C_N^2 \leqslant 2^{d+1} C_N$, and since $C_N < \infty$ (we have a trivial bound $C_N \lesssim N$ because the operator norm of the averages is uniformly bounded), this gives $C_N \leqslant 2^{d+1}$, which shows $\|Mf\|_{L^2(\mathbb{R}^d)} \leqslant 2^{d+1} \|f\|_{L^2(\mathbb{R}^d)}$.

21.6 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 21.13 (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 21.14. *Let $E \subset \mathbb{R}^d$ be a set of nonzero Lebesgue measure. Then for any non-zero $a_1, \ldots, a_N \in \mathbb{R}$ there exists $x \in E$ and $c \in \mathbb{R}$ such that*

$$a_1 x + c, \ldots, a_N x + c \in E.$$

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 $E \cap a_1^{-1}E \cap \cdots \cap a_N^{-1}E$, which means the set is nonempty. If $y \in E \cap a^{-1}E \cap \cdots \cap a_N^{-1}E$, then $y, a_1 y, \ldots, a_N y \in E$. $\qquad \square$

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 21.15. *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

$$\lim_{\delta \to 0} \fint_{B(x,\delta)} |f(y) - f(x)| \, dy \leqslant \limsup_{\delta \to 0} \fint_{B(x,\delta)} \left(|f(y) - p| + |p - f(x)| \right) \, dy \leqslant 2\varepsilon.$$

We then let $\varepsilon \to 0$. Since $f(x)$ is finite for almost all x, this completes the proof. $\qquad\square$

It is interesting to note that for any $f \in L^1_{loc}(\mathbb{R}^d)$, there is $g \in L^1_{loc}(\mathbb{R}^d)$ such that $f = g$ almost everywhere, and the Lebesgue set of g is maximal. One choice is to define

$$g(x) = \limsup_{\delta \to 0} \fint_{B(x,\delta)} f(y) \, dy.$$

Often the Lebesgue set of f is defined to be the Lebesgue set of g, when one wants to think of the Lebesgue set as a distributional invariant of f rather than depending solely on the pointwise behaviour.

21.7 Ergodic Averages

Ergodic theory studies the behaviour of measure preserving dynamical systems. By this, we mean, for a probability space X, a bimeasurable bijection $T : X \to X$ such that $|T^{-1}(E)| = |E|$ for all $E \subset X$. A classic example, for $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, is the shift map $T : \mathbb{T} \to \mathbb{T}$ given by $T(x) = x + \alpha$. The map T induces a linear operator T, such that $Tf(x) = f(Tx)$. That T is measure preserving implies that T is an isometry of $L^p(X)$ for any p. We begin with a result of Von Neumann.

Theorem 21.16 (Von-Neumann Ergodic Theorem). *Let* $T : H \to H$ *be a unitary operator. Then for any* $f \in H$, *the averages*

$$A_N f = \frac{f + \cdots + T^{N-1} f}{N}$$

converge as $N \to \infty$ *in* H.

398

Proof. If $Tf = f$, then it is clear that $A_N f \to f$ as $N \to \infty$. Conversely, if $f = Tg - g$ for some $g \in H$, then the averages telescope, and we obtain that $A_N f = N^{-1}(T^N g - g)$, which converges to zero since T is unitary. But these two cases turn out to be orthogonal complements of one another, i.e. we can write H as the orthogonal sum of the closures of the two subspaces

$$H_1 = \{f \in H : Tf = f\} \quad \text{and} \quad H_2 = \{f \in H : f = Tg - g \text{ for some } g \in H\}.$$

These spaces are clearly orthogonal; if $Tf_1 = f_1$ and $f_2 = Tg - g$, then

$$\langle f_1, g \rangle = \langle Tf_1, Tg \rangle = \langle f_1, Tg \rangle$$

and thus

$$\langle f_1, f_2 \rangle = \langle f_1, Tg \rangle - \langle f_1, g \rangle = 0.$$

To show these spaces are orthogonal complements, let f be orthogonal to the closure of H_2. Then $\langle f, Tf - f \rangle = 0$. But this means that

$$\langle f, Tf \rangle = \langle f, f \rangle = \langle Tf, Tf \rangle,$$

and so $\langle Tf, Tf - f \rangle = 0$. But this means that $\langle Tf - f, Tf - f \rangle = 0$, so $Tf = f$. $\qquad\square$

Now we move onto the mean ergodic theorem.

Theorem 21.17. *Suppose $T : X \to X$ is measure preserving and $1 \leqslant p < \infty$. Then if $f \in L^p(X)$, then $\{A_N f\}$ converges in $L^p(X)$.*

Proof. The operators $\{A_N\}$ are uniformly bounded on $L^p(X)$. Thus it suffices to verify the theorem for a dense subclass of functions f. So suppose $f \in L^\infty(X)$. The functions $\{A_N f\}$ converge in $L^2(X)$ and thus converge in $L^p(X)$ for any $1 \leqslant p < \infty$ by Hölder's inequality. The functions are also *bounded* in $L^\infty(X)$, so interpolation gives convergence in $L^p(X)$ for any $1 \leqslant p < \infty$. $\qquad\square$

The functions $A_N f$ do not necessarily converge in $L^\infty(X)$ if $f \in L^\infty(X)$. But they do converge *almost everywhere*, as we will prove using our theory of maximal functions we have developed. For functions f on X, define

$$Mf(x) = \sup_{N \geqslant 1} \left| \frac{1}{N} \sum_{n=0}^{N-1} T^n f(x) \right|.$$

399

We trivially have $\|Mf\|_{L^\infty(X)} \lesssim \|f\|_{L^\infty(X)}$, and we will now justify that $\|Mf\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$. The space X has no metric properties, so it seems surprising that we can use our previous theory of maximal functions here. But the idea is to *lift* our transform to a space where we *do* have such properties, thus obtaining a *transference principle* reducing the study of general measure preserving systems to particular ones.

Theorem 21.18. *Let $T : X \to X$ be a measure preserving system. Then for $f \in L^1(X)$,*

$$\|Mf\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}.$$

Proof. We have $M = \lim_{K\to\infty} M_K$, where

$$M_K f(x) = \sup_{1 \leqslant N < K} \frac{1}{N} \sum_{n=1}^{N} T^n f(x).$$

For any $K > 0$, define $F : X \times [2K] \to \mathbb{C}$ by setting

$$F(x, n) = T^n f(x).$$

The space $[K]$ is a space to which we can apply the Hardy-Littlewood theory of maximal functions we have developed, *uniformly in K*, if we equip it with the uniform measure, which is doubling with respect to the standard metric on $[K]$. Thus we obtain that for each $x \in X$,

$$\left\| \sup_{1 < N \leqslant K} N^{-1} \sum_{n=1}^{N} F(x, n+k) \right\|_{l_k^{1,\infty}[K]} \lesssim \|F(x, k)\|_{l_k^1[2K]}.$$

Apply the $L_x^{1,\infty}$ norm to the left hand side of the equation, and the L^1 norm to the right hand side. Since T is measure preserving, $\|F(x, k)\|_{L_x^1(X)l_k^1[2K]} = \|f\|_{L^1(X)}$. Thus

$$\left\| \sup_{1 < N \leqslant K} \frac{1}{N} \sum_{n=1}^{N} F(x, n+k) \right\|_{L^{1,\infty}(X\times[K])} \lesssim \|f\|_{L^1(X)}.$$

To work with the left hand side, notice the identity

$$\sup_{1 \leqslant N < K} \frac{1}{N} \sum_{n=1}^{N} T^{n+k} f(x) = \sup_{1 \leqslant N < K} \frac{1}{N} \sum_{n=1}^{N} T^n (T^k f)(x).$$

Thus the left hand side is equal to $\| \|M_K f\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$. Taking $K \to \infty$ yields the required inequality. $\square$

Corollary 21.19 (The Pointwise Ergodic Theorem). *Let $T : X \to X$ be a measure preserving system. If $f \in L^1(X)$, then $A_N f(x)$ converges for almost every $x \in X$.*

Proof. Our maximal function bounds thus imply that the result holds for a dense subclass of $L^1(X)$. The result is obvious for a function $f \in L^2(X)$ with $Tf = f$. The theorem is also true for $f = g - Tg$, where $g \in L^\infty(X)$. But the sums of such functions are dense in $L^2(X)$ by the orthogonal decomposition we just proved, and thus dense in $L^1(X)$. $\qquad\square$

21.8 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 $\mathbb{R}^d$ 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 $\delta > 0$,

$$\lim_{\alpha \to 0} \int_{|x| \geq \delta} |K_\alpha(x)|\, dx \to 0.$$

It requires only basic analysis to verify good kernel convergence.

Theorem 21.20. *If $\{K_\alpha\}$ is a good kernel, then for any absolutely integrable function f, $\lim_{\alpha \to 0} f * K_\alpha = f$ in $L^1(\mathbb{R}^d)$, and for any continuity point x of f, $\lim_{\alpha \to 0}(f * K_\alpha)(x) = f(x)$.*

Proof. Note that

$$\begin{aligned}
\|(f * K_\alpha) - f\|_1 &= \int |(f * K_\alpha)(x) - f(x)|\, dx \\
&= \int \left| \int K_\alpha(y)[f(x-y) - f(x)]\, dy \right| dx \\
&\leq \int |K_\alpha(y)| \|T_y f - f\|_1\, dy
\end{aligned}$$

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 $\varepsilon \to 0$. $\qquad \square$

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

$$\int_{|x| \geqslant \varepsilon} |K_\delta(x)| \leqslant \int_{\varepsilon}^{\infty} \int_{S^{d-1}} \frac{C\delta}{r} \, d\sigma dr = C\delta |S^{n-1}| \int_{\varepsilon}^{\infty} \frac{dr}{r} \leqslant \frac{C\delta |S^{n-1}|}{\varepsilon}$$

402

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

$$P_y(x) = y^{-1}P_1(xy^{-1})$$

403

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 21.21. *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. □

21.9 H^p Spaces On the Unit Disk

We briefly describe how the Hardy-Littlewood maximal principle can be used to develop the theory of Hardy spaces. Recall that if $f : \mathbb{D} \to \mathbb{C}$ is a harmonic function on the interior of the unit disk, and we define $f_r : \mathbb{T} \to \mathbb{C}$ by setting $f_r(t) = f(re^{2\pi i t})$, then for the *Poisson kernel*

$$P_t(x) = \frac{1 - t^2}{1 + t^2 - 2t \cos(2\pi x)} = \text{Re}\left(\frac{e^{2\pi i x} + t}{e^{2\pi i x} - t}\right),$$

we have for $0 \leqslant s < r < 1$, $f_s = f_r * P_{s/r}$. Applying Young's inequality with the fact $\{P_t\}$ is positive and with unit mass, we thus conclude that $s \mapsto \|f_s\|_{L^p(\mathbb{T})}$ is a monotone increasing function of s. We define $H^p(\mathbb{D})$ to be the space of all *holomorphic* (and thus harmonic) functions for which the norm

$$\|f\|_{H^p(\mathbb{D})} = \lim_{t \to 1} \|f_t\|_{L^p(\mathbb{T})}.$$

We have seen that the Poisson kernel is an approximation to the identity, so that we have a pointwise estimate $|f * P_t| \lesssim Mf$.

Theorem 21.22. *If $f \in H^p(\mathbb{D})$, then for $1 < p \leqslant \infty$, the function*

$$f_1 = \lim_{t \to 1} f_t$$

exists pointwise almost everywhere, and lies in $L^p(\mathbb{T})$. If $1 < p < \infty$, then $\{f_t\}$ also converges to f_1 in L^p norm.

Proof. Fix $0 < t < 1$. For $g \in L^{p^*}(\mathbb{T})$, we verify that

$$\langle f_r, g * P_t \rangle = \langle f_r * P_t, g \rangle = \langle f_{rt}, g \rangle = \langle f_t * K_r, g \rangle.$$

As $r \to 1$, $f_t * K_r$ converges to f_t in the L^p norm, so we conclude that there exists a quantity A_t such that as $r \to 1$,

$$\langle f_r, g * P_t \rangle \to A_t.$$

We have $|A_t| \lesssim \|f\|_{H^p(\mathbb{D})} \|g\|_{L^{p^*}(\mathbb{T})}$, uniformly in t. As $t \to 1$, $g * P_t$ converges to g in the $L^{p^*}(\mathbb{T})$ norm, and thus because $\{f_r\}$ is bounded in $L^p(\mathbb{T})$, uniformly in r, we conclude that $\langle f_r, g * P_t \rangle$ converges to $\langle f_r, g \rangle$ uniformly in r. But this is enough to justify the convergence of $\langle f_r, g \rangle$. Thus we have justified that there exists $f_1 \in L^p(\mathbb{T})$ such that $f_r \to f_1$ weakly in $L^p(\mathbb{T})$. But this means that

$$\langle f_t, g \rangle = \lim_{r \to 1} \langle f_t * P_r, g \rangle = \lim_{r \to 1} \langle f_r, g * P_t \rangle = \langle f_1, g * P_t \rangle = \langle f_1 * P_t, g \rangle.$$

Thus $f_t = f_1 * P_t$. From our results on approximations to the identity proved in this chapter, the remaining results are now immediate. $\square$

21.10 The Strong Maximal Function

TODO

21.11 The Tangential Poisson Maximal Function

TODO

Chapter 22

Aside: Differentiability of Measurable Functions

A simple consequence of the results of maximal functions considered in the last chapter is a kind of fundamental theorem of calculus. If $f \in L^1(\mathbb{R})$, then we can define $F \in C(\mathbb{R})$ 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 $f' \in L^1(\mathbb{R})$, and such that

$$f(t) = \int_{-\infty}^{t} f'(s)\, ds$$

holds for almost every $t \in \mathbb{R}$. 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 $P = a \leqslant t_0 \leqslant \cdots \leqslant t_n$, 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,

$$L(x) = \int_a^b \sqrt{x_1'(t)^2 + \cdots + x_n'(t)^2} \, dt$$

Considering rectifiable curves leads directly to the notion of a function with bounded variation.

Theorem 22.1. *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. $\qquad \square$

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 22.2. *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. □

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 22.3. *If f is a continuous function of bounded variation, then*

$$x \mapsto V(f, a, x) \quad x \mapsto V(x, b)$$

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

$$P = \{a = t_0 < \cdots < t_n = x\}$$

410

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

$$V(f, a, x_1) \geqslant V(f, Q) = V(f, P) - |f(x) - f(t_{n-1})| + |f(x_1) - f(t_{n-1})|$$
$$\geqslant V(f, P) - \varepsilon \geqslant V(f, a, x) - \varepsilon$$

A similar argument gives continuity on the right, and the continuity as the left bound of the interval changes. $\square$

To obtain the differentiation theorem for functions of bounded variation, we require a lemma of F. Riesz.

Lemma 22.4 (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. $\square$

Theorem 22.5. *If f is increasing and continuous on $[a, b]$, then f is differentiable almost everywhere. That is,*

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

411

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

$$|E_\gamma| \leqslant \sum(b_k - a_k) \leqslant \frac{1}{\gamma} \sum f(b_k) - f(a_k) \leqslant \frac{f(b) - f(a)}{\gamma}$$

412

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

$$R(b_{nm} - a_{nm}) \leqslant f(b_{nm}) - f(a_{nm}) \quad f(b_n) - f(a_n) \leqslant r(b_n - a_n)$$

then, because f is increasing,

$$|U_N| = \sum_{nm}(b_{nm} - a_{nm}) \leqslant \frac{1}{R}\sum_{nm}(f(b_{nm}) - f(a_{nm}))$$

$$\leqslant \frac{1}{R}\sum f(b_n) - f(a_n) \leqslant \frac{r}{R}\sum_n(b_n - a_n) \leqslant \frac{r}{R}|I_N|$$

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 22.6. *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

$$\int_a^b f'(x) \leqslant \liminf_{n\to\infty} \int_a^b g_n(x) = \liminf_{n\to\infty} n \int_a^b [f(x + 1/n) - f(x)]$$

$$= \liminf_{n\to\infty} n \left[\int_b^{b+1/n} f(x) - \int_a^{a+1/n} f(x) \right] = f(b) - f(a)$$

413

where the last equality follows because f is continuous. ∎

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.

22.1 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 $(a_1, b_1), \ldots, (a_n, b_n)$ 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

$$\int_a^b f'(x)\, dx = f(b) - f(a)$$

414

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 22.7. *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 22.8. *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 $B_1, \ldots, B_N$ 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 $B_1, \ldots, B_{n_0}$ 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 $B_1, \ldots, B_{n_0}$, 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 $B_1^1, \ldots, B_{n_1}^1$ 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 $B_{11}, \ldots, B_{1n_1}$ 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 $B_1, \ldots, B_{n_1}$, we still have a Vitali cover of E_1. Inductively, we can then find a disjoint family of balls $B_{k1}, \ldots, B_{kn_k}$ 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. $\qquad\square$

Theorem 22.9. *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. $\qquad\square$

Theorem 22.10. *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.

417

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. $\square$

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 22.11. *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. $\qquad\square$

This property of absolutely continuous functions is independent 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 22.12. *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. $\qquad\square$

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 22.13. *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. $\qquad\square$

22.2 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 **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

$$j_y(x) = \begin{cases} 0 & : x < y \\ \theta(y) & : x = y \\ 1 & x > y \end{cases}$$

420

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 22.14. *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 $x_1, x_2, \ldots$, 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 22.15. *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.

22.3 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 22.16. *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 $P = \{a \leqslant t_1 < \cdots < t_n \leqslant b\}$ 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 $a = t_0 < \cdots < t_N$, where G is constant on each (t_n, t_{n+1}), then

$$V(G, a, b) \geqslant \sum |G(t_n) - G(t_{n-1})| = \sum \left| \int_{t_{n-1}}^{t_n} g(t) \, dt \right|$$

$$= \sum \int_{t_{n-1}}^{t_n} |g(t)| \, dt = \int_a^b |g(t)| \, dt \geqslant \|f'\|_1 - \varepsilon$$

Letting $\varepsilon \to 0$ now gives the result. $\qquad\square$

423

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 22.17. *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. $\qquad \square$

22.4   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 22.18. *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

$$\left| \int f(x) \varphi'(x) \right| = \left| \lim_{h \to 0} \int f(x) \frac{\varphi(x+h) - \varphi(x)}{h} \right|$$
$$= \left| \lim_{h \to 0} \int \frac{f(x-h) - f(x)}{h} \varphi(x) \right| \leqslant A \|\varphi\|_\infty$$

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 $\qquad\qquad\qquad\qquad\square$

22.5 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

$$\lim_{\delta \to 0} \frac{1}{|B^{n-m}|} \frac{|K^\delta|}{\delta^m},$$

where B^d is the unit ball in $\mathbb{R}^d$, and $|B^d|$ denotes it's volume. 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 22.19. *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. $\square$

Theorem 22.20. *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

$$M^*(\Gamma) = \limsup_{\delta \to 0} \frac{|\Gamma|^\delta}{\alpha(n-1)\delta} \quad M_*(\Gamma) = \liminf_{\delta \to 0} \frac{|\Gamma|^\delta}{\alpha(n-1)\delta}$$

426

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 $I_1, \ldots, I_{M+1}$, 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

$$NL\left(2\delta/N + O(\delta\varepsilon/N + \delta^2 + \varepsilon/N^2)\right) = 2\delta L + O(\delta\varepsilon + \delta^2 N + \varepsilon/N)$$

427

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 $M(\Gamma) = L$. $\qquad\square$

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 22.21. *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 $|\Gamma^\delta| \lesssim \delta^{2-\alpha-1/\alpha}$. $\qquad\square$

22.6 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 22.22. *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

429

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 22.23. *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 22.24. *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. $\square$

22.7 Differentiability of Measures

Using the Besicovitch covering theorem, we can obtain some differentiability properties of Radon measure on $\mathbb{R}^d$. For two such measures μ and ν on $\mathbb{R}^d$, define

$$\overline{\frac{d\mu}{d\nu}}(x) = \limsup_{r \to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))}$$

and

$$\underline{\frac{d\mu}{d\nu}}(x) = \liminf_{r \to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))}.$$

The derivative is defined when both quantities agree with one another, i.e. precisely when the limit

$$\frac{d\mu}{d\nu}(x) = \lim_{r \to 0} \frac{\nu(B(x,r))}{\mu(B(x,r))}$$

exists.

Theorem 22.25. *Let μ and ν be Radon measures on $\mathbb{R}^d$. Then the derivative $d\mu/d\nu$ exists for ν almost every $x \in \mathbb{R}^d$, and for any Borel set $E \subset \mathbb{R}^d$,*

$$\int_E \frac{d\mu}{d\nu}(x)d\nu(x) \leqslant \mu(B),$$

with equality if μ is absolutely continuous with respect to ν. Moreover, if

$$\underline{\frac{d\mu}{d\nu}}(x) < \infty$$

for μ almost every $x \in \mathbb{R}^d$, then μ is absolutely continuous with respect to ν.

Proof. TODO: See covering lemma argument in 2.13 of Matilla. $\qquad\square$

Chapter 23

Singular Integral Operators

We now return to the study of kernel operators

$$Tf(x) = \int K(x, y) f(y) \, dy$$

where $K(x, y)$ is singular on the diagonal $\Delta = \{(x, y) : x = y\}$. Such operators are called *singular integrals*, and often arise in many parts of harmonic analysis, complex analysis, and the analysis of partial differential equations.

23.1   Examples of Singular Integral Operators

The prototypical example of a singular integral operator is the Hilbert transform on $\mathbb{R}$, i.e. the operator

$$Hf(x) = \int \frac{f(y)}{x - y} \, dy,$$

which has kernel

$$K(x, y) = \frac{1}{x - y}.$$

This kernel is singular on the diagonal, to the point where we cannot interpret the integral in the definition of the Hilbert transform in the Lebesgue sense, since the function $y \mapsto 1/y$ is not Lebesgue integrable (but only 'barely'). There are several ways to fix this, the most common is to interpret the integral as a *principal value*, i.e. noticing that for any Schwartz function $f \in \mathcal{S}(\mathbb{R})$, and any $x \in \mathbb{R}$, the limit

$$\lim_{\varepsilon \to 0} \int_{|y| > \varepsilon} \frac{f(y)}{y} \, dy$$

exists, which we call the principal value of the integral, often denoted by

$$\text{p.v} \left(\int \frac{f(y)}{y} \, dy \right).$$

This quantity gives rise to a bounded linear functional on $\mathcal{S}(\mathbb{R})$, i.e. a tempered distribution which we denote by p.v$(1/x)$. The Hilbert transform H is then precisely the operator given by convolution against this tempered distribution, i.e.

$$Hf(x) = (\text{p.v}(1/x) * f)(x)$$
$$= \lim_{\varepsilon \to 0} \int_{|y-x| \geqslant \varepsilon} \frac{f(y)}{x-y} \, dy.$$

As defined, this operator is now bounded from $\mathcal{S}(\mathbb{R})$ to $\mathcal{E}(\mathbb{R})$. One very useful trick to studying the L^p mapping properties of these operators is to take Fourier transforms.

Lemma 23.1. *The Fourier transform of p.v$(1/x)$ is $-i\pi \text{sgn}(x)$.*

Proof. Let $u(x) = \text{p.v}(1/x)$. By general properties of the Fourier transform $\hat{u}$ is smooth away from the origin, odd, and homogeneous of degree zero. Since $x \cdot u(x) = 1$, we conclude that

$$D_\xi \hat{u} = -2\pi i \delta.$$

The only odd distribution with this property is $\hat{u}(\xi) = -i\pi \text{sgn}(\xi)$. □

Since the Fourier transform of u is a *bounded function*, it follows from Parseval's identity that

$$\|Hf\|_{L^2} = \|u * f\|_{L^2} = \|\hat{u}\hat{f}\|_{L^2} \lesssim \|\hat{f}\|_{L^2} = \|f\|_{L^2}.$$

Thus H is a bounded operator on $L^2(\mathbb{R})$. The Calderon-Zygmund theory we will describe in this chapter uses this property to show that H is also weakly bounded on $L^1(\mathbb{R})$, and thus bounded on $L^p(\mathbb{R})$ for all $1 < p \leqslant 2$. By duality, it actually follows that H is bounded on $L^p(\mathbb{R})$ for all $1 < p < \infty$. This is a general characteristic of this method; the Calderon-Zygmund theory will give us weak boundedness in the L^1 norm, provided we already have some boundedness in the L^p norm for some $p > 1$, and we must use other techniques, such as orthogonality, to justify this initial boundedness.

Another example of a singular integral to which the Calderon-Zygmund theory will apply is the *Riesz transforms* $\{R_j\}$ on $\mathbb{R}^d$, for $1 \leqslant j \leqslant d$, given by convolution against the *Riesz Kernels*

$$r_j(x) = \text{p.v} \left(\frac{x_j}{|x|^{d+1}} \right),$$

i.e. we have $R_j f = r_j * f$. The singular nature again arises by the fact that r_j is singular at the origin. We also observe that r_j is homogeneous of degree $-d$, like the convolution kernel associated with the Hilbert transform was homogeneous of degree -1, which means the kernel is not Lebesgue integrable, but *only barely*, in the sense that a less negative homogeneity would be locally integrable at the origin, and a more negative homogeneity would be integrable away from the origin. Taking Fourier transforms is also useful to analyze the Riesz transform, since it's Fourier transform is a bounded function, which gives the same kind of L^2 boundedness we obtain for the Hilbert transform.

Lemma 23.2. *We have*

$$\hat{r}_j(\xi) = \frac{-i\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} \frac{\xi_j}{|\xi|},$$

an element of $L^\infty(\mathbb{R}^d)$, which implies R_j is bounded on $L^2(\mathbb{R}^d)$.

Proof. Proof: TODO. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In 1952, motivated by the analysis of homogeneous constant coefficient partial differential operators, Calderon-Zygmund introduced their more general theory of singular integrals, which includes the Riesz and Hilbert transforms as special cases. They analyzed singular integral operators on $\mathbb{R}^d$ of the form

$$Tf = k * f,$$

where k is a distribution which is homogeneous of degree $-d$, smooth away from the origin, and satisfies the cancellation condition

$$\int_{|x|=1} k(x)\, d\sigma(x) = 0.$$

Let us call such an operator a *classical Calderon-Zygmund operator*. If m is a bounded function on $\mathbb{R}^d$ which is homogeneous of degree zero and smooth away from the origin, and if k is the inverse Fourier transform of m, then k will be a

classical Calderon-Zygmund kernel. If $P(D)$ is a constant coefficient, homogeneous partial differential operator of order k, then $T = (-\Delta)^{-k/2} \circ P(D)$ will be an operator of this form, since it is given by convolution with the inverse Fourier transform of $m(\xi) = |\xi|^{-k} P(\xi)$. *More interestingly*, if P is *elliptic*, i.e. if the equation $P(\xi) = 0$ has no non-zero real solutions, then the operator T given by convolution with the inverse Fourier transform of $m(\xi) = |\xi|^k / P(\xi)$ will be of this form. Moreover, this operator has the property that if f and g are Schwartz, and $P(D)f = (-\Delta)^{k/2} g$, then $f = Sg$, i.e. S gives the inverse to $(-\Delta)^{-k/2} \circ P(D)$. The theory we give here will thus have the consequence that if $P(D)f = g$ and $g \in L^p(\mathbb{R}^d)$, then $f \in L_k^p(\mathbb{R}^d)$ – a *regularity estimate* for solutions to equations involving the partial differential equation $P(D)$.

The same method that Calderon and Zygmund introduced applies to a much more general family of operators given by a convolution kernel k. In particular, the theory can be applied to any measurable function $k : \mathbb{R}^d - \{0\} \to \mathbb{C}$ satisfying the following three conditions:

- $|k(x)| \lesssim 1/|x|^d$ for all $x \in \mathbb{R}^d$.

- For all $y \neq 0$,
$$\int_{|x| > 2|y|} |k(x) - k(x - y)| \, dx \lesssim 1.$$

- For all $0 < r < s < \infty$,
$$\int_{r < |x| < s} k(x) \, dx = 0.$$

The distribution k is then called a *Calderon-Zygmund kernel of convolution type*. One can then interpret k as a tempered distribution by considering the principal value at the origin. Given a pointwise differentiable function $k : \mathbb{R}^d - \{0\} \to \mathbb{C}$, the second condition is guaranteed if one has estimates of the form

$$|\nabla k(x)| \lesssim \frac{1}{|x|^{d+1}}$$

In this case, we call k a *strong Calderon-Zygmund kernel of convolution type*. The Fourier transform of a Calderon-Zygmund convolution kernel k is bounded, so the associated convolution operator T is automatically bounded on $L^2(\mathbb{R}^d)$, as we saw with the Hilbert and Riesz transforms.

Lemma 23.3. *Let k be a Calderon-Zygmund kernel. Then* $\|\hat{k}\|_{L^\infty(\mathbb{R}^d)} \lesssim A.$

Proof. For any $0 < r < s$, we will show that

$$m_{r,s}(\xi) = \int_{r<|x|<s} k(x)e^{-2\pi i \xi \cdot x} \, dx.$$

is uniformly bounded in L^∞, from which the result will follow. Let $a = \max(r, |\xi|^{-1})$, and let $b = \min(s, |\xi|^{-1})$. We have

$$\left| \int_{r<|x|<b} k(x)e^{-2\pi i \xi \cdot x} \right| = \left| \int_{r<|x|<|\xi|^{-1}} k(x)[e^{-2\pi i \xi \cdot x} - 1] \right|$$

$$\leqslant \int_{r<|x|<b} |\xi||x|k(x) \, dx$$

$$\leqslant A|\xi| \int_{0<|x|<|\xi|^{-1}} |x|^{1-d}$$

$$\lesssim A|\xi||\xi|^{-1} \lesssim A.$$

For large $|x|$, we utilize the fact that the exponential is oscillating sufficiently rapidly, and we must somehow use the smoothness of k. If we had a strong derivative condition we could integrate by parts, but we cannot do this in general with the weaker conditions assumed. Instead, we use a trick, using the fact that the exponential is $\xi/2|\xi|^2$ periodic to write

$$\int_{a<|x|\leqslant s} k(x)e^{-2\pi i \xi \cdot x} \, dx = - \int_{a<|x|\leqslant s} k(x)e^{-2\pi i \xi \cdot (x+\xi/2|\xi|^2)} \, dx$$

$$= - \int_{a<|x-\xi/2|\xi|^2|\leqslant s} k(x - \xi/2|\xi|^2)e^{-2\pi i \xi \cdot x}.$$

The region $\{x : a < |x - \xi/2|\xi|^2| \leqslant s\}$ does not differ from the set $\{x : a < |x| \leqslant s\}$ by very much. Indeed, if the first condition holds, then $|x| \geqslant |\xi|^{-1}/2$, and so if a point is in the first set but not the second, then

$$0.5|\xi|^{-1} \leqslant |x| \leqslant |\xi|^{-1}.$$

On this set, $|k(x-\xi/2|\xi|^2)| \leqslant A|\xi|^{-d}$, so together with this bound the total integral of $k(x - \xi/2|\xi|^2)$ over this region is at most $O(A)$. Similarily, if a point is in the second set but not the first, then

$$|x - \xi/2|\xi|^2| \geqslant |\xi|^{-1}/2,$$

437

and a similar analysis gives the integral of $k(x - \xi/2|\xi|^2)$ is at most $O(A)$. Thus we conclude that

$$2\left|\int_{a<|x|\leqslant s} k(x)e^{-2\pi i\xi\cdot x}\,dx\right|$$
$$\leqslant \left|\int_{a<|x|\leqslant s}(k(x) - k(x - \xi/2|\xi|^2))e^{-2\pi i\xi\cdot x}\right| + O(A).$$

But now the triangle inequality implies that

$$\left|\int_{a<|x|\leqslant s}(k(x) - k(x - \xi/2|\xi|^2))e^{-2\pi i\xi\cdot x}\right| \lesssim A,$$

which completes the proof. $\qquad\square$

If $P(D)$ is a homogeneous differential operator of order k such that the equation $P(\xi) = 0$ has no non-zero real roots, then $(-\Delta)^{-k/2} \circ P(D)$ will be an invertible operator, and it's inverse will be a Calderon-Zygmund operator of convolution type. More generally, if $P(D)$ is an elliptic differential operator, then for some $M > 0$, $P(\xi) = 0$ has no real solutions ξ with $|\xi| \geqslant M$, for some fixed $M > 0$. If $\chi \in C^\infty(\mathbb{R}^d)$ has support on $|\xi| \geqslant M$, then one can verify that the inverse Fourier transform of $m(\xi) = \chi(\xi)|\xi|^k/P(\xi)$ is a strong Calderon-Zygmund kernel k of convolution type. Convolution by k corresponds to a *parametrix* operator for the equation $P(D)f = g$, i.e. an inverse for the differential operator, modulo smoothing terms.

It is important to note that we do not even need to exploit the fact that the Calderon-Zygmund operators are given by convolution in order to obtain results, which is very helpful when applying these techniques to non Euclidean settings. We just need an *integrability* and *cancellation condition*. This can be done on any locally compact space X equipped with a quasi-distance function ρ, i.e. a function $\rho : X \to X \to [0, \infty)$ such that:

- $\rho(x, y) = 0$ holds if and only if $x = y$.

- $\rho(x, y) \sim \rho(y, x)$.

- $\rho(x, y) \lesssim \rho(x, z) + \rho(z, y)$.

Given ρ, we can define $B(x, \delta) = \{y \in X : \rho(x, y) < \delta\}$, and we make the further assumption that X is equipped with a Radon measure satisfying $|B(x, c\delta)| \lesssim_c$

$|B(x, \delta)|$. These are similar assumptions to those occuring in the analysis of the Hardy-Littlewood maximal operator in more general settings.

The operators we will consider will be an operator T defined formally by the equation

$$Tf(x) = \int K(x, y) f(y) \, dy,$$

where K is some measurable function defined away from the diagonal. More precisely, we make the apriori assumptions that T is bounded from $L^q(X)$ to itself, for some $q > 1$, and that the following properties hold

- (Integrability Conditions): For each $f \in L^q(X)$ with compact support, for almost every $x \notin \text{supp}(f)$, we have

$$Tf(x) = \int K(x, y) f(y) \, dy,$$

  where the right hand side is assumed absolutely integrable in y, and thus can be interpreted in a Lebesgue sense.

- (Cancellation Condition): There exists a constants $A > 0$ and $c > 1$ such that for any $\delta > 0$, any $y \in X$, and any $y' \in X$ with $\rho(y, y') < \delta$,

$$\int_{B(y, c\delta)^c} |K(x, y) - K(x, y')| \, dx \leqslant A.$$

The first condition ensures that the Schwartz kernel of T is well behaved away from the diagonal, and the second condition ensures that enough cancellation occurs near the diagonal for us to obtain useful results.

On $\mathbb{R}^d$, equipped with the standard metric balls, such assumptions are satisfied if K is a *singular kernel*, i.e. satisfying bounds of the form

$$|K(x, y)| \lesssim \frac{1}{|x - y|^d},$$

and Hölder continuity estimates of the form

$$|K(x, y) - K(x, y')| \lesssim \frac{|y - y'|^\sigma}{|x - y|^{d+\sigma}} \quad \text{if} \quad |x - y| \geqslant c|y - y'|,$$

where $0 < \sigma \leqslant 1$ is some fixed value. The first assumption ensures the integrability condition holds, and the Hölder continuity estimate ensures the cancellation

condition holds. The Hölder continuity is most easily checked by obtaining point-wise derivative estimates of the form

$$|\nabla_y K(x,y)| \lesssim \frac{1}{|x-y|^{d+1}},$$

for then the mean value theorem implies the Hölder continuity conditions with $\sigma = 1$. An operator associated with a kernel of this type is called a *strong* singular kernel.

In the more abstract setting of a quasi-distance space X, we can also identify a pointwise condition that implies the required integrability and cancellation conditions. Define the function

$$V(x,y) = \inf\{|B(y,\delta)| : x \in B(y,\delta)\}.$$

Then $V(x,y) \sim B(y,\rho(x,y)) \sim B(x,\rho(x,y))$. The extension of the bound

$$|K(x,y)| \lesssim \frac{1}{|x-y|^d}$$

on $\mathbb{R}^d$ thus becomes that on X,

$$|K(x,y)| \lesssim \frac{1}{V(x,y)}.$$

The Hölder continuity condition can be generalized as follows: consider a *Dini modulus of continuity*, i.e. a non-decreasing function η with $\eta(0) = 0$ and

$$\int_0^1 \eta(s)\frac{ds}{s} < \infty.$$

The condition becomes that for $\rho(x,y) \geq c\rho(y,y')$,

$$|K(x,y) - K(x,y')| \lesssim \eta\left(\frac{\rho(y,y')}{\rho(x,y)}\right) \frac{1}{V(x,y)}.$$

Let us now show that this condition implies the required cancellation condition, in which case we say that K is a *strong singular kernel*.

Lemma 23.4. *If*

$$|K(x,y) - K(x,y')| \lesssim \eta\left(\frac{\rho(y,y')}{\rho(x,y)}\right) \frac{1}{V(x,y)}.$$

then

$$\int_{B(y,c\delta)^c} |K(x,y) - K(x,y')| \lesssim 1.$$

440

Proof. We write

$$\int_{B(y,c\delta)^c} |K(x,y) - K(x,y')| = \sum_{k=0}^{\infty} \int_{B(y,2^{k+1}c\delta) - B(y,2^kc\delta)} |K(x,y) - K(x,y')|\, dx.$$

If $x \notin B(y, 2^kc\delta)$, then $\rho(x,y) \geq 2^kc\delta$, so that we know

$$|K(x,y) - K(x,y')| \lesssim \eta\left(c^{-1}2^{-k}\right) \frac{1}{V(x,y)} \lesssim \eta\left(c^{-1}2^{-k}\right) \frac{1}{|B_y(2^kc\delta)|}.$$

Thus we conclude that

$$\int_{B(y,2^{k+1}c\delta) - B(y,2^kc\delta)} |K(x,y) - K(x,y')| \lesssim |B(y,2^{k+1}c\delta)|\eta\left(c^{-1}2^{-k}\right) \frac{1}{|B(y,2^kc\delta)|}$$

$$\lesssim \eta(c^{-1}2^{-k}).$$

Summing over k gives that

$$\int_{B(y,c\delta)^c} |K(x,y) - K(x,y')| \lesssim \sum_k \eta(c^{-1}2^{-k}) \sim \int_0^1 \eta(s)\frac{ds}{s} \lesssim 1. \qquad \square$$

Given a operator T defined on $L^q(X)$, we say a singular kernel K is associated with T if, for any compactly supported $f \in L^q(X)$, and any $x \notin \text{supp}(f)$,

$$Tf(x) = \int K(x,y)f(y)\, dy,$$

where the integrability assumptions on K ensure this is integrable. Then we say T is a *singular integral operator*. If T is bounded on $L^2(X)$, we furthermore say T is a *Calderon-Zygmund operator*. Unlike in the case of convolution kernels, determining when a singular integral operator is a Calderon-Zygmund operator can be a tricky question – it does not automatically follow from the structure above. And to apply the theory described here, we *must* have some a priori L^q boundedness to begin with in order to apply the Calderon-Zygmund theory.

If T is an operator associated with a singular kernel K, then the behaviour of T is determined *away* from the diagonal, but the singular kernel K does not determine the operator T. For instance, if $b \in L^\infty(\mathbb{R}^d)$, then the pointwise multiplication operator $Tf = bf$ is a Calderon-Zygmund operator with a singular kernel vanishing away from the diagonal. This is the only degeneracy of the relation of a kernel to a Calderon-Zygmund operators.

441

Theorem 23.5. *If T is a singular integral operator bounded on $L^q(X)$, associated with the singular kernel $K(x, y) = 0$. Then there exists $b \in L^\infty(X)$ such that*

$$Tf = bf$$

for all $f \in L^2(X)$.

Proof. The operator T induces a linear functional λ on $C_c(X)$ by defining

$$\lambda(f) = \int_X Tf(x) \, dx.$$

For any compact set K, and any $f \in C_c(X)$ supported on K, the function Tf is supported on K, and so

$$|\lambda(f)| \leq |K|^{1/q'} \|Tf\|_{L^q(X)} \lesssim |K|^{1/q'} \|f\|_{L^q(X)} \lesssim_K \|f\|_{L^\infty(X)}.$$

Thus λ is a continuous linear functional on $C_c(X)$. Thus there exists a signed Radon measure μ on X such that for $f \in C_c(X)$,

$$\int_X Tf(x) \, dx = \int f(x) d\mu(x).$$

The measure μ is absolutely continuous, since for a finite measure set $E \subset X$,

$$\mu(E) = \int_X T \mathbb{I}_E(x) \, dx = \langle T \mathbb{I}_E, \mathbb{I}_E \rangle \lesssim |E|.$$

Thus there exists $b \in L^1_{\text{loc}}(X)$ such that for any finite measure $E \subset X$,

$$\int f(x) d\mu(x) = \int f(x) b(x) \, dx.$$

For any simple function s, one uses the inequality above to show that

$$\left| \int s(x) b(x) \, dx \right| \lesssim \|s\|_{L^1(X)}.$$

But this is only possible if $b \in L^\infty(X)$. Now for any two simple functiosn s_1 and s_2, we check that

$$\langle T s_1, s_2 \rangle = \int b(x) s_1(x) s_2(x) \, dx,$$

and by continuity, this is only possible if $Tf = bf$ for all $f \in L^q(X)$. $\qquad \square$

Of course, in general, for operators not defined by convolution, it is nontrivial to show that operators given by singular kernels are bounded on $L^q(\mathbb{R}^d)$ for $q > 1$, which is necessary in order to apply the theory we now develop. But the advantage is that we can apply the theory to study more general partial differential equations, for instance, in the theory of parametrices for elliptic partial differential equations with *non-constant* coefficients, via the theory of *pseudodifferential operators*.

23.2   The Calderon-Zygmund Decomposition

Let us now describe the technique developed by Calderon and Zygmund to understand singular integrals. Let T be an operator with a singular kernel. The main trick of the theory of Calderon-Zygmund is the introduction of the *Calderon-Zygmund decomposition*. The method allows us to take in a function f with small L^1 norm, and write $f = g + b$, where g, the *good part* of the operator, has been 'upgraded' to have both small L^1 norm, *and* small L^∞ norm. If we assume T is bounded on $L^p(\mathbb{R}^d)$ for some p, then this will imply that Tg is well behaved. On the other hand, the function b, the *bad part* of the operator, cannot be expected to have good L^p properties other than having small L^1 norm, but the Calderon-Zygmund technique *can* guarantee that the function b is 'locally oscillatory'. Combined with the local cancellation properties of the singular integral T, this will enable us to get weak L^1 boundedness.

Let us now describe this decomposition more precisely. The trick here is to exploit the compatibility of measure and metric that has already been exploited, for instance, in our proof of Hardy-Littlewood maximal bounds. We begin with a version of the decomposition obtained by dyadic methods.

Theorem 23.6. *Let $f \in L^1(\mathbb{R}^d)$ and set $\lambda > 0$. Then there exists a set $\mathcal{Q}$ of disjoint, dyadic cubes, and a decomposition*

$$f = g + \sum_{Q \in \mathcal{Q}} b_Q,$$

such that

$$\|g\|_{L^1(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)} \quad and \quad \|g\|_{L^\infty(\mathbb{R}^d)} \lesssim_d \lambda,$$

and for $Q \in \mathcal{Q}$, $supp(b_Q) \subset Q$, $\int b_Q = 0$, and $\|b_Q\|_{L^1(\mathbb{R}^d)} \lesssim_d \lambda|Q|$. We have

$$\bigcup \mathcal{Q} = \{x : M_\Delta f(x) > \lambda\},$$

where M_Δ is the maximal averaging operator over dyadc cubes. From the weak L^1 bound for this operator we thus obtain that

$$\sum_{Q \in \mathcal{Q}} |Q| \leqslant \frac{\|f\|_{L^1(\mathbb{R}^d)}}{\lambda}.$$

Proof. Set $c = 2^d$. Call a dyadic cube Q *bad* if

$$\fint_Q |f(x)| \, dx > \lambda.$$

Since $f \in L^1(\mathbb{R}^d)$, any cube with sidelength exceeding $\|f\|_{L^1(\mathbb{R}^d)}/\lambda$ is good. Thus every bad cube is contained in a *maximal* bad cube, and the union of these bad cubes is precisely $\{x : M_\Delta f(x) > \lambda\}$. Let $\mathcal{Q}$ be the set of all maximal bad cubes. Define

$$g(x) = \sum_{Q \in \mathcal{Q}} \mathbb{I}(x \in Q) \fint_Q f(y) \, dy + \mathbb{I}(M_\Delta f(y) \leqslant \lambda) f(y).$$

By the Lebesgue density theorem, $|f| \leqslant M_\Delta f$ almost everywhere. Thus $|f(y)| \leqslant \lambda$ for almost every y with $M_\Delta f(y) \leqslant \lambda$. Conversely, if Q is a maximal bad cube, then the parent cube Q' of Q is good, which means that

$$\fint_Q |f(y)| \, dy \leqslant 2^d \fint_{Q'} |f(y)| \, dy \leqslant 2^d \lambda.$$

Thus we conclude that $\|g\|_{L^\infty(\mathbb{R}^d)} \lesssim_d \lambda$. It is also easy to see that $\|g\|_{L^1(\mathbb{R}^d)} \leqslant \|f\|_{L^1(\mathbb{R}^d)}$. On the other hand, if we set

$$b_Q(x) = \mathbb{I}(x \in Q)[f(x) - \fint_Q f(y) \, dy],$$

then $\int b_Q = 0$, $\operatorname{supp}(b_Q) \subset Q$, and $\int |b_Q(x)| \leqslant 2 \int_Q |f(x)| \, dx \leqslant 2^{d+1} \lambda$. $\qquad \square$

Remark. On a general measure space X, for $f \in L^1(X)$ one can write $f = f\,\mathbb{I}_{x \leqslant \lambda} + f\,\mathbb{I}_{x > \lambda} = g + b$. We have

$$\|g\|_{L^1(X)} \leqslant \|f\|_{L^1(X)} \quad \text{and} \quad \|g\|_{L^\infty(X)} \leqslant \lambda.$$

And for the bad function,

$$\|b\|_{L^1(X)} \leqslant \|f\|_{L^1(X)} \quad \text{and} \quad |\operatorname{supp}(b)| \leqslant \|f\|_{L^1(X)}/\lambda.$$

Thus the Calderon-Zygmund decomposition obtains a similar version of this result, but accounting for additional metric structure, which gives the 'local oscillatory' property of the decomposition.

Remark. For each point $x \in \mathbb{R}^d$ contained in a bad cube, let T_x be the integer n such that the maximal bad cube containing x has sidelength 2^{-l}. Otherwise, let $T_x = \infty$. Then in the language of conditional expectations, $g = \mathbb{E}[f|T_x]$. Thus one can view the Calderon-Zygmund decomposition as a kind of non-probabilistic *stopping time* result.

The key idea in the proof above was to first decompose the open set $\Omega = \{x : M_\Delta f(x) > \lambda\}$ into disjoint dyadic cubes, which is another very useful geometric result, the *Whitney decomposition*.

Theorem 23.7. *Let $\Omega \subset \mathbb{R}^d$ be an open set. For any $K \geqslant 1$, there exists a decomposition Ω into a disjoint union of dyadic cubes Q, where for any $Q \in Q$,*

$$diam(Q) \sim Kd(Q, \Omega^c).$$

Proof. Let Q' denote the dyadic cubes Q contained in Ω such that

$$K\text{diam}(Q) \leqslant d(Q, \Omega^c) \leqslant 4K\text{diam}(Q).$$

These cubes cover Ω. Indeed, for each $x \in \Omega$, one can find a dyadic cube Q with diameter between $d(x, \Omega^c)/4K$ and $d(x, \Omega^c)/2K$ containing x. Then

$$
\begin{aligned}
d(Q, \Omega^c) &\geqslant d(x, \Omega^c) - \text{diam}(Q) \\
&\geqslant (1 - 1/4K) \cdot d(x, \Omega^c) \\
&\geqslant (2K - 1/2) \cdot \text{diam}(Q) \\
&\geqslant K\text{diam}(Q).
\end{aligned}
$$

Conversely,

$$d(Q, \Omega^c) \leqslant d(x, \Omega^c) \leqslant 4K\text{diam}(Q).$$

But now we can take a maximal subfamily Q of Q'. $\qquad\square$

Remark. Consider the decomposition Q in the last theorem. If Q_1, Q_2 are two nearby cubes in the decomposition, and

$$d(Q_1, Q_2) \leqslant \delta(\text{diam}(Q_1) + \text{diam}(Q_2))$$

for some $\delta > 0$, then

$$
\begin{aligned}
|d(Q_1, \Omega^c) - d(Q_2, \Omega^c)| &\leqslant d(Q_1, Q_2) + \text{diam}(Q_1) + \text{diam}(Q_2) \\
&\leqslant (1 + \delta)(\text{diam}(Q_1) + \text{diam}(Q_2)).
\end{aligned}
$$

Thus

$$\text{diam}(Q_2) \leqslant \frac{d(Q_2, \Omega^c)}{K} \leqslant \frac{d(Q_1, \Omega^c)}{K} + \frac{1+\delta}{K}\text{diam}(Q_1) + \frac{1+\delta}{K}\text{diam}(Q_2).$$

Rearranging, we find that for $\delta \leqslant K/2 - 1$,

$$\frac{\text{diam}(Q_2)}{2} \leqslant \frac{d(Q_1, \Omega^c)}{K} + (1+\delta)\text{diam}(Q_1) \leqslant (5+\delta)\text{diam}(Q_1),$$

and so

$$\text{diam}(Q_2) \leqslant (5+\delta) \cdot \text{diam}(Q_1).$$

By symmetry, we automatically obtain the two sided inequality

$$(5+\delta)^{-1} \cdot \text{diam}(Q_1) \leqslant \text{diam}(Q_2) \leqslant (5+\delta) \cdot \text{diam}(Q_1).$$

Thus balls which are close together also have comparable diameter.

The Calderon-Zygmund decomposition and Whitney decomposition are in some sense equivalent. Suppose we have found a construction of the Whitney decomposition without using the Calderon-Zygmund decomposition. Given a function f, set

$$\Omega = \{x : M_\Delta f(y) > \lambda\},$$

and let $\mathcal{Q}$ denote the Whitney decomposition. If $Q \in \mathcal{Q}$, then $d(Q, \Omega^c) \lesssim \text{diam}(Q)$. Thus we can find a point x with $d(x, Q) \lesssim \text{diam}(Q)$ and with $M_\Delta f(x) \leqslant \lambda$. Thus for a universal constant $C > 0$, $x \in C \cdot Q$, and so

$$\fint_Q |f(y)| \, dy \leqslant C^d \fint_{CQ} |f(y)| \, dy \leqslant C^d \lambda.$$

Writing $b_Q(x) = \mathbb{I}(x \in Q) \cdot [f(x) - (\fint_Q f(y) \, dy)]$ and $g = f - \sum b_Q$, we obtain an analogous decomposition to the one obtained before.

We will obtain a more general Calderon-Zygmund decomposition using this technique, i.e. first proving a Whitney type decomposition result. Certainly one cannot prove an exact analogy of the Whitney decomposition for *balls*, i.e. one cannot cover a general open set by disjoint open balls with radius comparable to their distance to the boundary of the open set. Nonetheless, we can obtain such a decomposition with the *bounded intersection property*, i.e. such that no point is contained in too many of the balls in the particular cover.

446

Theorem 23.8. *Let* $\Omega \subset \mathbb{R}^d$ *be an open set. For any* $K \geqslant 1$, Ω *is the union of a family of balls* $\mathcal{B}$, *where for each ball* $B \in \mathcal{B}$, *the radius* r_B *of this ball is comparable to* $Kd(Q, \Omega^c)$, *and each point in* Ω *is contained in* $O_d(1)$ *balls.*

Proof. Applying the last remark, without loss of generality assume $K \geqslant 2$, and pick $\delta = 1$. Consider the cubes Q in the decomposition above, and let $\mathcal{B}$ be a family of balls obtained by covering each cube Q in $\mathcal{Q}$ by a ball B with $\text{diam}(B) = \text{diam}(Q)$. If a point x is contained in a family of balls $B_1, \ldots, B_k$ corresponding to dyadic cubes $Q_1, \ldots, Q_k$, then $d(Q_i, Q_j) \leqslant \text{diam}(Q_i) + \text{diam}(Q_j)$. Thus we find that for any Q_i and Q_j

$$\text{diam}(Q_i) \leqslant 50 \cdot \text{diam}(Q_j).$$

Thus there is a common quantity integer n such that for all i, $2^n \leqslant \text{diam}(Q_i) \leqslant 50^2 \cdot 2^n$. The number of dyadic lattice points whose vertices have coordinates which can be written as a fraction with denominator 2^n, and which lie a distance at most $50^2 \cdot 2^n$ from x is $O_d(1)$, which shows $k \lesssim_d 1$, and thus we obtain the required bounded intersection property. $\square$

This decomposition using balls is the one most generalizable to more general situations where one is working in a space without quite as regular a tiling as the family of cubes. To obtain this generalization, we assume the slightly stronger assumptions we made in our analysis of the Hardy-Littlewood maximal function, i.e. we assume we are working on a locally compact topological space X equipped with a locally finite Radon measure such that the technical assumptions made there, and the slightly stronger *engulfing* and *doubling* conditions mentioned, hold true. To generalize the Whitney decompositon, we fix two additional constants c^* and c^{**} with $1 < c^* < c^{**}$ which will depend only on c_1 and c_2, and for a ball $B = B(x, \delta)$, we let $B^* = B(x, c^* \delta)$ and $B^{**} = B(x, c^{**} \delta)$.

Theorem 23.9. *Let* $\Omega \subset X$ *be an open set. Then there is a collection of pairwise disjoint balls* $\mathcal{B}$, *such that*

$$\bigcup_{B \in \mathcal{B}} B^* = \Omega, \quad \text{and} \quad B^{**} \cap \Omega^c \neq \varnothing \text{ for any } B \in \mathcal{B}.$$

It is simple to construct from these balls a disjoint family of sets Q_B *for each* $B \in \mathcal{B}$ *such that* $B \subset Q_B \subset B^*$, *and* $\bigcup Q_B = \Omega$, *which can be taken as a substitute for cubes in the original Whitney covering lemma.*

447

Remark. As with the Euclidean case above, the balls $\{B^* : B \in \mathcal{B}\}$, though not disjoint, have the bounded intersection property. This follows from a similar argument to that of the above argument. First we show intersecting balls have comparable radii. Suppose $B_1, B_2 \in \mathcal{B}$ and $B_1 \cap B_2 \neq \emptyset$, where $B_i = B(x_i, r_i)$, and $r_1 \leqslant r_2$. Then

$$B_1^{**} = B(x_1, c^{**} r_1) \subset B(x_2, c_2 c^{**} r_1).$$

Since $B_2^* = B(x_2, c^* r_2)$ is disjoint from Ω^c, whereas B_1^{**} is not disjoint from Ω^c it follows that $c^* r_2 \leqslant c_2 c^{**} r_1$, so that $r_2 \lesssim r_1$, and by symmetry, $r_1 \sim r_2$. Now if $B_1^*, \ldots, B_N^*$ are balls whose intersection contains a point x_0, where $B_i = B(x_i, r_i)$. Then the balls all have comparable radii, i.e. if $r = \max(r_1, \ldots, r_N)$ then $r \sim r_i$. Set $B = B(x_0, r)$. The doubling / engulfing property implies that $|B_i| \sim |B|$ for all i. But also $B_i \subset B(x_0, c_1 r)$ for all i, and since the $\{B_i\}$ are disjoint, we find that

$$N|B| \lesssim |B_1| + \cdots + |B_N| \leqslant |B(z, c_1 r)| \lesssim |B|.$$

Thus we find $N \lesssim 1$, i.e. so we have the bounded intersection property.

Proof. TODO: See Big Stein. $\qquad\square$

A consequence is a generalized Calderon-Zygmund decomposition.

Theorem 23.10. *Let $f \in L^1(X)$, and $\alpha > 0$, where*

$$\alpha > \frac{\|f\|_{L^1(X)}}{|X|}.$$

Then we can write $f = g + b$, where b has a decomposition as $\sum b_k$ associated with a family of balls $\{B_k\}$ such that:

- $\|g\|_{L^\infty(X)} \lesssim \alpha$.

- *The functions $\{b_k\}$ have disjoint support, b_k is supported on B_k, $\|b_k\|_{L^1(X)} \lesssim \alpha |B_k|$, and*

$$\int b_k(x) \, dx = 0.$$

- $\sum |B_k| \lesssim \alpha^{-1} \|f\|_{L^1(X)}$.

- *We can find disjoint sets $\{Q_k\}$ such that $Q_k \subset B_k$, b_k is supported on Q_k*

Proof. TODO: See Big Stein. $\qquad\square$

23.3 Calderon-Zygmund Kernels

We now prove our main result using the Calderon-Zygmund decomposition; given an operator T bounded in $L^q(X)$, and with a kernel K satisfying the integrability and cancellation conditions above, we will show that T is bounded on $L^p(X)$ for all $1 < p < \infty$. Interpolation shows it suffices to show T is weakly bounded on $L^1(X)$.

Theorem 23.11. *The operator T satisfies an estimate of the form*

$$\|Tf\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$$

for $f \in L^1(X) \cap L^q(X)$, so that T extends uniquely to an operator from $L^1(X)$ to $L^{1,\infty}(X)$, and by interpolation, from $L^p(X)$ to $L^p(X)$ for all $1 < p < q$.

Proof. Our goal is to bound the measure of the set

$$\{x : |Tf(x)| > \alpha\}.$$

It suffices to obtain the bound (by a density argument) for $f \in L^1(X) \cap L^q(X)$. Perform a Calderon-Zygmund decomposition at the scale α, writing $f = g + \sum b_k$, associated with a family of balls $\{B_k\}$. It suffices to obtain a bound on the measure of the sets

$$\{x : |Tg(x)| > \alpha/2\} \quad \text{and} \quad \{x : |Tb(x)| > \alpha/2\}.$$

First, we argue that $g \in L^q(X)$, and more precisely, that

$$\|g\|_{L^q(X)} \lesssim \alpha^{1-1/q} \|f\|_{L^1(X)}.$$

On $A = (\bigcup B_k)^c$, the function g arees with f, so we have $\|g\|_{L^\infty(A)} \lesssim \alpha \|f\|_{L^1(X)}$ and $\|g\|_{L^1(A)} \lesssim \|f\|_{L^1(A)}$, which we can interpolate to conclude that $\|g\|_{L^q(A)} \lesssim \alpha^{1-1/q} \|f\|_{L^1(X)}$. On the other hand, $|A^c| \lesssim \alpha^{-1} \|f\|_{L^1(X)}$, so combined with the bound $\|g\|_{L^\infty(A^c)} \lesssim \alpha \|f\|_{L^1(X)}$ we can again interpolate to yield $\|g\|_{L^q(A^c)} \lesssim \alpha^{1-1/q} \|f\|_{L^1(X)}$, and combining this with the previous bound completes the claim. Thus we have $\|Tg\|_{L^q(X)} \lesssim \alpha^{1-1/q} \|f\|_{L^1(X)}$, which conveniently yields via Chebyshev's inequality that the measure of the set

$$\{x : |Tg(x)| > \alpha/2\}$$

has measure at most $O(\alpha^{-1} \|f\|_{L^1(X)})$, which is precisely the bound we need.

It remains to analyze the set $\{x : |Tb(x)| > \alpha/2\}$. Each of the functions b_k differs from f on it's support Q_k by the average value of f on Q_k, and so these functions all lie in $L^q(X)$. If x_0 denotes the center of the ball B_k, then

$$\int_{(B_k^*)^c} |Tb_k(x)|\ dx = \int_{(B_k^*)^c} \left|\int K(x, x')b_k(x')\ dx'\right|\ dx$$

$$= \int_{(B_k^*)^c} \left|\int [K(x, x') - K(x, x_0)]b_k(x')\ dx'\right|\ dx$$

$$\leqslant \int_{(B_k^*)^c} \int |K(x, x') - K(x, x_0)||b_k(x')|\ dx'\ dx$$

$$\leqslant \left(\sup_{x' \in B_k} \int_{(B_k^*)^c} |K(x, x') - K(x, x_0)|\ dx\right) \|b_k\|_{L^1(X)}$$

$$\lesssim A\alpha|B_k|.$$

In particular, this means that

$$\|Tb\|_{L^1((\bigcup_k B_k^*)^c)} \lesssim \sum_k A\alpha|B_k| \lesssim A\|f\|_{L^1(X)}.$$

On the other hand, $\sum |B_k^*| \lesssim \alpha^{-1}\|f\|_{L^1(X)}$, so we conclude that by Chebyshev,

$$\{x : |Tb(x)| > \alpha/2\} \lesssim \alpha^{-1}A\|f\|_{L^1(X)} + \sum|B_k^*| \lesssim \alpha^{-1}\|f\|_{L^1(X)}.$$

Putting together all these bounds completes the proof. □

Remark. Let T be as above. If the kernel of T^* satisfies the integrability and cancellation conditions, then by duality, it is bounded on $L^{q^*}(\mathbb{R}^d)$. But the theorem above then implies that T^* is bounded on $L^r(\mathbb{R}^d)$ for $1 < r \leqslant q^*$. But this means that T is bounded on $L^p(\mathbb{R}^d)$ for $p \geqslant q$. Combined with the bounds in the previous theorem, we conclude that T is bounded on $L^p(\mathbb{R}^d)$ for all $1 < p < \infty$.

23.4   Truncation of Singular Integral Operators

Given a singular integral operator T on a measure space X satisfying the doubling and engulfing conditions with has considered, it is natural to consider the family of *truncated* operators $\{T_\varepsilon\}$, where T_ε is the operator with kernel

$$K_\varepsilon(x, y) = \begin{cases} K(x, y) & : \rho(x, y) \geqslant \varepsilon, \\ 0 & : \rho(x, y) < \varepsilon \end{cases}.$$

Our goal is to determine if estimates of the form

$$\|T_\varepsilon f\|_{L^q(X)} \lesssim \|f\|_{L^q(X)},$$

hold uniformly in ε. This is an intuitively stronger result than bounds on the singular integral, since under mild regularity assumptions, one has

$$\|Tf\|_{L^q(X)} = \lim_{\varepsilon \to 0} \|T_\varepsilon f\|_{L^q(X)}.$$

We can obtain such a result if we assume that K is a strong singular kernel.

Theorem 23.12. *Suppose T is an operator with a singular kernel K satisfying*

$$K(x,y) \lesssim \frac{1}{V(x,y)}.$$

If T is bounded on $L^q(X)$, then we have a bound

$$\|T_\varepsilon f\|_{L^q(X)} \lesssim \|f\|_{L^q(X)}$$

uniformly in $\varepsilon > 0$.

Proof. Since $|K(x,y)| \lesssim V(x,y)^{-1}$, for a fixed x, we have

$$\{y \in X : K(x,y) \geqslant \alpha\} \subset \{y \in X : V(x,y) \leqslant 1/\alpha\}$$
$$\subset \bigcup \{B(x,\delta) : y \in B(x,\delta) \text{ and } |B(x,\delta)| < 1/\alpha\}.$$

As δ increases, $B(x,\delta)$ increases, so by monotonicity, the measure of this set is at most $1/\alpha$. Thus

$$\|K(x,\cdot)\|_{L^{1,\infty}_y} \lesssim 1.$$

Moreover, if $\rho(x,y) \geqslant \varepsilon$, then $V(x,y) \gtrsim |B(x,\varepsilon)|$, so that

$$\|K_\varepsilon(x,\cdot)\|_{L^\infty_y} \lesssim_{x,\varepsilon} 1.$$

But this means $\|K_\varepsilon(x,\cdot)\|_{L^p_y} \lesssim_{x,\varepsilon} 1$ for all $1 < p \leqslant \infty$, and so by Hölder's inequality, we know that the formula

$$(T_\varepsilon f)(x) = \int K_\varepsilon(x,y) f(y) \, dy$$

is well defined for all $f \in L^q(X)$.

One useful trick is that to obtain that the operators T_ε are uniformly, it suffices to obtain uniform L^q estimates on the operators $S_\varepsilon = T - T_\varepsilon$. The advantage of switching to $\{S_\varepsilon\}$ is that they are *localized*, i.e. the support of $S_\varepsilon f$ is contained in a ε neighborhood of the support of f. If we let χ_δ be the indicator function of $B(x_0, \delta)$, for some fixed x_0, then we will show that

$$\|\chi_{a\varepsilon} S_\varepsilon f\|_{L^q(X)} \lesssim \|\chi_{b\varepsilon} f\|_{L^q(X)},$$

for some a and b, with $a < b$. This suffices to complete the proof. Inded, given any $\varepsilon > 0$, we can find a maximal disjoint collection of balls $\{B_i\}$ of radius $a\varepsilon/c$ in X, where $c > 1$ is the engulfing constant. This fact implies that the corresponding balls $\{B_i^*\}$ of radius $a\varepsilon$ cover X. Moreover, the balls $\{B_i^{**}\}$ of radius $b\varepsilon$ have the bounded overlap property. Indeed, if N is suitably large, and $x \in X$ lies in N balls $B_1^{**}, \ldots, B_N^{**}$ in the cover, then by the engulfing property, $B(x, bc\varepsilon)$ contains $B_1, \ldots, B_N$. Each of the balls $B_1, \ldots, B_N$ is close enough together that, by the doubling property, they all have comparable volume. Thus

$$|B(x, bc\varepsilon)| \geqslant \sum_{i=1}^{N} |B_i| \gtrsim N|B(x, bc\varepsilon)|,$$

which gives a contradiction for suitably large N. Thus for $f \in L^q(X)$, we have

$$\begin{aligned}
\|S_\varepsilon f\|_{L^q(X)} &= \left(\sum \|\chi_{a\varepsilon} S_\varepsilon f\|_{L^q(X)}^q \right)^{1/q} \\
&\lesssim \left(\|\chi_{b\varepsilon} f\|_{L^q(X)}^q \right)^{1/q} \\
&\lesssim \|f\|_{L^q(X)},
\end{aligned}$$

and so we obtain the required bound.

So let's establish the bound

$$\|\chi_{a\varepsilon} S_\varepsilon f\|_{L^q(X)} \lesssim \|\chi_{b\varepsilon} f\|_{L^q(X)},$$

If $b \gtrsim a + 1$, then we have

$$\chi_{a\varepsilon} S_\varepsilon f = \chi_{a\varepsilon} S_\varepsilon \chi_{b\varepsilon} f.$$

Next, if $d < b$ we write

$$\chi_{a\varepsilon} S_\varepsilon \chi_{b\varepsilon} f = \chi_{a\varepsilon} S_\varepsilon \chi_{d\varepsilon} f + \chi_{a\varepsilon} S_\varepsilon (\chi_{b\varepsilon} - \chi_{d\varepsilon}) f.$$

452

Let us assume that $a + d$ is small enough so that if $\rho(x, x_0) \leq a\varepsilon$ and $\rho(y, x_0) \leq d\varepsilon$, then $\rho(x, y) \leq \varepsilon$. This implies that

$$\chi_{a\varepsilon} S_\varepsilon \chi_{d\varepsilon} f = \chi_{a\varepsilon} S \chi_{d\varepsilon} f.$$

Thus

$$\|\chi_{a\varepsilon} S_\varepsilon \chi_{d\varepsilon} f\|_{L^q(X)} \lesssim \|\chi_{d\varepsilon} f\|_{L^q(X)} \lesssim \|\chi_{b\varepsilon} f\|_{L^q(X)}.$$

It thus suffices to analyze

$$U_\varepsilon f = \chi_{a\varepsilon} S_\varepsilon (\chi_{b\varepsilon} - \chi_{d\varepsilon}) f.$$

For $x \in \chi_{a\varepsilon}$, we have

$$U_\varepsilon f(x) = \int_{B(x,\varepsilon) \cap (B(x_0,b\varepsilon) - B(x_0,d\varepsilon))} K(x, y) f(y) \, dy.$$

If $y \in B(x, \varepsilon) \cap (B(x_0, b\varepsilon) - B(x_0, d\varepsilon))$, then

$$\rho(x, y) + \rho(x_0, x) \gtrsim \rho(x_0, y) \geq d\varepsilon.$$

Since $\rho(x_0, x) \leq a\varepsilon$, we conclude that there is $c > 0$ such that

$$\rho(x, y) \geq (cd - a)\varepsilon.$$

If $a < cd/2$, then $\rho(x, y) \geq (cd/2)\varepsilon$. Since $\rho(x_0, y) \lesssim \varepsilon$, we conclude that $\rho(x, y) \gtrsim \rho(x_0, y)$. But this means that $V(x, y) \gtrsim V(x_0, y)$, and so

$$|K(x, y)| \lesssim \frac{1}{V(x_0, y)} \lesssim \frac{1}{|B(x_0, d\varepsilon)|} \lesssim \frac{1}{|B(x_0, \varepsilon)|}.$$

But now Schur's Lemma type arguments imply that

$$\|U_\varepsilon f\|_{L^q(X)} \lesssim \|f\|_{L^q(X)}. \qquad \square$$

We can then also gain uniform weak L^1 estimates for the operators $\{T_\varepsilon\}$.

Theorem 23.13. *Suppose T is a Calderon-Zygmund operator with a singular kernel K satisfying*

$$K(x, y) \lesssim \frac{1}{V(x, y)},$$

and is bounded on $L^q(X)$. Then we have a bound of the form

$$\|T_\varepsilon f\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^q(X)}$$

uniformly in $\varepsilon > 0$.

Proof. We have already proved that T_ε is bounded in the L^q norm, uniformly in ε. To obtain the theorem. It thus suffices to show $\{T_\varepsilon\}$ satisfies the Calderon-Zygmund assumptions, uniformly in $\varepsilon > 0$, which is left as an exercise to the reader. $\square$

Since the operators $\{T_\varepsilon\}$ are uniformly bounded, then there exists a sequence $\{\varepsilon_j\}$ such that the operators $\{T_{\varepsilon_j}\}$ converge weakly to some operator T_0. Then $T - T_0$ is a singular integral operator with singular kernel 0, bounded on $L^q(X)$, and so we know there exists $b \in L^\infty(X)$ such that $(T - T_0)f = bf$ for all $f \in L^q(X)$. The full family $\{T_\varepsilon\}$ may not converge weakly, e.g. by considering a fractional integral operator of imaginary order.

Example. *Let T be the Fourier multiplier on $\mathbb{R}^d$ with symbol $|\xi|^{-it}$. Then T is given by convolution against a constant multiple of the function*

$$k(x) = |x|^{-d+it}.$$

T is bounded on $L^2(\mathbb{R}^d)$ because it's symbol is bounded. Moreover, we have $|k(x)| = 1/|x|^d$, and $|\nabla k(x)| \lesssim 1/|x|^{d+1}$, so k is a strong singular kernel. But it is not true for all functions f that $T_\varepsilon f$ converges as $\varepsilon \to 0$. Indeed, if $f \in C_c^\infty(\mathbb{R}^d)$, and for a fixed y, we let $S f(x, r)$ denote the spherical average of f on the sphere of radius r centered at x, then

$$
\begin{aligned}
\int_{\varepsilon_1 \leqslant |y| \leqslant \varepsilon_2} f(x - y)|y|^{-d+it} &= \int_{\varepsilon_1}^{\varepsilon_2} r^{it-1}(S f)(x, r) \, dr \\
&= \frac{\varepsilon_2^{it}(S f)(x, \varepsilon_2) - \varepsilon_1^{it}(S f)(x, \varepsilon_1)}{it} \\
&\quad - \frac{1}{it} \int_{\varepsilon_1}^{\varepsilon_2} r^{it}(S' f)(x, r) \, dr \\
&\approx \frac{\varepsilon_2^{it} - \varepsilon_1^{it}}{it} f(x) + O(\varepsilon_2).
\end{aligned}
$$

The oscillation here as ε_1 and ε_2 vary close to the origin imply that the operators $\{T_\varepsilon\}$ cannot converge. Nonetheless, the maximal operator T^* we now consider is still bounded for these truncations.

To determine more deeply the notion of whether $\{T_\varepsilon\}$ converges to T as $\varepsilon \to 0$, or at least, converges to an operator differing from T by a pointwise multiplication

operator, and especially when this convergence happens pointwise almost everywhere, it is natural to introduce the maximal operator

$$T^* f = \sup_{\varepsilon} |T_\varepsilon f|.$$

We will obtain this given that K is a *strong* singular kernel, i.e. satisfying a pointwise cancellation estimate of the form

$$|K(x,y) - K(x',y)| \leqslant \eta \left(\frac{\rho(x,x')}{\rho(x',y)} \right) \frac{1}{V(x',y)}$$

whenever $\rho(x',y) \gtrsim \rho(x,x')$.

Theorem 23.14. *Suppose T is a singular integral operator associated with a strong singular kernel $K(x,y)$, such that $K(y,x)$ is also a strong singular kernel. If T is bounded on $L^q(X)$, then*

$$T^* f \lesssim M(Tf) + Mf$$

and more generally, for any $r > 0$,

$$T^* f \lesssim_r M(|Tf|^r)^{1/r} + Mf.$$

Thus T^* is bounded on $L^p(X)$ for $1 < p \leqslant q$, and bounded on $L^{1,\infty}(X)$.

Proof. TODO: See Stein, End of Chapter 1. □

We note that under these assumptions, T and T_* are actually bounded on all L^p spaces, by using duality and noting that the adjoint of T is also a singular integral of the same type as T itself.

23.5   Calderon-Zygmund Operators on BMO

Let us review the main results we have obtained. We are given an operator T, with some a priori L^q boundedness, and with a kernel satisfying suitable cancellation conditions. The L^q boundedness does not automatically imply L^1 boundedness. This is because the mass of a given input could be spread out over a much larger region, in a way aligning with the L^q bounds, but such that the L^1 norm of the output is much larger than the L^1 norm of the input.

Remark. One can see heuristically why L^1 boundedness is equivalent to showing mass cannot be spread, by pretending general inputs are quasi-step functions, if a quasi-step function f with height H_0 and width W_0 is mapped under T to a quasi-step function g with height H_1 and width W_1, then the only way for us to have $\|f\|_{L^1(X)} \gg \|g\|_{L^1(X)}$ but still have $\|f\|_{L^q(X)} \lesssim \|g\|_{L^q(X)}$ is if $H_1 W_1 \gg H_0 W_0$ but $H_1 W_1^{1/q} \lesssim H_0 W_0^{1/q}$ is if

$$H_0 W_0 \ll H_1 W_1 \lesssim W_1^{1-1/q} H_0 W_0^{1/q},$$

which we can rearrange to read $W_0 \ll W_1$, i.e. the support of the quasi step function is vastly expanded.

We were able to show that the support of a function is not vastly expanded by using the Calderon-Zygmund decomposition, since, using the cancellation properties of a given kernel, we saw that if b is a function with mean zero, supported on a ball B, then we have bounds of the form

$$\|Tb\|_{L^1(X-B^*)} \lesssim \|b\|_{L^1(X)}.$$

This is sufficient to show that the operator does not significantly expand the support of the operator T. We were therefore able to use this to obtain weak L^1 bounds for the operator.

Now suppose we are given the 'dual' Calderon-Zygmund conditions, i.e. we are given an operator T, bounded on $L^q(X)$, and a measurable function K satisfying:

- (Integrability conditions): For each $f \in L^q(X)$ with compact support, and for almost every $x \notin \mathrm{supp}(f)$, we have

$$Tf(x) = \int K(x, y) f(y) \, dy,$$

  where the right hand side is assumed absolutely integrable in y, and thus can be interpreted in the Lebesgue sense.

- (Cancellation Condition $*$): There exists constants $A > 0$ and $c > 1$ such that for any $\delta > 0$, any $x \in X$, and any $x' \in X$ with $\rho(x, x') < \delta$,

$$\int_{B(y, c\delta)^c} |K(x, y) - K(x', y)| \, dy \leqslant A.$$

456

We might expect that T is then bounded on $L^p(X)$ for all $q < p < \infty$, and, assuming the usual integrability conditions one can just obtain this by duality, since the adjoint of T will satisfy the usual conditions. We want to work instead by trying to obtain 'weak L^∞ bounds' that we can interpolate. This is different from the usual weak norms; indeed, $L^{\infty,r} = L^\infty$ for all $0 < r \leqslant \infty$. A substitute is provided by the *BMO Norm*, or *Bounded mean oscillation norm*. One quirk is that it is only a seminorm: it is only defined modulo constants.

For a locally-integrable function $f : X \to \mathbb{C}$, we define

$$\|f\|_{\text{BMO}(X)} = \sup_B \inf_c \fint_B |f(x) - c| \, dx.$$

For any ball B, one can remove the infinum over c by choosing c_B to be the average value of f over B. We let $BMO(\mathbb{R}^d)$ be the Banach space of equivalence classes of functions with finite BMO norm, where we identify two functions f_0 and f_1 if $f_0 - f_1$ agrees with a constant function almost everywhere. We have a bound $\|f\|_{\text{BMO}(X)} \leqslant \|f\|_{L^\infty(X)}$, so every function in $L^\infty(\mathbb{R}^d)$ lies in $BMO(\mathbb{R}^d)$, but $BMO(\mathbb{R}^d)$ also contains unbounded functions, like $\log|x|$ (TODO Prove this).

We can now extend the definition of Tf to a general element of $L^\infty(\mathbb{R}^d)$, such that the output is defined as an element of the space V, which is the space $L^1_{\text{loc}}(X)$ of locally integrable functions, where we have quotiented out the constant functions. By duality, it suffices to define how Tf integrates against a general compactly supported, bounded function g with mean zero. If g is supported on the interior of a ball B, then we define

$$\langle Tf, g \rangle = \int \int_B K(x,y)f(y) \, dyg(x) \, dx + \int_{(B*)^c} [K(x,y) - K(c_x,y')]g(x)f(y).$$

The first quantity is well defined, and a continuous functional of g, since $f\mathbf{I}_B$ lies in $L^q(\mathbb{R}^d)$, and so $T(f\mathbf{I}_B)$ lies in $L^q(\mathbb{R}^d)$. The second quantity is well-defined and a continuous functional of g because of the cancellation properties, i.e.

$$\left| \int_{(B*)^c} [K(x,y) - K(c_x,y')]g(x)f(y) \right| \leqslant A\|f\|_{L^\infty}\|g\|_{L^\infty}.$$

The fact that this is invariant of the choice of ball B follows because g has mean zero. Thus Tf is well-defined as a locally integrable function. But actually, since our implicit constants are independent of B, we have actually shown that T is a continuous map from $L^\infty(X)$ into BMO(X).

Theorem 23.15. *We have* $\|Tf\|_{BMO(X)} \lesssim \|f\|_{L^\infty(X)}$.

In order to exploit the BMO norm, we must obtain tools which show such bounds are 'almost' as good as L^∞ bounds. One result comes from iterating Markov's inequality applied locally on balls at various scales, giving us the John-Nirenberg inequality.

Theorem 23.16 ((John-Nirenberg Inequality)). *For any X, there exists c such that if $f \in BMO(X)$. Then for any ball B, if $c_B = \fint_B f$, then we have*

$$\left|\{x \in B : |f(x) - c_B| \geqslant \lambda\|f\|_{BMO(X)}\}\right| \lesssim e^{-c\lambda}|B|.$$

Proof. Assume without loss of generality that $\|f\|_{BMO(X)} = 1$. Our goal is to iterate at various scales. Let $A(\lambda)$ be the best constant for which we have

$$\left|\{x \in B : f(x) \geqslant c_B + \lambda\}\right| \leqslant A(\lambda)|B|.$$

Markov's inequality, applied to the definition of the BMO norm, gives $A(\lambda) \leqslant 1/\lambda$, and a trivial estimate gives $A(\lambda) \leqslant 1$. We now use a variant of the Calderon-Zygmund decomposition. $\qquad\square$

23.6   Vector-Valued Operators

We can generalize pretty much all of what we have said to operators which act on *vector-valued* functions rather than *real-valued* functions. In other words, for a Banach space B, we consider the space $L^p(X, B)$ consisting of all functions $f : X \to B$ such that

$$\|f\|_{L^p(X,B)} = \left(\int_X \|f(x)\|_B^p \, dx\right)^{1/p}$$

is finite. We then fix two Banach spaces B_1 and B_2, and consider a kernel

$$K : X \times X \times B_1 \to B_2,$$

where we write the output of (x, y, a) as $K(x, y)a$. We consider an operator T, bounded from $L^q(X, B_1)$ to $L^q(X, B_2)$, and such that for any compactly supported $f \in L^q(X, B_1)$, and any $x \notin \text{supp}(f)$,

$$Tf(x) = \int K(x, y)f(y) \, dy,$$

458

where we assume the right hand side is absolutely integrable, i.e. that

$$\int \|K(x,y)f(y)\|_{B_2} \, dy < \infty.$$

Moreover, there is $A > 0$ and $c > 1$ such that for any $\delta > 0$, any $y \in X$, any $y' \in X$ with $\rho(y, y') < \delta$,

$$\int_{B(y,c\delta)^c} \|K(x,y) - K(x,y')\|_{B_1 \to B_2} \, dx \leqslant A,$$

where the norm here is the *operator norm*, viewing $K(x,y)$ as a bounded operator on B. This is satisfied, in particular, when $X = \mathbb{R}^d$, if

$$\|K(x,y)\|_{B_1 \to B_2} \leqslant |x - y|^{-d} \quad \text{and} \quad \|\nabla_y K(x,y)\|_{B_1 \to B_2} \lesssim |x - y|^{-(d+1)}.$$

The Calderon-Zygmund type proofs of the last section then apply, so we conclude that T is a bounded operator from $L^1(X, B_1)$ to $L^{1,\infty}(X, B_2)$, and thus for $1 < p \leqslant q$, bounded from $L^p(X, B_1)$ to $L^p(X, B_2)$.

Example. *Given a Calderon-Zygmund operator T, define a vector-valued operator on integer sequences of the functions by setting*

$$T(\{f_j\}) = \{Tf_j\}.$$

The kernel of this vector-valued operator is then given by $K(x,y)\{z_j\} = \{K_T(x,y)z_j\}$, where K_T is the kernel of T viewed as an operator on scalar valued functions. Now

$$\|K(x,y) - K(x,y')\|_{l^r(\mathbb{Z}) \to l^r(\mathbb{Z})} = |K_T(x,y) - K_T(x,y')|,$$

and so T is a Calderon-Zygmund operator on $L^p(X, l^r(\mathbb{Z}))$ for all $r > 0$. To apply the Calderon-Zygmund theory, we note that for $1 < r < \infty$, we obtain an apriori bound on T over $L^r(X, l^r(\mathbb{Z}))$ by Fubini and norm interchange. Thus Calderon-Zygmund theory gives that T is bounded from $L^1(X, l^r(\mathbb{Z}))$ to $L^{1,\infty}(X, l^r(\mathbb{Z}))$. In particular, for $1 < p \leqslant r$ we have

$$\left\|\left(\sum_j |Tf_j|^r\right)^{1/r}\right\|_{L^p(X)} \lesssim \left\|\left(\sum_j |f_j|^r\right)^{1/r}\right\|_{L^p(X)}.$$

If T^* is a Calderon-Zygmund operator, the result continues to hold for all $1 < p < \infty$. Here are also several more general results we follow from essentially the same argument:

459

- *For any measure space Y, if, for a function $f : X \times Y \to \mathbb{R}$, we define $Tf(x, y) = Tf_y(x)$, where $f_y(x) = f(x, y)$, then T is a Calderon-Zygmund operator on $L^p(X, L^r(Y))$, and we obtain that*

$$\|Tf\|_{L_x^{1,\infty} L_y^r} \lesssim \|f\|_{L_x^1 L_y^r}.$$

  Even more generally, we can consider a family of Calderon-Zygmund integral operators $\{T_z\}$ with kernels $\{K_z\}$, and provided that we have the cancellation condition

$$\int_{B(y, c\delta)^c} \sup_z |K_z(x, y) - K_z(x, y')| \, dx \leqslant A,$$

  the operator $Tf(x, y) = (T_y f_y)(x)$ is a Calderon-Zygmund integral operator on $L^p(X, L^r(Y))$ valued functions for $r > 0$, and for $1 < r < \infty$, we have an apriori $L^r(X, L^r(Y))$ bound by Fubini, and so

$$\|Tf\|_{L^{1,\infty}(X)L^r(Y)} \lesssim \|f\|_{L^1(X)L^r(Y)}.$$

- *For a Hilbert space H, let $C_r(H)$ be the space of all bounded operators $A : H \to H$ for which the quantity*

$$\|A\|_{C_r(H)} = Tr[(T^*T)^{r/2}]^{1/r}$$

  is finite. If T is a Calderon-Zygmund integral on X, then it's extension is a Calderon-Zygmund kernel on $C_r(H)$ valued functions for all $r > 0$. For $1 < r < \infty$, we have an apriori $L^r(X, C_r(H))$ bound (TODO: DO WE?)

Given any scalar-valued linear operator T, and a Banach space B, we can always extend T to an operator on B-valued functions by identifying the space of B valued functions, roughly speaking, with the tensor product of the space of scalar-valued functions and B. We thus extend T by defining

$$T(f \otimes v) = Tf \otimes v,$$

and then applying linearity to extend the definition to finite linear combinations of elementary tensor products. This kind of extension is precisely the one considered in the last example. If T is bounded from $L^p(X)$ to $L^q(Y)$, it is not necessarily true that it's extension is bounded from $L^p(X, B)$ to $L^q(Y, B)$. But if B is a Hilbert space, the result is always bounded.

Theorem 23.17. *If H is a Hilbert space, and T is a linear operator bounded from $L^p(X)$ to $L^q(Y)$ for some $0 < p, q \leqslant \infty$. Then T is bounded from $L^p(X, H)$ to $L^q(Y, H)$.*

Proof. Let $\{e_\alpha\}$ be a basis for H. We claim that for any $\{f_\alpha\}$, all but finitely many of which are non-zero, we have

$$\left\| T\left(\sum_\alpha f_\alpha \otimes e_\alpha\right) \right\|_{L^q(Y,H)} \lesssim \left\| \sum_\alpha f_\alpha \otimes e_\alpha \right\|_{L^p(Y,H)}.$$

In fact, this bound is equivalent to

$$\left\| \left(\sum_\alpha |Tf_\alpha|^2\right)^{1/2} \right\|_{L^q(Y)} \lesssim \left\| \left(\sum_\alpha |f_\alpha|^2\right)^{1/2} \right\|_{L^p(X)}.$$

Without loss of generality, we restrict ourselves to *real-valued* operators T. For any real valued $\lambda_1, \ldots, \lambda_n$, there is a constant C_r independent of n for which

$$\left(\sum_i \lambda_i^2\right)^{1/2} = C_r \, \mathbb{E}[|\lambda_1 X_1 + \cdots + \lambda_n X_n|^r]^{1/r} = \left\| \sum \lambda_i X_i \right\|_{L^r(\Omega)},$$

where $X_1, \ldots, X_n$ are independent and identically distributed normally distributed random variables in some probability space Ω. Let us first assume that $q \geqslant p$. Then

$$\left\| \left(\sum_\alpha |Tf_\alpha|^2\right)^{1/2} \right\|_{L^q(Y)} = C_q \left\| \sum_\alpha (Tf_\alpha) X_\alpha \right\|_{L^q(Y)L^q(\Omega)}$$

$$= C_q \left\| T\left(\sum X_\alpha f_\alpha\right) \right\|_{L^q(\Omega)L^q(Y)}$$

$$\lesssim C_q \left\| \sum X_\alpha f_\alpha \right\|_{L^q(\Omega)L^p(Y)}$$

$$\lesssim C_q \left\| \sum X_\alpha f_\alpha \right\|_{L^p(Y)L^q(\Omega)}$$

$$= \left\| \left(\sum_\alpha |f_\alpha|^2\right)^{1/2} \right\|_{L^p(X)}.$$

On the other hand, if $q < p$, one uses $L^p(\Omega)$, applies Hölder in the beginning and uses a similar argument, incurrence an error $C_q C_p^{-1}$ in the process. $\qquad\square$

Remark. The proof also gives results in Lorentz spaces, since the norm interchange trick remains true in that class as well.

By interpolation, we obtain a corollary of this result.

461

Corollary 23.18. *If T is a bounded linear operator from $L^p(X)$ to $L^q(Y)$ for $1 \leqslant p < \infty$, and r is contained in the interval between 2 and p, then*

$$\left\| \left(\sum_j |Tf_j|^r \right)^{1/r} \right\|_{L^q(Y)} \lesssim \left\| \left(\sum_j |f_j|^r \right)^{1/r} \right\|_{L^p(X)}.$$

One important application of these results is to *square functions*. For instance, we can use singular integrals to prove the Littlewood-Paley square function bounds.

Example. *Consider a function Φ with $\int \Phi = 0$. Set $\Phi_t = t^{-n} Dil_t \Phi$, and define*

$$(s_\Phi f)(x) = \left(\int_0^\infty |(f * \Phi_t)(x)|^2 \frac{dt}{t} \right)^{1/2},$$

If we now consider the vector-valued *operator T taking scalar-valued functions f to vector-valued functions over the Banach space $L^2(\mathbb{R}_+, x^{-1}dx)$ given by*

$$Tf(x, t) = (f * \Phi_t)(x),$$

then T is a Calderon-Zygmund operator over $L^2(\mathbb{R}_+, t^{-1}dt)$. It's kernel is

$$K(x, y)(t) = \Phi_t(x - y),$$

and we have the estimates

$$\begin{aligned}
\|\Phi_t(y)\|_{L^2(\mathbb{R}_+, t^{-1}dt)} &= \left(\int |K(y)(t)|^2 \, t^{-1} \, dt \right)^{1/2} \\
&= \left(\int t^{-2n} |\Phi \left(\frac{y}{t} \right)|^2 \, t^{-1} \, dt \right)^{1/2} \\
&= |y|^{-n} \left(\int t^{2n} |\Phi(t)|^2 \, t^{-1} \, dt \right) \\
&\lesssim |y|^{-n},
\end{aligned}$$

and

$$\begin{aligned}
\|\nabla K(y)\|_{L^2(\mathbb{R}_+, t^{-1}dt)} &= \left(\int |\nabla K(y)(t)|^2 \, t^{-1} \, dt \right)^{1/2} \\
&= \left(\int t^{-2n-2} |\nabla \Phi \left(\frac{x-y}{t} \right)|^2 \, t^{-1} \, dt \right)^{1/2} \\
&= |y|^{-n-1} \left(\int t^{2n} |\nabla \Phi(t)|^2 \, t^{-1} \, dt \right) \\
&\lesssim |y|^{-n-1}.
\end{aligned}$$

462

Thus we see T is a Calderon-Zygmund operator of convolution type. Since T is bounded on $L^2(\mathbb{R}^n, L^2(\mathbb{R}_+, t^{-1} dt))$, we immediately obtain that

$$\left\| \left(\int_0^\infty |(\Phi_t * f)(x)| \right)^{1/2} \right\|_{L^{1,\infty}(\mathbb{R}^n)} \lesssim \|f\|_{L^1(\mathbb{R}^n)},$$

and for $1 < p < \infty$,

$$\left\| \left(\int_0^\infty |(\Phi_t * f)(x)| \right)^{1/2} \right\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}.$$

Similar results apply to discrete analogues.

Example. *Consider bounds of the form*

$$\left\| \left(\sum_{k \in \mathbb{Z}} |(\Phi_{2^k} * f)(x)|^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}.$$

The kernel of this operator, mapping scalars to elements of $l^2(\mathbb{Z})$, is given by

$$K(x) = \{\Phi_{2^k}(x - y)\},$$

By translation invariance, to show this kernel gives a Calderon-Zygmund kernel, we must show that for $|x| \gtrsim \delta$ and $|y| < \delta$,

$$\int_{|x| \gtrsim \delta} \|K(x - y) - K(x)\|_{l^2(\mathbb{Z})} \, dx \lesssim 1.$$

We will be able to do this assuming that Φ is C^1, and satisfies bounds of the form

$$|\Phi(x)| + |\nabla \Phi(x)| \lesssim 1/|x|^{n+1}.$$

We note that this assumption, together with the mean value theorem, guarantees that

$$|\Phi_{2^k}(x - y) - \Phi_{2^k}(x)| \lesssim \frac{\delta}{2^{k(n+1)} \langle x/2^k \rangle^{n+1}},$$

and

$$|\Phi_{2^k}(x - y) - \Phi_{2^k}(x)| \lesssim \frac{1}{2^{kn} \langle x/2^k \rangle^{n+1}}.$$

463

This implies that

$$\|K(x - y) - K(x)\|_{l^2(\mathbb{Z})} \lesssim \delta \log(|x|/\delta)^{1/2}|x|^{-(n+1)},$$

and this is sufficient to conclude the required cancellation bound. To obtain the bounds above, we thus need to obtain an apriori bound for the operator. But we can obtain this over $L^2(\mathbb{R}^d)$, because

$$\left\| \left(\sum\nolimits_{k \in \mathbb{Z}} |(\Phi_{2^k} * f)(x)|^2 \right)^{1/2} \right\|_{L^2(\mathbb{R}^n)} = \left(\sum\nolimits_{k \in \mathbb{Z}} \|\Phi_{2^k} * f\|^2_{L^2(\mathbb{R}^d)} \right)^{1/2},$$

Our assumptions imply that

$$|\hat{\Phi}(\xi)| \lesssim \min(|\xi|^{1/2}, |\xi|^{-1}),$$

and thus

$$\sum\nolimits_{k \in \mathbb{Z}} |\hat{\Phi}(2^k \xi)|^2 \lesssim 1.$$

Applying Parseval, we thus conclude that

$$\begin{aligned}
\left(\sum\nolimits_{k \in \mathbb{Z}} \|\Phi_{2^k} * f\|^2_{L^2(\mathbb{R}^d)} \right)^{1/2} \\
= \left(\int_{\mathbb{R}^d} \sum\nolimits_{k \in \mathbb{Z}} |\hat{\Phi}(2^k \xi)|^2 |\hat{f}(\xi)|^2 \, d\xi \right)^{1/2} \\
\lesssim \|\hat{f}\|_{L^2(\mathbb{R}^d)} \\
\lesssim \|f\|_{L^2(\mathbb{R}^d)}.
\end{aligned}$$

Calderon-Zygmund theory thus gives $L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R}^d, l^2(\mathbb{Z}))$ bounds and $L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d, l^2(\mathbb{Z}))$ bounds for $1 < p < \infty$. Thus we've proved an upper *bound on the Littlewood-Paley Square Function.*

Provided that $\sum_{k \in \mathbb{Z}} |\Phi_{2^k}(x)|^2 = 1$, we can prove a converse to this bound. To simplify notation, define $\Delta_j f = \Psi_{2^j} * f$.

Proof. Similar analyis to above shows that

$$\left\| \left(\sum\nolimits_j |\Delta_j f|^2 \right)^{1/2} \right\|_{L^2(\mathbb{R}^d)} \gtrsim \|f\|_{L^2(\mathbb{R}^d)},$$

Applying Cauchy-Schwartz and Hölder's inequality gives that

$$
\begin{aligned}
\left\langle \sum_j \Delta_j^* \Delta_j f, g \right\rangle &= \sum_j \langle \Delta_j f, \Delta_j g \rangle \\
&= \int_{\mathbb{R}^d} \sum_j (\Delta_j f)(x)(\Delta_j g)(x) \\
&\leqslant \int_{\mathbb{R}^d} \left(\sum_j |(\Delta_j f)(x)|^2 \right)^{1/2} \left(\sum_j |(\Delta_j g)(x)|^2 \right)^{1/2} \\
&\leqslant \left\| \left(\sum_j |\Delta_j f|^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^d)} \left\| \left(\sum_j |\Delta_j g|^2 \right)^{1/2} \right\|_{L^{p'}(\mathbb{R}^d)} \\
&\lesssim \left\| \left(\sum_j |\Delta_j f|^2 \right)^{1/2} \right\|_{L^p(\mathbb{R}^d)} \|g\|_{L^{p'}(\mathbb{R}^d)}.
\end{aligned}
$$

Taking suprema over g with $\|g\|_{L^{p'}(\mathbb{R}^d)} \leqslant 1$ gives

$$
\|f\|_{L^p(\mathbb{R}^d)} = \|\sum_j \Delta_j^* \Delta_j f\|_{L^p(\mathbb{R}^d)} \lesssim \|(\sum_j |\Delta_j f|^2)^{1/2}\|_{L^p(\mathbb{R}^d)}. \qquad \square
$$

Remark. A similar analysis works if $\sum \Phi_{2^k}(x) = 1$ for $x \neq 0$, since then we have $f = \sum_j \Delta_j f$, and we can apply Hölder's inequality twice instead of applying Cauchy-Schwartz.

A similar argument gives results for *vector-valued Littlewood Paley*, i.e. bounding quantities of the form

$$
\|\Delta_k(f_j)\|_{L_x^p l_j^r l_k^2} \lesssim \|f_j\|_{L_x^p l_j^r}
$$

for $1 < r < \infty$ and $1 < p < \infty$. The proof is a simple generalization.

Chapter 24

Fourier Multiplier Operators

Our aim in this chapter is to study the boundedness of *Fourier multiplier operators*. Given a function $m : \mathbb{R}^d \to \mathbb{C}$, known as a *symbol*, we want to associate a multiplier operator T, sometimes denoted $m(D)$, which when applied to a function $f : \mathbb{R}^d \to \mathbb{C}$ should be formally given by the equation

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

In maximum generality, for any tempered distribution m on $\mathbb{R}^d$ we can define T as a continuous operator from $\mathcal{S}(\mathbb{R}^d)$ 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 $T f = K * f$. 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 $\mathcal{S}(\mathbb{R}^d)$ (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* $\mathbb{R}$, *consider a rough cutoff* $\mathbb{I}_{[-1,1]}$. *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* $\mathbb{R}$, *we consider the Fourier multiplier*

$$m(\xi) = -i \cdot sgn(\xi).$$

Then m(D) is the Hilbert transform.

Example. *In* $\mathbb{R}^d$, *we consider the Fourier multiplier*

$$m_R(\xi) = \mathbb{I}(|\xi| \leqslant R).$$

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* $\mathbb{I}_S(D)$. *In the engineering literature these multipliers are called* ideal low pass filters.

Example. *In this chapter, it is natural to renormalize the differentiation operators* $D^\alpha : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ *so that for* $f \in \mathcal{S}(\mathbb{R}^d)$,

$$\widehat{D^\alpha f} = \xi^\alpha \hat{f}.$$

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 $T : C_c^\infty(\mathbb{R}^d) \to C_{\text{loc}}^\infty(\mathbb{R}^d)$ 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

$$\|Tf\|_{L^q(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

for any $f \in \mathcal{S}(\mathbb{R}^d)$, then T is a Fourier multiplier operator. To prove this, we apply the theory of Sobolev embeddings.

Theorem 24.1. *Suppose $1 \leqslant p, q \leqslant \infty$, and $T : \mathcal{S}(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is a linear map commuting with translations and satisfies*

$$\|Tf\|_{L^q(\mathbb{R}^d)} \leqslant \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in \mathcal{S}(\mathbb{R}^d)$. Then T is a Fourier multiplier operator.

Proof. For any $f \in \mathcal{S}(\mathbb{R}^d)$, $Tf \in W^{q,n}(\mathbb{R}^d)$ for any $n > 0$. To see this, we note that for any $h > 0$ and $k \in \{1, \ldots, d\}$, 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 $L^p(\mathbb{R}^d)$. Thus by continuity of f, Tf has a strong derivative $T(D^k f)$ in $L^q(\mathbb{R}^d)$. Induction shows Tf has strong derivatives of all orders. The last lemma shows that $Tf \in C(\mathbb{R}^d)$, and

$$\begin{aligned}
\|Tf\|_{L^\infty(\mathbb{R}^d)} &\lesssim \sum_{|\alpha| \leqslant n+1} \|D^\alpha(Tf)\|_{L^q(\mathbb{R}^d)} \\
&= \sum_{|\alpha| \leqslant n+1} \|T(D^\alpha f)\|_{L^q(\mathbb{R}^d)} \\
&\lesssim \sum_{|\alpha| \leqslant n+1} \|D^\alpha f\|_{L^q(\mathbb{R}^d)}.
\end{aligned}$$

The map $f \mapsto Tf(0)$ is thus a continuous operator on $\mathcal{S}(\mathbb{R}^d)$, and therefore defines a tempered distribution Λ. Translation invariance shows that $Tf = \Lambda * f$, and setting $m = \widehat{\Lambda}$ completes the proof. $\qquad\square$

Remark. It follows from this argument that if $T : \mathcal{S}(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is a linear operator commuting with translations satisfying a bound

$$\|Tf\|_{L^q(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)},$$

then for any $f \in \mathcal{S}(\mathbb{R}^d)$, $Tf \in C^\infty_{\text{loc}}(\mathbb{R}^d)$ and is slowly increasing, as is all of it's derivatives.

24.1 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, $\widehat{m(D)f} = m\hat{f}$ is supported on S. Frequency localization has the additional feature that $m(D)f$ is smooth (analytic even), since it's Fourier transform is compactly supported. One major advantage of localizing in frequency is that it enables us to control the behaviour of the derivatives 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 localization useful especially useful in problems involving derivatives. The uncertainty principle also gives another heuristic application of this principle; if m is supported on a cube centred at the origin with sidelengths $R_1, \ldots, R_d$, then the function $m(D)f$ will be roughly speaking, locally constant on 'dual rectangles' with sidelength $1/R_1, \ldots, 1/R_d$. More generally, if we instead choose m supported on a cube centered at ξ_0 with sidelengths $R_1, \ldots, R_n$, 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 since the behaviour of certain operators can be exploited more clearly once inputs or outputs have been localized in frequency.

Let us consider some examples. If we consider the Fourier multiplier $\mathbb{I}_{[-R,R]}$ in $\mathbb{R}^1$, then this multiplier corresponds to the kernel

$$K(x) = \int_{-R}^{R} e^{2\pi i \xi x} = \frac{\sin(2\pi Rx)}{\pi x} = 2R \cdot \text{sinc}(2\pi Rx),$$

i.e. a sinc function. If f is supported on an interval I, then it follows from a simple estimate that

$$|(K * f)(x)| \lesssim \frac{\|f\|_{L^1(\mathbb{R})}}{d(x, I)}.$$

Thus we see that we can localize in frequency while remaining somewhat localized in space, but with some additional fuzziness that decays away from this interval at a rate of $1/x$. Unfortunately,this is often not enough to obtain useful estimates, e.g. the fuzziness does not even lie in $L^1(\mathbb{R})$. In $\mathbb{R}^d$, if we write $K = K_1 \otimes \cdots \times K_d$ be the kernel associated with the rectangle multiplier, then we have a similar $1/x$ decay estimate (though we do get up to $1/x^d$ decay in directions not close to being parallel with any axis, an error that still remains non integrable).

We can do better if we use a smooth cutoff function, i.e. we choose a smooth non-negative function m compactly supported on a cube I of sidelength L which equals one on the interior third of I, and satisfies

$$\|D^\alpha m\|_{L^\infty(\mathbb{R}^d)} \lesssim_\alpha L^{-|\alpha|}$$

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 $n > 0$, $|K(x)| \lesssim_n L^{d-n} |x|^{-n}$, which decreases away from the origin much faster than for a rough cutoff. This implies that we have the decay estimates

$$|(K * f)(x)| \lesssim_n \frac{L^{d-n} \|f\|_{L^1(\mathbb{R}^d)}}{d(x, J)^n}.$$

for f supported on a cube J, and all $n > 0$. Thus the majority of the mass of $K * f$ is concentrated in an $O(1/L)$-thickening of the interval J, so spatial localization is preserved provided the sidelengths of J are $\Omega(1/L)$. This means the 'fuzziness' outside of the interval J here is fast decaying in a quantitative manner, which is often enough that it can be safely ignored. Thus frequency localization in this setting also preserves spatial localization as long as we respect the uncertainty principle, i.e. we do not care about localization in space to an accuracy finer than $\Omega(1/L)$.

24.2 L^p Regularity

We now wish to study conditions on m which guarantee bounds of the form

$$\|m(D)f\|_{L^q(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

for all $f \in \mathcal{S}(\mathbb{R}^d)$. 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.

Example. *Let $k_s(x) = |x|^{-s}$, for $0 < s < d$ be the* Riesz kernel. *Then $k_s \in L^{r,\infty}(\mathbb{R}^d)$ for $r = d/s$, so the weak-type variant of Young's inequality implies that*

$$\|k_s * f\|_{L^q(\mathbb{R}^d)} \lesssim_{p,q} \|f\|_{L^p(\mathbb{R}^d)}$$

provided $1/p - 1/q = 1 - s/d$. Analysing the scaling properties of each side of this equation shows these are the only exponents for which this inequality holds. Now it turns out that $\hat{k}_s = C_{d,s} k_{d-s}$ for some constants $C_{d,s}$, and so it follows that $k_s \in M^{p,q}(\mathbb{R}^d)$ when $1/p - 1/q = s/d$.

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 $\|m\|_{M^{p,q}(\mathbb{R}^d)}$ denote the operator norm of $m(D)$ from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}(\mathbb{R}^d)^*$, via the $L^p(\mathbb{R}^d)$ norm in the input, and via the $L^q(\mathbb{R}^d)$ norm in the output, i.e. it is the smallest quantity $\|m\|_{M^{p,q}(\mathbb{R}^d)}$ such that

$$\|m(D)f\|_{L^q(\mathbb{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in \mathcal{S}(\mathbb{R}^d)$. If $1 \leqslant p < \infty$, then the Hahn-Banach density theorem shows every element of $M^{p,q}(\mathbb{R}^d)$ extends uniquely to an operator from $L^p(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$ with the same operator norm. If $p = \infty$, then every such operator extends to a map from $L^\infty(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$ with the same operator norm, but this extension need not be unique. We let $M^{p,q}(\mathbb{R}^d)$ be the set of tempered distributions for which the bound is finite. For simplicity, we also let $M^p(\mathbb{R}^d)$ denote $M^{p,p}(\mathbb{R}^d)$. All of these spaces are Banach spaces. By symmetries of the Fourier transform, it is easy to check that translations, modulations, and dilations all preserve the space $M^{p,q}(\mathbb{R}^d)$, though dilations need not preserve the $M^{p,q}$ norm unless $p = q$. Thus we have a family of Banach Algebras $M^p(\mathbb{R}^d)$. On these spaces, translation, dilation, conjugation, and modulation are all isometries, so this space is highly symmetric.

Littlewood's principle tells us that the only interesting Fourier multipliers in $M^{p,q}(\mathbb{R}^d)$ occur with 'the larger exponent on the left'.

Theorem 24.2. *Fix $1 \leqslant q < p \leqslant \infty$, and suppose m is a tempered distribution on $\mathbb{R}^d$ satisfying a uniform bound*

$$\|m(D)f\|_{L^q(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in \mathcal{S}(\mathbb{R}^d)$. Then $m = 0$. In other words, $M^{p,q}(\mathbb{R}^d) = 0$ for $q < p$.

Proof. Suppose $m \neq 0$ and $q < p$. Then there is $f_0 \in \mathcal{S}(\mathbb{R}^d)$ with $m(D)f_0 \neq 0$. Thus $m(D)f_0$ lies in $L^q(\mathbb{R}^d)$. Fix a large integer N and pick $x_1, \ldots, x_N \in \mathbb{R}^d$ separated far enough apart that

$$\left\| \sum_{n=1}^{N} \mathrm{Trans}_{x_n} f_0 \right\|_{L^p(\mathbb{R}^d)} \gtrsim N^{1/p} \|f_0\|_{L^p(\mathbb{R}^d)}$$

and

$$\left\| \sum_{n=1}^{N} \mathrm{Trans}_{x_n} m(D) f_0 \right\|_{L^q(\mathbb{R}^d)} \sim N^{1/q} \|m(D)f_0\|_{L^q(\mathbb{R}^d)} \lesssim N^{1/q} \|f_0\|_{L^p(\mathbb{R}^d)}.$$

Translation invariance of convolution shows $N^{1/q} \lesssim N^{1/p}$, which is impossible for suitably large N. Thus $m = 0$. $\qquad\square$

Example. *We claim that $M^{2,2}(\mathbb{R}^d) = L^\infty(\mathbb{R}^d)$, in the sense that the two spaces consist of the same family of distributions, and the norms on both spaces are equal, i.e. $\|m\|_{M^{2,2}(\mathbb{R}^d)} = \|m\|_{L^\infty(\mathbb{R}^d)}$. To see this, suppose either $m \in M^{2,2}(\mathbb{R}^d)$ or $m \in L^\infty(\mathbb{R}^d)$, and 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 $\Phi \in L^2(\mathbb{R}^d)$, $\Phi \cdot m \in L^2(\mathbb{R}^d)$. But since $1/\Phi \in L^\infty_{loc}(\mathbb{R}^d)$, this implies that $m \in L^1_{loc}(\mathbb{R}^d)$.

Now we know m is locally integrable, we can obtain the general result by Parseval's inequality, since a bound of the form

$$\|m(D)f\|_{L^2(\mathbb{R}^d)} \leqslant C\|f\|_{L^2(\mathbb{R}^d)}$$

holds for all $f \in \mathcal{S}(\mathbb{R}^d)$ if and only if

$$\|m \cdot g\|_{L^2(\mathbb{R}^d)} \leqslant C\|g\|_{L^2(\mathbb{R}^d)}$$

for all $g \in \mathcal{S}(\mathbb{R}^d)$. The former bound is equivalent to $\|m\|_{M^{2,2}(\mathbb{R}^d)} \leqslant C$, and the second bound is equivalent to $\|m\|_{L^\infty(\mathbb{R}^d)} \leqslant C$ by Hölder's inequality.

For any tempered distribution m and $f, g \in \mathcal{S}(\mathbb{R}^d)$, 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. In particular, this means that if R is the reflection operator $[Ru](x) = u(-x)$, then

$$
m(D)f = [(R \circ m^*(D) \circ R)f^*]^*.
$$

In particular, for any $0 < p \leqslant \infty$, $\|m(D)f\|_{L^p(\mathbb{R}^d)} = \|m(D)^* Rf^*\|_{L^p(\mathbb{R}^d)}$.

Theorem 24.3. *For any $1 \leqslant p \leqslant q < \infty$ and any tempered distribution m,*

$$
\|m\|_{M^{p,q}(\mathbb{R}^d)} = \|m\|_{M^{q*,p*}(\mathbb{R}^d)}.
$$

Proof. It will suffice to prove by duality that

$$
\|m\|_{M^{q*,p*}(\mathbb{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbb{R}^d)}.
$$

Assume without loss of generality that $\|m\|_{M^{p,q}(\mathbb{R}^d)} < \infty$. Given $f \in \mathcal{S}(\mathbb{R}^d)$, we know $m(D)f \in C^\infty_{\text{loc}}(\mathbb{R}^d) \cap L^q(\mathbb{R}^d)$. If $p \neq \infty$, then

$$
\|m(D)f\|_{L^{p*}(\mathbb{R}^d)} = \sup \left\{ \int m(D)f(x)g(x)\, dx : g \in \mathcal{S}(\mathbb{R}^d), \|g\|_{L^p(\mathbb{R}^d)} < \infty \right\}.
$$

Given any such $g \in \mathcal{S}(\mathbb{R}^d)$, we have by Hölder's inequality,

$$
\begin{aligned}
\int m(D)f(x)g(x)\, dx &= \int f(x)m^*(D)g(x)\, dx \\
&\leqslant \|f\|_{L^{q*}(\mathbb{R}^d)} \|m^*(D)g\|_{L^q(\mathbb{R}^d)} \\
&\leqslant \|m\|_{M^{p,q}(\mathbb{R}^d)} \|f\|_{L^{q*}(\mathbb{R}^d)}.
\end{aligned}
$$

Thus we conclude that

$$
\|m(D)f\|_{L^{p*}(\mathbb{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbb{R}^d)} \|f\|_{L^{q*}(\mathbb{R}^d)}
$$

473

and thus $\|m\|_{M^{q*,p*}(\mathbb{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbb{R}^d)}$. On the other hand, if $p = \infty$, then $q = \infty$ (or else the inequality is trivial). If $\|m\|_{M^{\infty,\infty}(\mathbb{R}^d)} < \infty$, then it follows that for any $f \in \mathcal{S}(\mathbb{R}^d)$, $m(D)f \in C^{\infty}(\mathbb{R}^d)$. It therefore follows by smoothness that

$$|m(D)f(0)| \leqslant \|m(D)f\|_{L^{\infty}(\mathbb{R}^d)} \leqslant \|m\|_{M^{\infty,\infty}(\mathbb{R}^d)} \|f\|_{L^{\infty}(\mathbb{R}^d)}.$$

A consequence of this is that for all $f \in \mathcal{S}(\mathbb{R}^d)$,

$$\left| \int \widehat{m}(x) f(x) \, dx \right| \leqslant \|m\|_{M^{\infty,\infty}(\mathbb{R}^d)} \|f\|_{L^{\infty}(\mathbb{R}^d)}.$$

This means $\widehat{m}$ is a distribution of order zero, and is therefore equal to some Radon measure μ. Moreover, the bound actually implies that μ is a finite Borel measure, with total variation at most $\|m\|_{M^{\infty,\infty}(\mathbb{R}^d)}$. But then Young's inequality implies that

$$\|\widehat{m} * f\|_{L^1(\mathbb{R}^d)} \leqslant \|\widehat{m}\|_{M(\mathbb{R}^d)} \|f\|_{L^1(\mathbb{R}^d)} \leqslant \|m\|_{M^{\infty,\infty}(\mathbb{R}^d)} \|f\|_{L^1(\mathbb{R}^d)}.$$

Thus $\|m\|_{M^{1,1}(\mathbb{R}^d)} \leqslant \|m\|_{M^{\infty,\infty}(\mathbb{R}^d)}$, which completes the proof of duality. $\qquad\square$

We also have a rescaling law which follows from the standard rescaling properties of the Fourier transform, i.e. for any matrix $A \in \mathrm{GL}(d)$,

$$\|m \circ A\|_{M^{p,q}(\mathbb{R}^d)} = |\det(A)|^{-|1/p - 1/q|} \|m\|_{M^{p,q}(\mathbb{R}^d)}.$$

In particular, the quantities $\|m\|_{M^p(\mathbb{R}^d)}$ are invariant under scaling.

In particular, if $1 \leqslant p \leqslant \infty$ and $m \in M^p(\mathbb{R}^d)$, then also $m \in M^{p*}(\mathbb{R}^d)$ and so Riesz-interpolation implies $m \in M^2(\mathbb{R}^d)$. Thus if we are studying L^p to L^p boundedness for any $1 \leqslant p \leqslant \infty$, we may restrict our attention to Fourier multipliers with a bounded symbol. More generally, this interpolation approach shows that $M^p(\mathbb{R}^d)$ is a larger space of multipliers the closer p is to two. Thus $M^2(\mathbb{R}^d) = L^{\infty}(\mathbb{R}^d)$ is the largest family, and $M^1(\mathbb{R}^d) = M^{\infty}(\mathbb{R}^d)$ is the smallest family of multipliers bounded on some $M^p(\mathbb{R}^d)$. It turns out we can also completely characterize this space of multipliers, it is the space of Fourier transforms of finite Radon measures.

Example. *The only remaining space which we can completely characterize are the spaces $M^{1,q}(\mathbb{R}^d)$ (and thus $M^{q*,\infty}$), where $1 \leqslant q \leqslant \infty$. For $q > 1$, we have*

$$M^{1,q}(\mathbb{R}^d) = \widehat{L^q(\mathbb{R}^d)}$$

and $M^{1,1}(\mathbb{R}^d) = \widehat{M(\mathbb{R}^d)}$, the set of all finite Borel measures. Moreover, in the case $q > 1$ we have

$$\|m\|_{M^{1,q}(\mathbb{R}^d)} = \|\widehat{m}\|_{L^q(\mathbb{R}^d)}$$

and for $q = 1$,

$$\|m\|_{M^{1,1}(\mathbb{R}^d)} = \|\widehat{m}\|_{M(\mathbb{R}^d)}.$$

Given such an $m \in M^{1,q}(\mathbb{R}^d)$, if $K = \widehat{m}$, and $\{\phi_\varepsilon\}$ is a family of standard mollifiers, then

$$\|K * \phi_\varepsilon\|_{L^q(\mathbb{R}^d)} = \|m(D)(\phi_\varepsilon)\|_{L^q(\mathbb{R}^d)} \lesssim \|\phi_\varepsilon\|_{L^1(\mathbb{R}^d)} = 1.$$

Applying the Banach-Alaoglu theorem, the family $\{K * \phi_\varepsilon\}$ has a weak $*$ convergent subsequence in $L^q(\mathbb{R}^d)^{**}$. Thus there is $\lambda \in L^q(\mathbb{R}^d)^{**}$ 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 $K = \lambda \in (L^q(\mathbb{R}^d))^{**}$. If $q > 1$, then $(L^q(\mathbb{R}^d))^{**} = L^q(\mathbb{R}^d)$, with the same norm, and $(L^1(\mathbb{R}^d))^{**} = M(\mathbb{R}^d)$.

Another case where the multiplier problem is trivial is for $m \in L^\infty(\mathbb{R}^d)$ with $\widehat{m} \geqslant 0$. This is related to the fact that the study of kernels which are non-negative and homogeneously distributed is also trivial, though the fact that we are working over a space with infinite measure complicates the analysis somewhat. To argue more precisely, by Lemma 4.12, $\|m\|_{L^\infty(\mathbb{R}^d)} = \|\widehat{m}\|_{L^1(\mathbb{R}^d)}$, and so for $1 \leqslant p \leqslant \infty$,

$$\|\widehat{m}\|_{L^1(\mathbb{R}^d)} = \|m\|_{M^2} \leqslant \|m\|_{M^p} \leqslant \|m\|_{M^1} = \|\widehat{m}\|_{L^1(\mathbb{R}^d)},$$

so $\|m\|_{M^p} = \|\widehat{m}\|_{L^1(\mathbb{R}^d)}$ for all $p \in [1, \infty]$.

We have a version of Young's inequality for multipliers. A scaling analysis shows that we can only have an inequality

$$\|m_1 * m_2\|_{M^{p,q}(\mathbb{R}^d)} \leqslant \|m_1\|_{L^{r_1}(\mathbb{R}^d)} \|m_2\|_{L^{r_2}(\mathbb{R}^d)}$$

if $1 + (1/p - 1/q) = 1/r_1 + 1/r_2$.

Lemma 24.4. *If $m_1 \in L^{r_1}(\mathbb{R}^d)$ and $m_2 \in L^{r_2}(\mathbb{R}^d)$, then*

$$\|m_1 * m_2\|_{M^{p,q}(\mathbb{R}^d)} \leqslant \|m_1\|_{L^{r_1}(\mathbb{R}^d)} \|m_2\|_{L^{r_2}(\mathbb{R}^d)}$$

for $1 \leqslant p, q \leqslant \infty$, and $1 \leqslant r_1, r_2 \leqslant 2$, and $1 + |1/p - 1/q| = 1/r_1 + 1/r_2$.

Proof. We apply complex interpolation. Young's inequality implies $m_1 * m_2 \in L^r$ for $r = 1/|1/p - 1/q|$. In particular, if $p = q = 2$, then $m_1 * m_2 \in L^\infty(\mathbb{R}^d)$, which means

$$\|m_1 * m_2\|_{M^2(\mathbb{R}^d)} = \|m_1 * m_2\|_{L^\infty(\mathbb{R}^d)} \leqslant \|m_1\|_{L^{r_1}(\mathbb{R}^d)} \|m_2\|_{L^{r_2}(\mathbb{R}^d)}.$$

475

If $p = q = 2$,

If $p = 1$, then $2 - 1/q = 1/r_1 + 1/r_2$ and $1 \leqslant r_1, r_2 \leqslant 2$. We may then apply Hölder's inequality and the Hausdorff Young inequality to conclude that

$$
\begin{aligned}
\|m_1 * m_2\|_{M^{1,q}(\mathbb{R}^d)} &= \|\widehat{m_1 * m_2}\|_{L^q(\mathbb{R}^d)} \\
&= \|\widehat{m_1}\widehat{m_2}\|_{L^q(\mathbb{R}^d)} \\
&\leqslant \|\widehat{m_1}\|_{L^{r_1^*}(\mathbb{R}^d)} \|\widehat{m_2}\|_{L^{r_2^*}(\mathbb{R}^d)} \\
&\leqslant \|m_1\|_{L^{r_1}} \|m_2\|_{L^{r_2}(\mathbb{R}^d)}.
\end{aligned}
$$

Next, for $p = q = 2$, we have $1/r_1 + 1/r_2 = 1$, and so we can again apply Hausdorff Young to conclude that

$$
\begin{aligned}
\|m_1 * m_2\|_{M^2(\mathbb{R}^d)} &= \|m_1 * m_2\|_{L^\infty(\mathbb{R}^d)} \\
&\leqslant \|m_1\|_{L^{r_1}(\mathbb{R}^d)} \|m_2\|_{L^{r_2}(\mathbb{R}^d)}.
\end{aligned}
$$

For the remaining cases, given an arbitrary $f \in \mathcal{S}(\mathbb{R}^d)$, and $m_1, m_2 \in C_c^\infty(\mathbb{R}^d)$, the quantity

$$
\|(m_1 * m_2)f\|_{L^p(\mathbb{R}^d)} = \left(\int \left| \int (m_1 * m_2)(\xi)\hat{f}(\xi)e^{2\pi i \xi \cdot x} \, d\xi \right|^p \, dx \right)^{1/p}
$$

is holomorphic in p, and so we can apply complex interpolation to prove that

$$
\|(m_1 * m_2)f\|_{L^q(\mathbb{R}^d)} \leqslant \|m_1\|_{L^{r_1}(\mathbb{R}^d)} \|m_2\|_{L^{r_2}(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)}. \qquad \square
$$

Essentially for any other family of multipliers, it is still incredibly difficult to determine conditions when, especially when $d > 1$. For instance, it still remains a major open question in harmonic analysis, for $d > 2$, to determine the values of $p \in [1, \infty]$ and $\delta > 0$ for which the multiplier

$$
m^\delta(\xi) = (1 - |\xi|^2)_+^\delta = \max((1 - |\xi|^2)^\delta, 0)
$$

lies in $M^p(\mathbb{R}^d)$, a problem known as the *Bochner-Riesz conjecture*.

The difficulty which causes m^δ to be unbounded is that it is singular on the boundary of the unit sphere, which is a large, curved set, upon which m^δ has δ degrees of regularity in the direction tangential to the sphere. 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

476

of this phenomenon occurs if $m \in \mathcal{S}(\mathbb{R}^d)$. Then $\hat{m} \in \mathcal{S}(\mathbb{R}^d)$, hence integrable, and by Young's convolution inequality, it follows that for any $p \leqslant q$, $\|m\|_{M^{p,q}(\mathbb{R}^d)} \lesssim 1$, and the induced map from $\mathcal{S}(\mathbb{R}^d)$ to $M^{p,q}(\mathbb{R}^d)$ is continuous. Another example is the following calculation: if $\Omega \subset \mathbb{R}^d$ is bounded and open, C is a fixed constant, and $L : \mathbb{R}^d \to \mathbb{R}^d$ 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 $|D^k(\phi \circ L)(x)| \leqslant C$ for all $x \in \Omega$. If m is a bump function adapated to $L(\Omega)$, then it follows that $\|m\|_{M^{p,q}(\mathbb{R}^d)} \lesssim_{C,\Omega} 1$.

It is important to note that it is mainly qualitative information that is of interest here. This is because $\|m \circ A\|_{M^p(\mathbb{R}^d)} = \|m\|_{M^p(\mathbb{R}^d)}$ for any multiplier m and invertible linear transformation A. Thus despite the fact that a singularity can become quantitatively more singular as we focus in on it by introducing an appropriate transformation A, the $M^p(\mathbb{R}^d)$ norm stays the same.

Example. *Consider the sawtooth function $m(\xi) = sgn(\xi) \max(1 - |\xi|, 0)$. For each $R > 0$, consider $m_R(\xi) = sgn(\xi) \max(1 - |\xi/R|, 0)$. Then m_R converges pointwise almost everywhere to $m_\infty(\xi) = sgn(\xi)$, and actually $m_R(D)$ converges to $m_\infty(D)$ distributinally. Here $m_\infty(D)$ is just the Hilbert transform. Now $\|m_R\|_{M^p(\mathbb{R}^d)} = \|m\|_{M^p(\mathbb{R}^d)}$ for all $p \in [1, \infty]$, so we might expect that $\|m\|_{M^p(\mathbb{R}^d)} = \|m_\infty\|_{M^p(\mathbb{R}^d)}$. Certainly the distributional limiting process implies that*

$$\|m_\infty\|_{M^p(\mathbb{R}^d)} \leqslant \|m\|_{M^p(\mathbb{R}^d)}.$$

Conversely, $m(\xi) = a(\xi)m_\infty(\xi)$, where $a(\xi) = \max(1 - |\xi|, 0)$ is a tent function. Now $\hat{a}(x) = sinc(x)^2$ is non-negative, so in particular, this means that for any p, $\|a\|_{M^p(\mathbb{R}^d)} = \|a\|_{L^\infty(\mathbb{R}^d)} = 1$. But this means that

$$\|m\|_{M^p(\mathbb{R}^d)} \leqslant \|a\|_{M^p(\mathbb{R}^d)}\|m_\infty\|_{M^p(\mathbb{R}^d)} \leqslant \|m_\infty\|_{M^p(\mathbb{R}^d)},$$

which completes the proof.

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 $m(\xi) = isgn(\xi)$, is singular at $\xi = 0$, but still satisfies the bounds

$$\|Hf\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in L^p(\mathbb{R}^d)$ 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,

$$\| \mathbb{I}_I \|_{M^p(\mathbb{R}^d)}, \| \mathbb{I}_I \|_{M^p(\mathbb{R}^d)} \lesssim_p 1.$$

Now suppose $m \in L^\infty(\mathbb{R}^d)$ has *bounded variation*, which means the quantity

$$V(m) = \sup_{\xi_1 < \cdots < \xi_N} \sum_{i=1}^{N-1} |m(\xi_{i+1}) - m(\xi_i)|.$$

is finite. Then m has countably many discontinuities, and the variation upper bound prevents m from being too nonsmooth. In this case, we can obtain an operator norm bound.

Theorem 24.5. *For any symbol m on $\mathbb{R}$, and any $1 < p < \infty$,*

$$\|m\|_{M^p(\mathbb{R})} \lesssim_p \|m\|_{L^\infty(\mathbb{R})} + V(m).$$

Proof. For each n, pick $\xi_1, \ldots, \xi_{N_n}$ such that

$$\sum_{i=1}^{N-1} |m(\xi_{i+1}) - m(\xi_i)| \geqslant V(m) - 1/n.$$

If we define

$$m_n = m(\xi_1)\, \mathbb{I}_{(-\infty,\xi_1)} + \sum_{i=1}^{N_n-1} m(\xi_i)\, \mathbb{I}_{(\xi_i,\xi_{i+1})} + m(\xi_N)\, \mathbb{I}_{(\xi_N,\infty)}$$

Then $\|m - m_n\|_{L^1(\mathbb{R})} \leqslant 1/n$. But this means that $m_n(D)$ converges to $m(D)$ weakly as operators on $\mathcal{S}(\mathbb{R})$, and so

$$\|m\|_{M^p(\mathbb{R})} \leqslant \limsup_{n \to \infty} \|m_n\|_{M^p(\mathbb{R})}.$$

Now we can rewrite

$$m_n(\xi) = m(\xi_1)\, \mathbb{I}_{(-\infty,\xi_1)} + \sum_{i=1}^{N-1} [m(\xi_i) - m(\xi_{i+1})]\, \mathbb{I}_{(\xi_1,\xi_i)}(\xi) + m(\xi_N)\, \mathbb{I}_{(\xi_N,\infty)}.$$

Since the multiplier operators corresponding to indicator functions of intervals are uniformly bounded in $M^p(\mathbb{R})$ for $1 < p < \infty$, we conclude that

$$\|m_n\|_{M^p(\mathbb{R})} \lesssim_p |m(\xi_1)| + \sum_{i=1}^{N-1} |m(\xi_i) - m(\xi_{i+1})| + |m(\xi_N)| \leqslant \|m\|_{L^\infty(\mathbb{R})} + V(m).$$

Taking $n \to \infty$ gives $\|m\|_{M^p(\mathbb{R})} \lesssim_p \|m\|_{L^\infty(\mathbb{R})} + V(m)$. $\qquad\square$

478

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 $m \in L^\infty(\mathbb{R}^d)$, 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 $C_{\text{loc}}^\infty(\mathbb{R}^d - \{0\})$, and satisfies

$$|D_\xi^\beta m(\xi)| \lesssim_\beta |\xi|^{-\beta} \tag{24.1}$$

  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 (24.1) for all β with $|\beta| \leq N$.

- For some integer $N > d/2$, m has weak derivatives in $L_{\text{loc}}^2(\mathbb{R}^d)$ up to order N, and satisfies an L^2 average estimate

$$\sup_{R>0} \left(\fint_{|\xi| \leq 2R} |\xi|^{2\beta} |D_\xi^\beta m(\xi)|^2 \, d\xi \right)^{1/2} \lesssim 1.$$

- For some non-zero, smooth radial test function φ, compactly supported away from the origin, define

$$m_t = \varphi \text{Dil}_{1/t} m.$$

Thee condition is that

$$\sup_{t>0} \|m_t\|_{L_\alpha^2} < \infty \tag{24.2}$$

for some $\alpha > d/2$ (not necessarily an integer). The supremum above is the *Besov norm* $\|m\|_{B_\alpha^{2,\infty}}$.

- For some φ as above, and for some $\varepsilon > 0$, if $k_t = \hat{m}_t$, then

$$\sup_{R>0} \int |k_t(x)| \langle x \rangle^\varepsilon \, dx < \infty.$$

479

If the second last point holds, then Parseval's inequality implies that

$$\sup_t \|k_t \langle x \rangle^\alpha\|_{L^2} < \infty,$$

to which the last point holds by Hölder's inequality for any ε smaller than $\alpha - d/2$. This result is closely related to the theory of singular integrals. In particular, under the first two assumptions, the multiplier has a convolution kernel which is a Calderon-Zygmund kernel.

Theorem 24.6. *Consider $m \in L^\infty(\mathbb{R}^d)$ 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 $f \in \mathcal{S}(\mathbb{R}^d)$,

$$\|m(D)f\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}.$$

Moreover, we have a bound $\|m(D)f\|_{L^{1,\infty}(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}$, which, together with the fact that $m \in M^2(\mathbb{R}^d)$ 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

$$\|\nabla K_n\|_{L^1(\mathbb{R}^d)} \lesssim \|K_n\|_{L^1(\mathbb{R}^d)} \lesssim 1.$$

480

This implies that for all $y \in \mathbb{R}^d$, uniformly in n,

$$\int |K_n(x+y) - K_n(x)| \, dx \lesssim |y|.$$

For any $f \in \mathcal{S}(\mathbb{R}^d)$, we have

$$m(D)f = \sum_{n=-\infty}^{\infty} (\mathrm{Dil}_{2^n} m_n)(D),$$

where the sum converges absolutely in $L^p(\mathbb{R}^d)$. 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

$$\|K_n * f\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

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 $\mathrm{Dil}_{1/2^n} K_n$ 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

481

the same centre and twice the width. Then

$$\int_{(Q^*)^c} 2^{nd} |(\mathrm{Dil}_{1/2^n} K_n) * b| \leqslant 2^{nd} \int_{(Q^*)^c} \left| \int_Q K_n(2^n(x-y))b(y)\, dy \right| dx$$

$$= 2^{nd} \int_{(Q^*)^c} \left| \int_Q [K_n(2^n(x-y)) - K_n(2^n x)]b(y)\, dy \right| dx$$

$$\leqslant 2^{nd} \int_Q \int_{(Q^*)^c} |b(y)||K_n(2^n(x-y)) - K_n(2^n x)|\, dx\, dy$$

$$= 2^n \int_Q |b(y)||y|\, dy$$

$$\lesssim 2^n R \cdot \|b\|_{L^1(\mathbb{R}^d)}.$$

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

$$\int_{(Q^*)^c} 2^{nd} |(\mathrm{Dil}_{1/2^n} K_n) * b| \leqslant 2^{nd} \int_{(Q^*)^c} \int_Q |K_n(2^n(x-y))||b(y)|\, dy\, dx$$

$$\leqslant 2^{nd} \int_{|z| \geqslant R} \int_Q |K_n(2^n z)||b(y)|\, dy\, dz$$

$$\leqslant (1 + R/2^n)^{-\varepsilon} \|b\|_{L^1(\mathbb{R}^d)}$$

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

$$\int_{(Q^*)^c} |K * b| \lesssim_\varepsilon \|b\|_{L^1(\mathbb{R}^d)}.$$

Now given a general $f \in \mathcal{S}(\mathbb{R}^d)$, we apply a Calderon-Zygmund decomposition, fixing $\alpha > 0$, and writing

$$f = g + \sum_{k=1}^{\infty} b_k$$

where $\|g\|_{L^1(\mathbb{R}^d)} + \sum_{k=1}^{\infty} \|b_k\|_{L^1(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}$, $\|g\|_{L^\infty(\mathbb{R}^d)} \lesssim \alpha$, and there exists a family of disjoint cubes $\{Q_k\}$ such that b_k is supported on Q_k, $\int b_k = 0$, and $\sum_{k=1}^{\infty} |Q_k| \lesssim \alpha^{-1} \|f\|_{L^1(\mathbb{R}^d)}$. If $\Omega = \bigcup_{k=1}^{\infty} Q_k$, then

$$\int_{\Omega^c} |K * (\sum_{k=1}^{\infty} b_k)| \lesssim \sum_{k=1}^{\infty} \|b_k\|_{L^1(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}.$$

482

Thus

$$\left|\{x \in \Omega^c : |K * (\sum_{k=1}^{\infty} b_k)| \geqslant \alpha/2\}\right| \lesssim \|f\|_{L^1(\mathbb{R}^d)}/\alpha,$$

and since $|\Omega| \leqslant \|f\|_{L^1(\mathbb{R}^d)}/\alpha$,

$$\left|\{x \in \mathbb{R}^d : |K * (\sum_{k=1}^{\infty} b_k)| \geqslant \alpha/2\}\right| \lesssim \|f\|_{L^1(\mathbb{R}^d)}/\alpha.$$

Since α was arbitrary, we conclude that $\|K * (\sum b_k)\|_{L^{1,\infty}(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}$. Next, we use the fact that K is bounded on L^2 (since m is bounded) to conclude that

$$\|K * g\|_{L^2(\mathbb{R}^d)} \lesssim \|g\|_{L^2(\mathbb{R}^d)} \lesssim \alpha^{1/2} \|g\|_{L^1(\mathbb{R}^d)}^{1/2}$$

and so

$$\left|\{x \in \mathbb{R}^d : |(K * g)(x)| \geqslant \alpha/2\}\right| \lesssim \|K * g\|_{L^2(\mathbb{R}^d)}^2 \alpha^{-2} \lesssim \|g\|_{L^1(\mathbb{R}^d)} \alpha^{-1}.$$

But now we know $\|K * g\|_{L^{1,\infty}(\mathbb{R}^d)} \lesssim \|g\|_{L^1(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}$. We sum and get $\|K * f\|_{L^{1,\infty}(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}$. □

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 $M^p(\mathbb{R}^d)$ 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 $L^1(\mathbb{R}^d)$ to $L^{1,\infty}(\mathbb{R}^d)$, and thus imply the resultant bounds for the limiting convolution kernel K.

Here is a basic application of the result, a kind of *Schauder estimate* for the fractional Laplacian.

Theorem 24.7. *Suppose $(-\Delta)^{k/2} u \in L^p(\mathbb{R}^d)$ for some $1 < p < \infty$. Then $\partial^\alpha u \in L^p$ for any multi-index α with $|\alpha| = k$.*

Proof. The Fourier multiplier operator whose symbol $m(\xi)$ is a constant multiple of $\xi^\alpha/|\xi|^k$ maps $(-\Delta)^{k/2}u$ to $\partial^\alpha u$. Moreover, m is homogeneous of degree zero, and C^∞ away from the origin, and thus satisfies bounds of the form

$$|\partial^\beta m(\xi)| \lesssim |\xi|^{-\beta}$$

for all $\beta > 0$. But this means that the multiplier operator $m(D)$ is bounded from $L^p(\mathbb{R}^d)$ to $L^p(\mathbb{R}^d)$ for all $1 < p < \infty$, and thus we have

$$\|\partial^\alpha u\|_{L^p(\mathbb{R}^d)} \lesssim \|(-\Delta)^{k/2}u\|_{L^p(\mathbb{R}^d)}. \qquad \square$$

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 $M^{p,q}(\mathbb{R}^d)$ with a function lies in $M^{p,q}(\mathbb{R}^d)$. One can obtain such a result, provided the function is integrable.

Theorem 24.8. *If $u \in L^1(\mathbb{R}^d)$, then $\|m * u\|_{M^{p,q}(\mathbb{R}^d)} \leqslant \|m\|_{M^{p,q}(\mathbb{R}^d)}\|u\|_{L^1(\mathbb{R}^d)}$.*

Proof. Let $K = \hat{m}$. If $v = \hat{u}$, 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 $\|\mathrm{Trans}_\xi m\|_{M^{p,q}(\mathbb{R}^d)} = \|m\|_{M^{p,q}(\mathbb{R}^d)}$, the result follows. $\qquad\square$

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 24.9. *If $m \in M^{p,q}(\mathbb{R}^d)$, and $1 \leqslant p \leqslant q \leqslant 2$, $M^{p,q}(\mathbb{R}^d) \subset L^q(\mathbb{R}^d)^*$.*

Proof. TODO $\qquad\square$

On the other hand, if $q > 2$, then $M^{p,q}(\mathbb{R}^d)$ contains distributions of positive order. We will show an explicit examples for large $d \geqslant 4$, and non constructive examples otherwise.

Example. *Recall that the surface measure σ on the sphere in $\mathbb{R}^d$ satisfies*

$$|\hat{\sigma}(\xi)| \lesssim_d |\xi|^{-\frac{d-1}{2}}.$$

Consider the multiplier corresponding to the distribution $\Lambda = \partial\sigma/\partial r$, which is not a distribution of order zero. On the other hand, $|\hat{\Lambda}(\xi)| \lesssim_d |\xi|^{-\frac{d-3}{2}}$. Provided $d \geqslant 4$, this implies that $\|\hat{\Lambda}\|_{L^q(\mathbb{R}^d)} < \infty$ for $(d-3)/2q > d$, i.e. for $q > 2d/(d-3) = 2 + 6/(d-3)$, and thus $\|\Lambda\|_{M^{1,q}(\mathbb{R}^d)} < \infty$ for this range. For each $\varepsilon > 0$, if we take d appropriately large, apply duality, and iterpolate, for each $\varepsilon > 0$, we obtain a distribution Λ of positive order, which lies in $M^{p,q}(\mathbb{R}^d)$ for all (p,q) with $p < 2 - \varepsilon$ and $q > 2 + \varepsilon$.

Example. *Our other, non constructive examples apply Baire category arguments. Fix $\eta \in C_c^\infty(\mathbb{R}^d)$, and for any $\lambda > 0$, let $m_\lambda(\xi) = \eta(\xi)e^{2\pi i\lambda|\xi|^2}$. We claim that if $K_\lambda = \widehat{m_\lambda}$, $\|K_\lambda\|_{L^\infty(\mathbb{R}^d)} \lesssim \lambda^{-d/2}$. This follows from a simple oscillatory integral argument. Also K_λ will be essentially supported on a ball of radius $O(\lambda)$ at the origin, so one can show from this that $\|K_\lambda\|_{L^1(\mathbb{R}^d)} \lesssim \lambda^{d/2}$. Thus $\|m_\lambda\|_{M^{1,\infty}} \lesssim \lambda^{-d/2}$ and $\|m_\lambda\|_{M^{1,1}} = \|m_\lambda\|_{M^{\infty,\infty}} \lesssim \lambda^{d/2}$. Interpolating gives for $q > 2$ and $q \geqslant p$, a bound of the form $\|m_\lambda\|_{M^{p,q}(\mathbb{R}^d)} \lesssim \lambda^{-\varepsilon}$ for some $\varepsilon > 0$. If $M^{p,q}(\mathbb{R}^d)$ solely contained distributions of order zero, if $\chi \in C_c^\infty(\mathbb{R}^d)$ is a bump function equal to one in a neighborhood of the support of η, then the operator $m \mapsto \chi m$ would be a bounded operator from $M^{p,q}(\mathbb{R}^d)$ to $M(\mathbb{R}^d)$ by the closed graph theorem. But this would imply that*

$$\|\eta\|_{L^1(\mathbb{R}^d)} = \|\chi m_\lambda\|_{M(\mathbb{R}^d)} \lesssim \|m_\lambda\|_{M^{1,q}(\mathbb{R}^d)} \lesssim \lambda^{-\varepsilon},$$

which certainly cannot be true. Thus $M^{p,q}(\mathbb{R}^d)$ contains distributions of positive order for $q > 2$ and $p < 2$.

De Leeuw's theorem shows slices of continuous $d + 1$ dimensional multipliers are bounded by the original mutiplier, so that we might view multipliers in higher dimensions as more complicated than lower dimensional multipliers.

Theorem 24.10. *Let $m \in C(\mathbb{R}^{d+1})$. Fix $\xi_0 \in \mathbb{R}$, and define $m_0 \in C(\mathbb{R}^d)$ by setting*

$$m_0(\xi) = m(\xi, \xi_0).$$

Then for any $1 \leqslant p \leqslant \infty$, $\|m_0\|_{M^p(\mathbb{R}^d)} \leqslant \|m\|_{M^p(\mathbb{R}^{d+1})}$.

Proof. Without loss of generality, assume $\xi_0 = 0$. For $\lambda > 0$ set

$$L(\xi_1, \ldots, \xi_{d+1}) = (\xi_1, \ldots, \xi_d, \xi_{d+1}/\lambda).$$

Then

$$\|m \circ L_\lambda\|_{M^p(\mathbb{R}^{d+1})} = \|m\|_{M^p(\mathbb{R}^{d+1})}.$$

Take $\lambda \to \infty$. Since m is continuous, $m \circ L_\lambda$ converges to $m \circ L_\infty$ pointwise, where $L_\infty(\xi_1, \ldots, \xi_{d+1}) = (\xi_1, \ldots, \xi_d, 0)$. On the other hand,

$$\|m \circ L_\infty\|_{M^p(\mathbb{R}^d)} = \|m_0\|_{M^p(\mathbb{R}^d)}.$$

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

$$\|m \circ L_\infty\|_{M^p(\mathbb{R}^d)} \leqslant \limsup_{\lambda \to \infty} \|m \circ L_\lambda\|_{M^p(\mathbb{R}^d)}.$$

But this follows from a simple weak convergence argument; for any $f, g \in \mathcal{S}(\mathbb{R}^{d+1})$, the dominated convergence theorem implies that

$$\lim_{\lambda \to \infty} |\langle (m \circ L_\lambda)(D)f, g \rangle| = \langle (m \circ L_\infty)(D)f, g \rangle. \qquad \square$$

Example. *If the cone multiplier*

$$m_\lambda(\xi, \eta) = \left(1 - \frac{|\xi|^2}{|\eta|^2} \right)^\lambda_+$$

lies in $M^p(\mathbb{R}^{d+1})$, then the corresponding Bochner-Riesz multiplier

$$m_\lambda(\xi) = \left(1 - |\xi|^2 \right)^\lambda_+$$

lies in $M^p(\mathbb{R}^d)$.

If m is no longer continuous, this theorem only holds for *almost all* slices of the function.

Theorem 24.11. *Fix $m \in M^p(\mathbb{R}^{d+1})$, and for each λ, let $m_\lambda(\xi) = m(\xi, \lambda)$ be defined on $\mathbb{R}^d$. Then for almost every $\lambda \in \mathbb{R}$, $\|m_\lambda\|_{M^p(\mathbb{R}^d)} \leqslant \|m\|_{M^p(\mathbb{R}^{d+1})}$.*

Proof. By duality and the multiplication formula for the Fourier transform,

$$\left| \int m(\xi) \widehat{f}(\xi) \widehat{g}(\xi) \, d\xi \right| \leqslant \|m\|_{M^p(\mathbb{R}^{d+1})} \|f\|_{L^p(\mathbb{R}^{d+1})} \|g\|_{L^q(\mathbb{R}^{d+1})},$$

486

where q is the dual of p. If $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ then

$$\left| \int m_\lambda(\xi) \widehat{f_1}(\xi) \widehat{f_2}(\lambda) \widehat{g_1}(\xi) \widehat{g_2}(\lambda) \, d\xi \, d\lambda \right| \leqslant \|m\|_{M^p(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^q(\mathbb{R}^d)}.$$

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

$$\|J\|_{L^\infty(\mathbb{R})} \leqslant \|J\|_{M^p(\mathbb{R})} \leqslant \|m\|_{M^p(\mathbb{R}^{d+1})} \|f_1\|_{L^p(\mathbb{R}^d)} \|g_1\|_{L^p(\mathbb{R}^d)}.$$

But this implies precisely that for almost every $\lambda \in \mathbb{R}$,

$$|J(\lambda)| \leqslant \|f_1\|_{L^p(\mathbb{R}^d)} \|g_1\|_{L^p(\mathbb{R}^d)},$$

which completes the proof (modulo a separability argument). $\qquad\square$

24.3   Transference

One can view the last result of the last section as a *transference theorem*, showing that the study of the boundedness of multipliers becomes harder in higher dimensions (since a higher dimensional result implies a corresponding lower dimensional result for the slices). Our goal now is to analyze a similar correspondence between the theory of multipliers on $\mathbb{R}^d$ and the theory of multipliers on $\mathbb{T}^d$.

Let us briefly describe the analogous properties of multipliers on $\mathbb{T}^d$ that we have proved for $\mathbb{R}^d$. There exists a theory of Schwartz and tempered sequences $\{a_m : m \in \mathbb{Z}^d\}$ analogous to that of Schwartz functions and tempered distributions on $\mathbb{T}^d$ and $\mathbb{R}^d$, such that the Fourier transform induces an isomorphism between Schwartz functions on $\mathbb{T}^d$ and Schwartz sequences on $\mathbb{Z}^d$, and tempered distributions on $\mathbb{T}^d$ with tempered sequences on $\mathbb{Z}^d$. One can take the convolution in both spaces. If $T : C^\infty(\mathbb{T}^d) \to C^\infty(\mathbb{T}^d)$ is any continuous operator commuting with translations, then there exists a distribution Λ on $\mathbb{T}^d$ such that $Tf = \Lambda * f$ for any

$f \in C^\infty(\mathbb{T}^d)$. Applying the Fourier transform, for any such operator there is also a tempered sequence $\{a_m : m \in \mathbb{Z}^d\}$ such that

$$\widehat{Tf}(n) = a_n \hat{f}(n),$$

so we have a theory of Fourier multipliers on $\mathbb{T}^d$. For any $1 \leqslant p, q \leqslant \infty$, we let $M^{p,q}(\mathbb{Z}^d)$ denote the family of all tempered sequences $\{a_m\}$ such that the induced Fourier multiplier operator T is bounded from the $L^p(\mathbb{T}^d)$ norm to the $L^q(\mathbb{T}^d)$ norm on smooth functions, with the associated operator norm giving the space a topology. Similarily, we define $M^p(\mathbb{Z}^d) = M^{p,q}(\mathbb{Z}^d)$. We have $M^{p,q}(\mathbb{Z}^d) = M^{q^*,p^*}(\mathbb{Z}^d)$ for all $1 \leqslant p, q \leqslant \infty$. As for multipliers on $\mathbb{R}^n$, we have $M^2(\mathbb{Z}^d) = l^\infty(\mathbb{Z}^d)$, and $M^{1,q}(\mathbb{Z}^d) = M^{q^*,\infty}(\mathbb{Z}^d) = \widehat{L^q(\mathbb{T}^d)}$ for $q > 1$, and $M^{1,1}(\mathbb{Z}^d) = M^{\infty,\infty}(\mathbb{Z}^d) = \widehat{M(\mathbb{T}^d)}$.

Our goal is to relate the boundedness of multipliers on $\mathbb{T}^d$ to the boundedness of associated multipliers on $\mathbb{R}^d$, and vice versa. To begin with, we show that bounded multipliers on $\mathbb{R}^d$ induce bounded multipliers on $\mathbb{T}^d$.

Theorem 24.12. *Suppose $m \in M^p(\mathbb{R}^d)$, and every point in $\mathbb{R}^d$ is a Lebesgue point of m. Then for all $R > 0$, the sequence $a_R(n) = m(n/R)$ lies in $M^p(\mathbb{Z}^d)$, and $\|a_R\|_{M^p(\mathbb{Z}^d)} \leqslant \|m\|_{M^p(\mathbb{R}^d)}$.*

Remark. Every function can be altered on a set of measure zero in order to satisfy the assumptions of this theorem, as the theory of the Hardy-Littlewood maximal function indicates.

Proof. Without loss of generality, by rescaling we may assume that $R = 1$. The case $p = 1$ and $p = \infty$ follows from Poisson sumation and the characterizations of $M^1(\mathbb{R}^d)$ and $M^1(\mathbb{Z}^d)$. Thus we may assume $1 < p < \infty$.

Here the proof follows from a general identity, that for any pair of trigonometric periodic polynomials f and g, if $\Phi_\delta(x) = e^{-\pi\delta|x|^2}$, if T is a Fourier multiplier on $\mathbb{R}^d$ with symbol $m(\xi)$, and S is a Fourier multiplier on $\mathbb{Z}^d$ with symbol $m(n)$, then for any $\alpha, \beta > 0$ with $\alpha + \beta = 1$,

$$\lim_{\delta \to 0} \delta^{d/2} \int_{\mathbb{R}^d} T\{f\Phi_{\alpha\delta}\} \cdot \overline{g\Phi_{\beta\delta}} = \int_{\mathbb{T}^d} Sf \cdot \overline{g} \, dx.$$

To obtain the identity, it suffices to assume by linearity that $f(x) = e^{2\pi i n_1 \cdot x}$ and $g(x) = e^{2\pi i n_2 \cdot x}$ for two integers n_1 and n_2. If $n_1 = n_2 = n$, we therefore have to prove that the left hand side is equal to $m(n)$, and if $n_1 \neq n_2$, we have to prove the left hand side is equal to zero. By the multiplication formula, the left hand side is

$$\lim_{\delta \to 0} \delta^{d/2} (\delta\alpha)^{-d/2} (\delta\beta)^{-d/2} \int_{\mathbb{R}^d} m(\xi) \cdot \Phi_{1/\alpha\delta}(\xi - n_1) \Phi_{1/\beta\delta}(\xi - n_2) = \lim_{\delta \to 0} (\alpha\beta\delta)^{-d/2} \int_{\mathbb{R}^d} m(\xi) \Phi_{1/\alpha\delta}(\xi - n_1) \Phi_{1/\beta\delta}(\xi -$$

488

If $n_1 = n_2 = n$, then

$$\Phi_{1/\alpha\delta}(\xi - n_1)\Phi_{1/\beta\delta}(\xi - n_2) = (\alpha\beta\delta)^{-n/2}\Phi_{1/\alpha\beta\delta}(\xi - n),$$

is an approximation to the identity as $\delta \to 0$, yielding the result. On the other hand, if $n_1 \neq n_2$, then the two Gaussian functions are primarily supported on disjoint sets, which yields the result in this case as well.

Returning to the general proof, we note that it suffices to prove by density and duality that for any trigonometric polynomials f and g,

$$\left| \int S f(x)\overline{g(x)}\, dx \right| \leqslant \|m\|_{M^p(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^{p*}(\mathbb{R}^d)}.$$

But the identity above gives that

$$\begin{aligned}
\left| \int S f(x)\overline{g(x)}\, dx \right| &\leqslant \lim_{\delta \to 0} \delta^{d/2} \left| \int \int_{\mathbb{R}^d} T\{f\Phi_{\alpha\delta}\} \cdot \overline{g\Phi_{\beta\delta}} \right| \\
&\leqslant \lim_{\delta \to 0} \delta^{d/2} \|m\|_{M^p(\mathbb{R}^d)} \|f\Phi_{\alpha\delta}\|_{L^p(\mathbb{R}^d)} \|g\Phi_{\beta\delta}\|_{L^{p*}(\mathbb{R}^d)} \\
&= \|m\|_{M^p(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{T}^d)} \|g\|_{L^{p*}(\mathbb{T}^d)}. \qquad \square
\end{aligned}$$

It is clearly not true that if $\{m(n) : n \in \mathbb{Z}^d\}$ gives an element of $M^p(\mathbb{Z}^d)$, then m is in $M^p(\mathbb{T}^d)$, since m can be poorly behaved away from the integers. But if $\{m(n/R) : n \in \mathbb{Z}^d\}$ is uniformly in $M^p(\mathbb{Z}^d)$, we can prove that m is in $M^p(\mathbb{T}^d)$ given weak continuity assumptions.

Theorem 24.13. *Suppose m is Riemann integrable, $1 \leqslant p \leqslant \infty$, and we define $a_R(n) = m(n/R)$. Then*

$$\|m\|_{M^p(\mathbb{R}^d)} \leqslant \sup_{R > 0} \|a_R\|_{M^p(\mathbb{Z}^d)}.$$

Proof. Fix $f, g \in C_c^\infty(\mathbb{R}^d)$. Then for suitably large R, $\mathrm{Dil}_{1/R} f$ and $\mathrm{Dil}_{1/R} g$ are supported in $[-1/2, 1/2]^d$. Define the periodic functions

$$\tilde{f}_R(x) = \sum_{n \in \mathbb{Z}^d} \mathrm{Dil}_{1/R} f(x - n) \quad \text{and} \quad \tilde{g}_R(x) = \sum_{n \in \mathbb{Z}^d} \mathrm{Dil}_{1/R} g(x - n).$$

Then

$$\int m(D)f(x)\overline{g}(x) = \int m(\xi)\widehat{f}(\xi)\overline{\widehat{g}(\xi)}$$

$$= \lim_{R\to\infty} R^{-d} \sum_{n\in\mathbb{Z}^d} b(n/R)\widehat{f}(n/R)\overline{\widehat{g}(n/R)}$$

$$= \lim_{R\to\infty} R^d \sum_{n\in\mathbb{Z}^d} b(n/R)\widehat{\tilde{f}_R}(n)\overline{\widehat{\tilde{g}_R}(n)}$$

$$\leqslant \lim_{R\to\infty} R^d \|a_R\|_{M^p(\mathbb{Z}^d)} \|\tilde{f}_R\|_{L^p(\mathbb{T}^d)} \|\tilde{g}_R\|_{L^{p*}(\mathbb{T}^d)}$$

$$\leqslant \lim_{R\to\infty} \|a_R\|_{M^p(\mathbb{Z}^d)} \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^{p*}(\mathbb{R}^d)}.$$

$\square$

Example. *The Hilbert transform H is a Fourier multiplier with symbol $i\,\mathrm{sgn}(\xi)$. Since it is homogeneous, Riemann integrable, and every point is a Lebesgue point, the theory of boundedness from $L^p(\mathbb{R}^d)$ to $L^p(\mathbb{R}^d)$ is equivalent to the boundedness of the periodic version of the Hilbert transform from $L^p(\mathbb{T}^d)$ to $L^p(\mathbb{T}^d)$.*

Similarily, we see that the convergence in $L^p(\mathbb{T}^d)$ of various Fourier series (Dirichlet sums, spherical summation, square summation, and so on) are equivalent to the associated problems of the convergence of the Fourier transform in $L^p(\mathbb{R}^d)$.

Many problems in harmonic analysis deal not only with the boundedness of Fourier multipliers, but also the boundedness of a maximal function associated with a family of multipliers (e.g. the almost everywhere convergence of Fourier series is connected with the properties of the maximal function $Mf = \sup_{R>0} S_R f$, where S_R is the partial summation operator. Let us proof a transference result for maximal functions associated with dilations of multiplier operators.

Theorem 24.14. *Fix a bounded, Riemann integrable function m on $\mathbb{R}^d$ such that every $\xi \in \mathbb{R}^d$ is a Lebesgue point of m. Define the maximal functions $Mf = \sup_{R>0} m(D/R)f$ on $\mathbb{R}^d$, and the maximal function $Nf = \sup_{R>0} T_R f$, where T_R is the Fourier multiplier operator on $\mathbb{T}^d$ associated with the sequence $a_R(n) = m(n/R)$. Then M is bounded from the $L^p(\mathbb{R}^d)$ norm to itself if and only if N is bounded from the $L^p(\mathbb{R}^d)$ norm to itself.*

Proof. It suffices to show that a uniform family of estimates of the form

$$\|\sum_{i=1}^{K} c_i m(D/R_i)f\|_{L^p(\mathbb{R}^d)} \lesssim \sum |c_i|$$

490

uniform in K and $R_1, \ldots, R_K$, is equivalent to a uniform family of estimates of the form

$$\|\sum_{i=1}^{K} c_i T_{R_i} f\|_{L^p(\mathbb{R}^d)} \lesssim \sum |c_i|,$$

which is uniform in K and $R_1, \ldots, R_K$. In the first case, we may assume without loss of generality that $f \in C_c^\infty(\mathbb{R}^d)$, and in the second, that f is a trigonometric polynomial. The proof then follows by essentially the analogous techniques to above. $\qquad \square$

Remark. One can also show that transference holds between maximal functions bounded merely from $L^p(\mathbb{R}^d)$ to $L^{p,\infty}(\mathbb{R}^d)$, using the duality of Lorentz spaces in much the same way.

Here are some other examples of transference.

Example. *TODO: ELABORATE ON THIS. For a given family of coefficients $\{a_m : m \in \mathbb{Z}^d\}$, consider the periodization operator*

$$Tf = \sum_{m \in \mathbb{Z}^d} a_m Trans_m f.$$

Then, roughly speaking, T is a Fourier multiplier operator with symbol $m(\xi) = \sum a_m e^{-2\pi i \xi \cdot m}$, where this sum must be interpreted in a non-standard sense if the sum $\{a_m\}$ does not converge, which means we might end up obtaining a tempered distribution m rather than a function. But certainly T is well defined for any compactly supported input, since then the sum is locally finite. Let us assume T is bounded on these inputs in the L^p norm for some $1 < p < \infty$ with operator norm A, and thus extends uniquely to a bounded operator from $L^p(\mathbb{R}^n)$ to itself. If we consider any family of coefficients $\{c_m : m \in \mathbb{Z}^d\}$, and consider, for $\varepsilon < 1$, the function

$$f(x) = \sum_{m \in \mathbb{Z}^d} c_m \varepsilon^{-d/p} \mathbb{I}\left(x \in \prod_{i=1}^{d} [m_i, m_i + \varepsilon]\right),$$

then $\|f\|_{L^p(\mathbb{R}^d)} = \|c\|_{l^p(\mathbb{Z}^d)}$. If $\|c_m\|_{l^p(\mathbb{Z}^d)}$, it follows by continuity that it must be true that

$$Tf(x) = \sum_{k \in \mathbb{Z}^d} \sum_{m+n=k} c_m a_n \varepsilon^{-d/p} \mathbb{I}\left(x \in \prod_{i=1}^{d} [k_i, k_i + \varepsilon]\right)$$

$$= \sum_{k \in \mathbb{Z}^d} (a * c)(k) \varepsilon^{-d/p} \mathbb{I}\left(x \in \prod_{i=1}^{d} [k_i, k_i + \varepsilon]\right).$$

It follows that

$$\|Tf\|_{L^p(\mathbb{R}^d)} = \|a * c\|_{l^p(\mathbb{Z}^d)}.$$

Thus we conclude that $\|a * c\|_{l^p(\mathbb{Z}^d)} \leqslant A\|c\|_{l^p(\mathbb{Z}^d)}$. Conversely, suppose convolution with a is bounded from $l^p(\mathbb{Z}^d)$ to $l^p(\mathbb{Z}^d)$ with norm A. Then

$$\left(\sum_{k\in\mathbb{Z}^d} |Tf(x+k)|^p\right)^{1/p} = \left(\sum_{k\in\mathbb{Z}^d} \left|\sum_m a_m f(x+k-m)\right|^p\right)^{1/p} \leqslant A\left(\sum_k |f(x+k)|^p\right)^{1/p}.$$

But

$$\begin{aligned}
\|Tf\|_{L^p(\mathbb{R}^d)} &= \left(\int_{[0,1]^d}\left(\sum_{k\in\mathbb{Z}^d} |Tf(x+k)|^p\right)\right)^{1/p} \\
&\leqslant A\left(\int_{[0,1]^d} \sum_k |f(x+k)|^p\right)^{1/p} \\
&= A\|f\|_{L^p(\mathbb{R}^d)}.
\end{aligned}$$

Thus we conclude that T is bounded if and only if the associated discrete convolution operator is bounded. Thus we have shown studying the boundedness of multiplier operators associated to symbols with domain $\mathbb{T}^d$, e.g. convolution operators from $l^p(\mathbb{Z}^d)$ to $l^q(\mathbb{Z}^d)$, via Fourier multiplier operators with periodic symbols in the domain $\mathbb{R}^d$. More precisely, we can define $M^{p,q}(\mathbb{Z}^d)$ to be the set of all distributions f on $\mathbb{T}^d$ such that convolution with $\hat{f}$ gives a bound from $L^p(\mathbb{Z}^d)$ to $L^q(\mathbb{Z}^d)$ in a suitable domain so we can interpret this convolution, e.g. for compactly supported sequences. It then follows that $M^p(\mathbb{Z}^d)$ is isomorphic to the family of periodic symbols in $M^p(\mathbb{R}^d)$.

Example. *Next, consider $m \in M^p(\mathbb{R}^d)$ with compact support in $[0,1]^d$, and consider the periodization*

$$m_P(\xi) = \sum_{n\in\mathbb{Z}^d} m(\xi - n).$$

We claim that $m_P \in M^p(\mathbb{R}^d)$. TODO (Look up Jodeit Paper?).

Chapter 25

Sobolev Spaces

Chapter 26

Paraproducts

The *paradifferential calculus* is a technique to study the Sobolev behaviour of a product $f_1 f_2$ by breaking up the product into a sum of *paraproducts*, i.e. the part of a pointwise product where we have restricted the frequencies interacting in the product. A basic example occurs in the analysis of the derivatives

$$\nabla(fg) = (\nabla f)g + f(\nabla g).$$

If we could 'invert' the derivative operator, we could write

$$fg = \nabla^{-1}\{(\nabla f)g\} + \nabla^{-1}\{f(\nabla g)\}.$$

A rigorous interpretation of the terms on the right hand side will be provided by *paraproducts*. The behaviour of the function $\nabla^{-1}\{(\nabla f)g\}$ will largely be determined by the interactions of the large frequency components of f (which dominate when taking derivatives) with g, and the behaviour of the function $\nabla^{-1}\{f(\nabla g)\}$ will be determined by the interactions of f with the large frequency components of g. One utility of the theory will allow us to find rigorous heuristics that justify that the 'highest order derivatives dominate' when determining the behaviour of some high order derivative $\nabla^k\{fg\}$, i.e. justifying that $\nabla^k\{fg\} \sim \nabla^k\{f\}g + f\nabla^k\{g\}$, as well as more general variants when ∇^k is replaced with a general pseudodifferential operator.

To start our study, for simplicity we begin with the 'constant coefficient case', i.e. we study translation-invariant multi-linear operators T. Such operators have a Fourier multiplier representation, i.e. there exists a multiplier m such that

$$T\{f_1, \ldots, f_k\}(x) = \int m(\xi)\widehat{f_1}(\xi_1) \ldots \widehat{f_k}(\xi_k)e^{2\pi i(\xi_1 + \cdots + \xi_k)\cdot x} \, d\xi.$$

An analogue of the Hörmander-Mikhlin multiplier theorem holds, though is not always as applicable as desired. A *Coifman-Mayer symbol* is a function m : $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ such that

$$|(\nabla_\xi^\alpha m)(\xi)| \lesssim |\xi|^{-|\alpha|}$$

for all α.

We begin with a basic argument which indicates the need for a general theory.

Theorem 26.1. *For functions f and g,*

$$\|fg\|_{W^{1,1.5}(\mathbb{R}^3)} \lesssim \|f\|_{W^{1,2}(\mathbb{R}^3)} \|g\|_{W^{1,2}(\mathbb{R}^2)}.$$

Proof. A density argument allows us to assume that f and g are both Schwartz functions. By homogeneity symmetry, we may assume $\|f\|_{W^{1,2}(\mathbb{R}^3)} = \|g\|_{W^{1,2}(\mathbb{R}^2)} = 1$. Then we know that

$$\|f\|_{L^2(\mathbb{R}^3)}, \|\nabla f\|_{L^2(\mathbb{R}^3)}, \|g\|_{L^2(\mathbb{R}^3)}, \|\nabla g\|_{L^2(\mathbb{R}^3)} \lesssim 1.$$

Sobolev embeddings also imply that

$$\|f\|_{L^6(\mathbb{R}^3)}, \|g\|_{L^6(\mathbb{R}^3)} \lesssim 1.$$

We begin by making the reduction

$$\begin{aligned}
\|fg\|_{W^{1,1.5}(\mathbb{R}^3)} &\lesssim \|fg\|_{L^{1.5}(\mathbb{R}^3)} + \|\nabla\{fg\}\|_{L^{1.5}(\mathbb{R}^3)} \\
&\leqslant \|fg\|_{L^{1.5}(\mathbb{R}^3)} + \|(\nabla f)g\|_{L^{1.5}(\mathbb{R}^3)} + \|f(\nabla g)\|_{L^{1.5}(\mathbb{R}^3)}.
\end{aligned}$$

We now use Hölder's inequality in three separate ways:

- $\|fg\|_{L^{1.5}(\mathbb{R}^3)} \lesssim 1$, since $f \in L^2(\mathbb{R}^3)$, and $g \in L^6(\mathbb{R}^3)$.

- $\|(\nabla f)g\|_{L^{1.5}(\mathbb{R}^3)}$, since $\nabla f \in L^2(\mathbb{R}^3)$ and $g \in L^6(\mathbb{R}^3)$.

- $\|f(\nabla g)\|_{L^{1.5}(\mathbb{R}^3)}$, since $f \in L^6(\mathbb{R}^3)$ and $\nabla g \in L^2(\mathbb{R}^3)$.

Combining these bounds completes the proof. $\square$

As an example of this result, let f and g be bump functions adapted to balls of radius $1/N$ and $1/M$, for $N \geqslant M \geqslant 1$. Then the Fourier support of f is concentrated on frequencies of magnitude $\sim N$, and the Fourier support of g is concentrated on frequencies of magnitude $\sim M$. Then

$$\|f\|_{L^2(\mathbb{R}^3)} \sim N^{-3/2} \quad \text{and} \quad \|g\|_{L^2(\mathbb{R}^3)} \sim M^{-3/2},$$

495

and
$$\|\nabla f\|_{L^2(\mathbb{R}^3)} \sim N^{-1/2} \quad \text{and} \quad \|\nabla g\|_{L^2(\mathbb{R}^3)} \sim M^{-1/2}.$$

This is consistent with the heuristic that ∇f behaves like Nf, and ∇g behaves like Mg, and, since $N, M \geqslant 1$, shows the derivatives dominate in the Sobolev norm. In the worst case, where the supports of f and g are nested, fg is a bump function adapted to the ball of radius $1/N$, having Fourier support on frequencies $\sim N$, in which case $\|fg\|_{L^{1.5}(\mathbb{R}^3)} \sim N^{-2}$, $\|\nabla(fg)\|_{L^{1.5}(\mathbb{R}^3)} \sim N^{-1}$, and the inequality above to be proved is of the form
$$N^{-1} \lesssim N^{-1/2} M^{-1/2}.$$

We thus see that the inequality is only tight when $N \sim M$.

Chapter 27

Time Frequency Analysis

In harmonic analysis, it is often useful to study a function $f : \mathbb{R}^d \to \mathbb{C}$ via it's Fourier transform $\widehat{f} : \mathbb{R}^d \to \mathbb{C}$. 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 $\mathbb{R}^d \times \mathbb{R}^d$, 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 $\mathbb{R}^d \times \mathbb{R}^d$ 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 $\{x\} \times \mathbb{R}^d$. Similarily, the phase portrait of a pure plane wave $x \mapsto e^{2\pi i \xi \cdot x}$ is $\mathbb{R}^d \times \{\xi\}$, 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. Similarly, 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

$$\{(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d : (x/\lambda, \lambda\xi) \in S\}$$

  More generally, given a linear transformation T, the phase portrait of $f \circ T^{-1}$ is equal to

$$\{(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d : (T(x), T^{-t}(\xi)) \in S\}.$$

  The rescaling above is a special case. In particular, if $T \in O(d)$, then $f \circ T^{-1}$ has phase portrait

$$\{(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d : (T(x), T(\xi)) \in S\}.$$

- The phase portrait of $\widehat{f}$ is equal to

$$\{(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d : (\xi, -x) \in S\},$$

  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.

27.1   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 27.1. *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 $|x| \geqslant$*

$$(\pi_{I \times J} f)(x) \lesssim_n |I|^{1-n} |J|^{1/2} \|f\|_{L^2(\mathbb{R}^d)} d(x, J)^{-n}$$

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 $\pi_{I \times J} f = K_I * (\psi_J \cdot f)$.

$$\left(\int |K_I(x-y)|^2 |\psi_J(y)|^2 \right)^{1/2} \|f\|_{L^2(\mathbb{R}^d)}$$

$\square$

27.2 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 $n \in \mathbb{Z}$. 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*

$$|A_r f| = \left| \sum r^{|n|} \hat{f}(n) e_n(t) \right| \leqslant \|f\|_1 \sum r^{|n|} = \|f\|_1 \left(1 + \frac{2}{1-r} \right)$$

Thus the map is bounded.

To understand whether the truncations $S_N f$ of f converge to f in the L^p norms, rather than pointwise, we turn to the analysis of an operator which is the core of the divergence issue, known as the *Hilbert transform*. It is a Fourier multiplier operator H associated with the coeficients

$$\lambda_n = \frac{\text{sgn}(n)}{i} = \begin{cases} +1/i & n > 0 \\ 0 & n = 0 \\ -1/i & n < 0 \end{cases}$$

Because

$$[|n| \leqslant N] = \frac{\text{sgn}(n+N) - \text{sgn}(n-N)}{2} + \frac{[n = N] + [n = -N]}{2}$$

we conclude

$$S_n f = \frac{i \left(e_{-n} H(e_n f) - e_n H(e_{-n} f) \right)}{2} + \frac{\hat{f}(n) e_n + \hat{f}(-n) e_{-n}}{2}$$

Since the operators $f \mapsto \hat{f}(n) e_n$ 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 $\widehat{f * g} = \hat{f}\hat{g}$ 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 27.2. *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,

$$\|Tf\|_2^2 = \sum \left|\widehat{Tf}(n)\right|^2 = \sum |\lambda_n|^2 \left|\hat{f}(n)\right|^2$$

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 27.3. *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 $z = e^{it}$,

$$K_r(z) = \sum_{n \in \mathbb{Z}} \frac{\text{sgn}(n)}{i} r^{|n|} z^n = K * P_r$$

Since we know the Hilbert transform is continuous in $L^2(\mathbf{T})$, we can conclude that, in particular, for any C^∞ function f,

$$Hf = \lim_{r \to 1} K * (P_r * f) = \lim_{r \to 1} (K * P_r) * f = \lim_{r \to 1} K_r * f$$

So it suffices to determine the limit of the K_r. We find that

$$\sum_{n=1}^{\infty} \frac{(rz)^n - (r\bar{z})^n}{i} = \frac{r}{i} \left(\frac{1}{\bar{z} - r} - \frac{1}{z - r} \right) = \frac{r}{i} \frac{z - \bar{z}}{|z|^2 - 2r\text{Re}(z) + r^2}$$

$$= \frac{2r \sin(t)}{1 - 2r\cos(t) + r^2} = \frac{4r \sin(t/2) \cos(t/2)}{(1 - r)^2 + 4r \sin^2(t/2)}$$

$$= \cot(t/2) + O\left(\frac{(1-r)^2}{t^3} \right)$$

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

$$\left| \int_{|t| \geqslant \varepsilon} [K_r(t) - \cot(t/2)] f(t) \right| \lesssim (1-r)^2 \|f\|_\infty \int_{|t| \geqslant \varepsilon} \frac{dt}{|t|^3} \lesssim \frac{(1-r)^2 \|f\|_\infty}{\varepsilon^2}$$

$$\left| \int_{|t| < \varepsilon} K_r(t) f(t)\, dt \right| \leqslant \int_0^\varepsilon |K_r(t)| |f(t) - f(-t)|$$

$$\lesssim \int_0^\varepsilon |t K_r(t)| |f'(0)| \lesssim \frac{|f'(0)|}{(1-r)^2} \int_0^\varepsilon t^2 \lesssim \|f'\|_\infty \frac{\varepsilon^3}{(1-r)^2}$$

502

$$\left| \int_{|t|<\varepsilon} \cot(t/2)f(t)\, dt \right| \lesssim \int_0^\varepsilon \frac{|f(t) - f(-t)|}{t} \lesssim \varepsilon f'(0)$$

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,

$$(Hf)(x) = \lim_{r\to 1} \int \cot(t/2)f(x-t)\, dt$$

27.3 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
$A_r(f)(t) =$

503

27.4 Conjugate Fourier Series

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

$$\sum_{n=-\infty}^{\infty} \hat{f}(n)e_n(t) = \text{Re}\left(\hat{f}(0) + 2\sum_{n=1}^{\infty} \hat{f}(n)e_n(t) \right)$$

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

$$\text{Im}\left(\hat{f}(0) + 2\sum_{n=1}^{\infty} \hat{f}(n)e_n(t) \right) = \Re\left(-i \sum_{n=-\infty}^{\infty} \text{sgn}(n)\hat{f}(n)e_n(t) \right)$$

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

Chapter 28

Higher Dimensional Fourier Analysis

The Fourier transform in dimensions greater than one have strange phenomena, which is not so obvious from the perspective given by the one dimensional theory. For instance, consider the Theorem of Marcel Riesz, which says that if we define the Dirichlet summation operator

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

then for $1 < p < 2$, if $f \in L^p(\mathbb{R}^d)$, $S_R f \to f$ in $L^p(\mathbb{R}^d)$. The result follows because the operators $\{S_R\}$ are uniformly bounded, and dilation symmetries show this is equivalent to showing that S_1 is a bounded operator on $L^p(\mathbb{R}^d)$. Tensorization shows that Riesz's theorem continues to work in higher dimensions for the *square summation operators*

$$S_R^{\text{sq}} f(x) = \int_{-R}^{R} \cdots \int_{-R}^{R} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi,$$

i.e. the operator S_1^{sq} is bounded on $L^p(\mathbb{R}^d)$ for $1 < p < \infty$. On the other hand, the *spherical summation operators*

$$S_R f(x) = \int_{|\xi| \leqslant R} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi$$

act very differently than their one dimensional counterpart. In particular, Riesz's theorem completely fails in this situation. The *ball multiplier* $S = S_1$ is *only*

bounded on $L^p(\mathbb{R}^d)$ for $p = 2$ (in which case the bound follows trivially from Plancherel). The unboundedness in $L^p(\mathbb{R}^d)$ for $1 < p < \infty$ follows from a clever result of Fefferman (1971).

Theorem 28.1. *The ball multiplier S is* only *bounded on $L^2(\mathbb{R}^d)$.*

Proof. We rely on a geometric construction, together with the uncertainty principle. Duality implies that it suffices to show it is unbounded on $L^p(\mathbb{R}^d)$ for $p > 2$. Thus we must find functions f which are 'broad', or 'spread out', and show that Sf can have parts which are large on a 'narrow', or 'concentrated' set. Our first observation is that the bad part of the multiplier occurs on the boundary of the unit ball B; the multiplier is C^∞ away from this boundary, so behaves very nicely with functions with Fourier support away from the boundary of the unit ball. Thus, without loss of generality, we might as well try and find functions whose Fourier support is concentrated near ∂B.

For any $R > 0$, we construct a function f_R supported on the annulus given by the equation $1 - 1/10R \leqslant |\xi| \leqslant 1 + 1/R$. To apply the uncertainty principle, we find $\sim R^{d-1}$ rectangles $\{\Theta\}$, which have dimensions R^{-1} in the direction tangent to ∂B, but have dimension $R^{-1/2}$ in the $d - 1$ directions tangent to the boundary of the sphere, and such that Θ has one tenth of it's mass in B, and half it's mass outside of B. We fix a compactly supported density function $\phi \in C_c^\infty(\mathbb{R}^d)$ on the unit ball, and rescale it to density functions $\{\phi_\Theta\}$ supported on Θ with magnitude roughly $R^{(d+1)/2}$ on Θ. For any $x_1, \ldots, x_N$, the uncertainty principle tells us that the function χ_{T_Θ}, with Fourier transform equal to

$$e^{-2\pi i x_\Theta \cdot \xi} \phi_\Theta(\xi),$$

has mass concentrated on the dual rectangle T_Θ centered at x_Θ and with magnitude roughly $R^{(d+1)/2}$ on T_Θ. The action of S on χ_{T_Θ} is to cut the Fourier support of Θ by a tenth. Thus we lose the 90% of the Fourier mass of Θ, but our support is also made ten times thinner in the tangential direction. As a result, $S\chi_{T_\Theta}$ will now have mass concentrated on a tube T_Θ^*, which is ten times longer than T_Θ.

Let T_Θ^+ be the portion of T_Θ^* which lies at the opposite end of T_Θ^* to T_Θ. Our construction now relies on *Kakeya* like phenomena. For any $\varepsilon > 0$, there exists a large R_0 such that for $R \geqslant R_0$, we can pick $\{x_\Theta\}$ such that tubes $\{T_\Theta\}$ are disjoint from one another, but such that the tubes $\{T_\Theta^+\}$ have large overlap, in the sense that

$$\Big| \bigcup_\Theta T_\Theta^+ \Big| \leqslant \varepsilon \Big| \bigcup_\Theta T_\Theta \Big|.$$

506

The supports of $S\chi_\Theta$ thus have lots of overlap on this set, but it is difficult to control the sum $\sum S\chi_\Theta$ since these functions might have different signs where they overlap. To fix this, we define

$$f_R = \sum_\Theta \varepsilon_\Theta \chi_\Theta$$

where $\{\varepsilon_\Theta\}$ are independent $\{-1, +1\}$ valued Bernoulli random variables, because Khintchine's inequality implies that we have the pointwise inequality, for any $1 < p < \infty$, of the form

$$\left(\mathbb{E} \left| \sum_\Theta \varepsilon_\Theta \chi_\Theta \right|^p \right)^{1/p} \sim \left(\sum_\Theta |\chi_\Theta|^2 \right)^{1/2}.$$

It follows that, since the tubes T_Θ are disjoint,

$$\mathbb{E} \|f_R\|^p_{L^p(\mathbb{R}^d)} \sim \int \left(\sum_\Theta |\chi_\Theta(x)|^2 \right)^{p/2} dx$$

$$\sim \left(R^{\frac{d+1}{2}} \right)^p \left| \bigcup_\Theta T_\Theta \right|.$$

Similarily, Khintchine's inequality can again be applied to

$$S f_R = \sum_\Theta \varepsilon_\Theta S \chi_\Theta.$$

to conclude that

$$\mathbb{E} \|S f_R\|^p_{L^p(\mathbb{R}^d)} \sim \int \left(\sum_\Theta |S\chi_\Theta(x)|^2 \right)^{p/2} dx$$

$$\gtrsim \left(R^{\frac{d+1}{2}} \right)^p \int \left(\sum_\Theta \mathbb{I}_{T_\Theta^+}(x) \right)^{p/2} dx.$$

If S was bounded on $L^p(\mathbb{R}^d)$, we would conclude that

$$\int \left(\sum_\Theta \mathbb{I}_{T_\Theta^+}(x) \right)^{p/2} dx \lesssim \left| \bigcup_\Theta T_\Theta \right|.$$

507

The trivial bound

$$\int \left(\sum_{\Theta} \mathbb{I}_{T_{\Theta}^+}(x) \right) \, dx \sim \left| \bigcup_{\Theta} T_{\Theta} \right|$$

combined with the fact that the sum is supported on the finite measure set $\bigcup_{\Theta} T_{\Theta}^+$ yields that

$$\left| \bigcup_{\Theta} T_{\Theta}^+ \right|^{-(p/2-1)} \left| \bigcup_{\Theta} T_{\Theta} \right|^{p/2} \lesssim \int \left(\sum_{\Theta} \mathbb{I}_{T_{\Theta}^+}(x) \right)^{p/2} \, dx.$$

Thus we find that

$$\left| \bigcup_{\Theta} T_{\Theta} \right|^{p/2-1} \lesssim \left| \bigcup_{\Theta} T_{\Theta}^+ \right|^{p/2-1}.$$

Because of the Kakeya like specification of these sets, this is impossible in the range $p > 2$ specified. $\qquad \square$

We thus see that understanding the behaviour of S is closely connected to Kakeya phenomena. Indeed, one can generalize this problem to determining the behaviour of the *Bochner-Riesz multipliers* S_{δ}, which are the Fourier multipliers with symbol

$$m^{\delta}(\xi) = (1 - |\xi|)_+^{\delta}.$$

For $\delta = 0$, S_{δ} is just the ball multiplier above. As δ increases, the singularity on the boundary of the unit ball becomes less and less bad, so it might be expected that S_{δ} becomes bounded on a larger range of L^p spaces. Determining the precise range under which these operators are bounded is a major open problem of harmonic analysis, and is closely connected to a deeper understanding of Kakeya like phenomena.

Chapter 29

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 absolute values, or various spatial decomposition strategies, fail completely to give tight estimates. A typical oscillatory integral is of the form

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

where $a : X \to \mathbb{C}$ is the *amplitude function*, and $\phi : X \to \mathbb{R}$ is the *phase*. The value λ is a positive 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 obtain more precise controls on this asymptotic decay more precisely.

Example. *The most basic example of an oscillatory integral is the Fourier transform, where for each function $f \in L^1(\mathbb{R})$, and each $\lambda \in \mathbb{R}$, we consider the quantity*

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

Thus in the oscillatory integral defining the Fourier transform of f, the function f plays the role of the amplitude, the phase function is $\phi(x) = 2\pi x$. 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. And indeed, in what follows we will often assume the boundedness of a certain number of derivatives of the 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 (see the analysis of the Airy function below for an example of this technique). However, such a method is not in vogue nowadays, since 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 $\psi(x_0)e^{i\lambda\phi(x_0)}$, times the volume of the region around x_0 where ϕ deviates by $\approx 1/\lambda$, to the asymptotic decay of $I(\lambda)$ as $\lambda \to \infty$, provided that we treat $2\pi = 1$ (we mean this facetiously, i.e. because some powers of 2π tend to show up in actual asymptotic formulas for these quantities, but which are usually irrelevant to a given analysis)

29.1 One Dimensional Theory

Let us begin with a simple example of an oscillatory integral, i.e.

$$I(\lambda) = \int_J e^{i\lambda\phi(x)} \, dx,$$

where J is a closed interval, and $\phi : J \to \mathbb{R}$ 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 $\lambda \to \infty$ due to cancellation. 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 29.1 (Esseén Concentration Inequality). *Let $\phi : J \to \mathbb{R}$ be a Borel measurable function, and for each $\lambda \in \mathbb{R}$, set*

$$I(\lambda) = \int_J e^{2\pi i\lambda\phi(x)} \, dx.$$

Then for any $\varepsilon > 0$,

$$\sup_{a\in\mathbb{R}} |\{x \in J : |\phi(x) - a| \leqslant \varepsilon\}| \lesssim \varepsilon \int_0^{1/\varepsilon} |I(\lambda)| \, d\lambda,$$

510

where the implicit constant is independent 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 independent of the function ϕ. If ψ is an integrable function supported on $[0, 1]$, then Fubini's theorem implies

$$\int_0^1 \psi(\lambda) I(\lambda) \, d\lambda = \int_0^1 \int_0^1 \psi(\lambda) e^{i\lambda\phi(x)} \, d\lambda \, dx$$
$$= \int_0^1 \widehat{\psi}(-\phi(x)/2\pi) \, dx.$$

In particular, reversing this chain of identities gives the bound

$$\left| \int_0^1 \widehat{\psi}(-\phi(x)/2\pi) \, dx \right| \leqslant \|\psi\|_{L^\infty[0,1]} \int_0^1 |I(\lambda)| \, d\lambda.$$

If we choose a bounded function ψ such that $\widehat{\psi}$ is non-negative, and bounded below on $[-1, 1]$, 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. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Remark. Recently, I have seen results that the *reverse* of this phenomenon sometimes holds, 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 a decay which is independent of the function ϕ, as the next example shows, if the integrand oscillates at a wavelength $1/2\pi\lambda$.

Example. *Fix a positive integer λ_0, and let $\phi_{\lambda_0}(x) = 2\pi x + f(\lambda_0 x)/\lambda_0$, where $f : \mathbb{R} \to \mathbb{R}$ is a smooth, 1-periodic function, such that $\|f'\|_{L^\infty(\mathbb{R})} \leqslant \pi$, and*

$$\int_0^1 e^{i(x+f(x))}\, dx \neq 0.$$

Then for each $x \in \mathbb{R}$, $1/2 \leqslant |\phi'_{\lambda_0}(x)| \leqslant 2$, and in particular, is bounded independently of λ_0. Since

$$\phi_{\lambda_0}(x + 1/\lambda_0) = \phi_{\lambda_0}(x) + 2\pi/\lambda_0,$$

we find $e^{i\lambda_0\phi_{\lambda_0}(x)}$ is $1/\lambda_0$ periodic. In particular, this means

$$\begin{aligned}
I(\lambda_0) &= \int_0^1 e^{i\lambda_0\phi_{\lambda_0}(x)}\, dx \\
&= \lambda_0 \int_0^{1/\lambda_0} e^{i(\lambda_0 x + f(\lambda_0 x))}\, dx \\
&= \int_0^1 e^{i(x+f(x))}\, dx.
\end{aligned}$$

Thus $I(\lambda_0; \phi_{\lambda_0}) \sim 1$, despite a uniform lower bound on the derivatives of the family of functions $\{\phi_{\lambda_0}\}$.

One option to avoid this counterexample is to place a uniform upper bound on ϕ'', in addition to lower bounding ϕ'. This yields a positive result.

Theorem 29.2. *Let $\phi : J \to \mathbb{R}$ 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

$$\int_0^1 e^{i\phi(x)} \, dx = \int_0^1 \frac{1}{i\phi'(x)} \frac{d}{dx} \left(e^{i\phi(x)} \right) \, dx$$

$$= \left(\frac{e^{i\phi(1)}}{i\phi'(1)} - \frac{e^{i\phi(0)}}{i\phi'(0)} \right) - \int_0^1 \frac{d}{dx} \left(\frac{1}{i\phi'(x)} \right) e^{i\phi(x)}.$$

Now

$$\frac{d}{dx} \left(\frac{1}{\phi'(x)} \right) = -\frac{\phi''(x)}{\phi'(x)^2},$$

so taking in absolute values completes the proof. $\qquad\square$

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 expect to obtain any better decay than $1/\lambda$ for general phase functions satisfying the assumptions of this theorem.

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 the counterexample above. This again leads an important result, called the *Van der Corput Lemma*.

Lemma 29.3 (Van der Corput). *Let $\phi : \mathbb{R} \to \mathbb{R}$ 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.

513

Proof. We may rescale so that $\lambda = 1$ and $J = [0, 1]$. Then the same integration by parts shows that

$$\int_0^1 e^{i\phi(x)} \, dx = \left(\frac{e^{i\phi(b)}}{i\phi'(1)} - \frac{e^{i\phi(0)}}{i\phi'(0)} \right) + \frac{1}{i} \int_0^1 \frac{d}{dx} \left(\frac{1}{i\phi'(x)} \right) e^{i\phi(x)} \, dx.$$

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,

$$\left| \int_J \frac{d}{dx} \left(\frac{1}{i\phi'(x)} \right) e^{i\phi(x)} \, dx \right| \leqslant \int_J \left| \frac{d}{dx} \left(\frac{1}{i\phi'(x)} \right) \right| dx$$
$$= \left| \int_J \frac{d}{dx} \left(\frac{1}{i\phi'(x)} \right) \right| dx$$
$$= \frac{1}{\phi'(b)} - \frac{1}{\phi'(a)}.$$

Again, this term is $O(1/A)$. $\qquad \square$

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 29.4. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be smooth, and suppose there is some $k \geqslant 2$ such that $|\partial^k \phi(x)| \geqslant A$ 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 independent of J.

Proof. We perform an induction on k, the case $k = 1$ already proven. By scale invariance, we may assume $\lambda = 1$. Now the k'th derivative bound implies the function $\partial^{k-1}\phi$ is monotone, so for each $\alpha > 0$, outside an interval of length at most $O(\alpha/A)$, $|\partial^{k-1}\phi(x)| \geqslant \alpha$. 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 $|I(\lambda)| \lesssim_k A^{-1/k}$. $\qquad \square$

Remark. If $\partial^{k-1}\phi$ 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 Esseen concentration 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.

$$I(\lambda) = \int_{-\infty}^{\infty} a(x)e^{i\lambda\phi(x)} \, dx.$$

The Van der Corput lemma also applies here.

Lemma 29.5. *Fix $k \geqslant 1$, and suppose $\phi : \mathbb{R} \to \mathbb{R}$ is a smooth phase function. If a is supported on J, $|\partial^k\phi(x)| \geqslant A$ for all $x \in J$, with the additional assumption that ϕ' is monotone if $k = 1$. Then*

$$|I(\lambda)| \lesssim_k \frac{\|a\|_{L^\infty(\mathbb{R})} + \|a'\|_{L^1(\mathbb{R})}}{(A\lambda)^{1/k}}.$$

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

$$I_0(x) = \int_0^x e^{i\lambda\phi(t)} \, dt.$$

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

$$\int_0^1 a(x)e^{i\lambda\phi(x)} \, dx = \int_0^1 a(x)I_0'(x) \, dx$$
$$= [a(1)I_0(1) - a(0)I_0(0)] - \int_0^1 a'(x)I_0(x) \, dx.$$

Now

$$|a(1)I_0(b) - a(0)I_0(a)| \lesssim \frac{\|a\|_{L^\infty(\mathbb{R})}}{(A\lambda)^{1/k}}$$

515

and

$$\left| \int_0^1 a'(x) I_0(x) \, dx \right| \lesssim_k \frac{\|a'\|_{L^1(\mathbb{R})}}{(A\lambda)^{1/k}}.$$

Putting these two estimates together completes the proof. □

Remark. Since the bound above isn't dependent 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

$$\sup_{a<b} \left| \int_a^b a(x) e^{i\lambda\phi(x)} \, dx \right| \lesssim_k \frac{\|a\|_{L^\infty(\mathbb{R})} + \|a'\|_{L^1(\mathbb{R})}}{(A\lambda)^{1/k}}.$$

By taking a limit in an appropriate sense, one can thus obtain an estimate on an oscillatory integral with an infinite range.

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 $\lambda \to \infty$ under the assumption that ϕ' is nonvanishing. This is the principle of *nonstationary phase*.

Theorem 29.6. *Let J be an interval, and fix $N \geqslant 0$. Consider a real-valued phase function $\phi \in C^{N+1}(\mathbb{R})$, with $\phi'(x) \neq 0$ for all $x \in J$, and an amplitude $a \in C^N(\mathbb{R})$ with $\mathrm{supp}(a) \subset J$. Then*

$$|I(\lambda)| \lesssim_{N,\phi} |J| \cdot \lambda^{-N} \max_{0 \leqslant k \leqslant N} \left\| \frac{\partial^k a}{|\partial\phi|^{2N-k}} \right\|_{L^\infty(J)},$$

where the implicit constant is uniformly bounded over a bounded family of phases ϕ in the Banach space $C^{N+1}(J)$, modulo constant functions.

Proof. We shall prove by induction on N that we can write

$$I(\lambda) = \frac{1}{\lambda^N} \int \left(\sum_k \frac{(\partial^k a)(x)}{(\partial\phi)(x)^k} Q_{N,k}(\partial\phi, \dots, \partial^{N+1}\phi) \right),$$

where $Q_{N,k}$ is a polynomial in the first $N+1$ derivatives of ϕ. This result is clearly true for $N = 0$. Applying integration by parts shows, for a fixed N and k, that

$$\int \frac{(\partial^k a)}{(\partial\phi)(x)^k} Q_{N,k}(\partial\phi, \dots, \partial^{N+1}\phi) e^{2\pi i\lambda\phi(x)} \, dx$$

$$= \frac{1}{i\lambda} \int \frac{(\partial^k a)(x)}{(\partial\phi)(x)^{k+1}} Q_{N,k}(\partial\phi, \dots, \partial^{N+1}\phi) \partial_x e^{i\lambda\phi(x)} \, dx$$

$$= \frac{-1}{i\lambda} \int \partial_x \left\{ \frac{(\partial^k a)(x)}{(\partial\phi)(x)^{k+1}} Q_{N,k}(\partial\phi, \dots, \partial^{N+1}\phi) \right\}.$$

516

Applying the product rule to the derivative gives that this is a sum of terms that fit into the sum above. □

Remark. Differentiation under the integration sign shows

$$\left(\frac{d}{d\lambda}\right)^n I(\lambda) = i^n \int \phi(x)^n a(x) e^{i\lambda\phi(x)} \, dx.$$

Since $\phi^n a$ satisfies the same assumptions that a does, 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

$$I(\lambda) = \int_{-\infty}^{\infty} a(x) e^{i\lambda x^2} \, dx,$$

where $a \in C_c^\infty(\mathbb{R})$. If a is nonzero in a neighborhood of the origin, and if we consider the phase $\phi(x) = x^2$, then

$$|\{x \in \mathbb{R} : |\phi(x)| \leqslant 1/\lambda\}| \lesssim \lambda^{-1/2}.$$

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 will now prove to be true for this oscillatory integral. A dyadic decomposition about the origin, combined with a single integration by parts at each scale gives this decay. However, we can obtain actually obtain a much more precise asymptotics.

Theorem 29.7. *Let $a \in \mathcal{S}(\mathbb{R})$. Then*

$$I(\lambda) \sim \pi^{1/2} e^{i\pi/4} \sum_{k=0}^{\infty} \frac{(i/4)^k}{k!} (\partial^{2k} a)(0) \cdot \lambda^{-k-1/2}$$

We can interpret this formula using the symbol calculus*: the right hand side gives an asymptotic expansion of a symbol of order $-1/2$, and so the result says that I is a symbol of order $-1/2$, and gives an expansion for this symbol. More precisely, we have that*

$$\left(\frac{d}{d\lambda}\right)^M \left\{ I(\lambda) - \pi^{1/2} e^{i\pi/4} \sum_{k<N} \frac{(i/4)^k}{k!} (\partial^{2k} a)(0) \cdot \lambda^{-k-1/2} \right\}$$

517

is $O_{N,M}(\lambda^{-1/2-N-M})$, where the implicit constant is uniformly bounded over a family of amplitudes a if

$$\int |\xi|^{2N} |(\partial_\xi^M \widehat{a})(\xi)| \, d\xi$$

is uniformly bounded. This holds, in particular, if the set of functions $|x|^M \cdot a(x)$ forms a bounded subset of $H^{2N+s}(\mathbb{R})$, for some $s > 1/2$, in particular, if the functions a are bounded in $C_c^{2N+1}(\mathbb{R})$. As a special case, if a equals one in a neighborhood of the origin, then

$$I(\lambda) \sim \pi^{1/2} e^{i\pi/4} \lambda^{-1/2}.$$

Moreover, we conclude that the map $a \mapsto I_a$ is a bounded map from $\mathcal{S}(\mathbb{R})$ to the set $\mathcal{S}_c^{-1/2}(\mathbb{R})$ symbols of order $-1/2$.

Remark. In many problems in analysis, it causes no problem to work with smooth functions in $C^\infty(\mathbb{R})$, modulo the set $\mathcal{S}(\mathbb{R})$ of Schwartz functions. One can then treat this asymptotic formula as uniquely identifying the oscillatory integral, since any two functions with the same asymptotics differ by a Schwartz function.

Proof. Applying the multiplication formula for the Fourier transform, noting that the distributional Fourier transform of $e^{i\lambda x^2}$ is

$$e^{i\pi/4}(\pi/\lambda)^{1/2} e^{-i(\pi^2/\lambda)\xi^2}.$$

Thus the multiplication formula for the Fourier transform tells us that

$$I(\lambda) = e^{i\pi/4}(\pi/\lambda)^{1/2} \int \widehat{a}(\xi) e^{-i(\pi^2/\lambda)\xi^2} \, d\xi.$$

For any $N > 0$, Taylor's theorem tells us that

$$e^{-i(\pi^2/\lambda)\xi^2} = \sum_{k<N} \frac{(-i\pi^2\xi^2)^k}{k!} \lambda^{-k} + O_N\left(\frac{|\xi|^{2N}}{\lambda^N}\right).$$

Thus

$$I(\lambda) = \sum_{k<N} e^{i\pi/4} \frac{(-i)^k \pi^{2k+1/2}}{k!} \lambda^{-k-1/2} \int \xi^{2k} \widehat{a}(\xi) \, d\xi$$

$$+ O_N\left(\lambda^{-N} \int |\xi|^{2N} |\widehat{a}(\xi)| \, d\xi\right)$$

$$= \pi^{1/2} e^{i\pi/4} \sum_{k<N} \frac{(i/4)^k}{k!} (\partial^{2k} a)(0) + R,$$

518

where the remainder term R is bounded as required by the Lemma. One then obtains the asymptotic formula for the derivative of I by noting that

$$\left(\frac{d}{d\lambda}\right)^M I(\lambda) = \int (-2\pi i x^2)^M a(x) e^{-2\pi i \lambda x^2}\, dx,$$

which reduces to the study of the case $M = 0$. $\qquad\qquad\qquad\square$

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 29.8. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a smooth phase with a single, non-degenerate critical point x_0, and let $a : \mathbb{R} \to \mathbb{R}$ 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*

$$I(\lambda) \sim e^{i\lambda\phi(x_0)} \lambda^{-1/2} \sum_{n=0}^{\infty} c_n \lambda^{-n}.$$

The implicit constants in the bounds

$$\left| I(\lambda) - e^{i\lambda\phi(x_0)} \lambda^{-1/2} \sum_{n<N} c_n \lambda^{-n} \right| \lesssim \lambda^{-1/2-N}$$

are uniformly bounded provided that the family of amplitude functions form a bounded subset of $C_c^{2N+1}(\mathbb{R})$, the family of phase functions ϕ are bounded in $C^{2N+3}(\mathbb{R})$, and the functions

$$\frac{|\phi'(x)|}{|x - x_0|}$$

are uniformly bounded from below on the common domains of the amplitude functions. We can explicitly calculate that

$$c_0 = (2\pi)^{1/2} \sqrt{\frac{i}{\phi''(x_0)}} \cdot a(x_0).$$

More generally, there exists an order two differential operator L, with non-constant coefficients depending on the first two derivatives of the phase function ϕ, such that for any $n > 0$,

$$c_n = (2\pi)^{1/2} \sqrt{\frac{i}{\phi''(x_0)}} \frac{(i/2)^n}{n!} L^n a(x_0).$$

519

The proof will show that the best way to obtain these coefficients is often to switch coordinates so that we can use the asymptotic expansion for an oscillatory integral with quadratic phase.

Proof. Translating our integration variables if necessary, we may assume that the critical point x_0 lies at the origin. A coordinate change $x \mapsto -x$ means we may assume that $\phi''(0) > 0$. Partitioning the support of a, applying the principle of nonstationary phase away from the origin, we may assume without loss of generality that $\phi'(x) \neq 0$ for all x on the support of a, and thus the function

$$f(x) = \text{sgn}(x)\phi(x)^{1/2}.$$

is a diffeomorphism on the support of a. By the change of variables formula, there exists a smooth, compactly supported function $a_0(y)$ such that

$$I(\lambda) = \int a(x)e^{i\lambda\phi(x)}\, dx = \int a_0(y)e^{i\lambda y^2}\, dy.$$

We can actually write

$$a_0(y) = 2y \cdot \frac{(a \circ f^{-1})(y)}{(\phi' \circ f^{-1})(y)}.$$

Thus we can apply the previous theorem to conclude that there exists a sequence of constants $\{c_n\}$, depending solely on the derivatives of a_0 at the origin, such that I is a symbol of order $-1/2$, and

$$I(\lambda) \sim e^{i\pi/4}\pi^{1/2} \sum_{k=0}^{\infty} \frac{c_k}{k!}\lambda^{-k-1/2}.$$

For each $k > 0$, we have

$$c_k = (i/4)^k(\partial^{2k}a_0)(0).$$

If we define an operator

$$Lu = \partial\{u \circ f^{-1}\} \circ f,$$

then $(\partial^{2k}a_0)(0) = L^{2k}\{fa(\phi')^{-1}\}(0)$. By the chain rule, we can write

$$(Lu)(y) = \frac{u'(y)}{f'(y)} = 2u'(y)\phi(y)^{1/2}\phi'(y)^{-1}.$$

This is a smooth operator

In particular, for any N, we have

$$|I(\lambda) - \sum_{k=0}^{N} c_k \lambda^{-k-1/2}| \lesssim_N \lambda^{-N-3/2},$$

where the implicit constant depends only on the first $2N + 4$ derivatives of $\{a_0\}$. In particular, for a fixed ϕ, and a fixed support, we conclude the bound depends only on upper bounds of the first $2N + 4$ derivatives of a. Now in particular, we have that, by L'Hopital's rule,

$$\begin{aligned}
a_0(0) &= \lim_{x \to 0} 2f(x) \frac{a(x)}{\phi'(x)} \\
&= 2a(0) \lim_{x \to 0} \frac{f'(x)}{\phi''(x)} \\
&= \frac{a(0)}{\phi''(0)} \lim_{x \to 0} \frac{\phi'(x)}{\phi(x)^{1/2}} \\
&= \frac{a(0)}{\phi''(0)} \lim_{x \to 0} \frac{x\phi''(0) + O(x^2)}{x\phi''(0)^{1/2} + O(x^2)} \\
&= \frac{a(0)}{\phi''(0)^{1/2}}.
\end{aligned}$$

Thus we conclude that

$$\begin{aligned}
c_0 &= (2\pi)^{1/2} e^{i\pi/4} a_0(0) \\
&= (2\pi)^{1/2} \sqrt{i/\phi''(0)} a(0). \qquad \Box
\end{aligned}$$

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*

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

We want to use the method of stationary phase to determine the decay of $J_m(r)$ as $r \to \infty$. We preceed as above, letting

$$a(\theta) = e^{im\theta} \quad \text{and} \quad \phi(\theta) = -\sin\theta.$$

Then the integral has a stationary point at $\theta = -\pi/2$ and $\theta = \pi/2$, and these stationary points are nondegenerate. Thus we have

$$|J_m(r)| = O_m(r^{-1/2}).$$

More precisely, let us write $a = a_1 + a_2 + a_3$, where a_1 is supported in a small neighbourhood of $-\pi/2$, a_2 is supported in a small neighborhood of $\pi/2$, and a_3 is supported away from $-\pi/2$ and $\pi/2$. Then $I_{a_1}(r)$ and $I_{a_2}(r)$ are oscillatory integrals with a unique nondegenerate stationary point. Thus we can write

$$I_{a_1}(r) = (2\pi)^{-1/2} e^{-i\pi/4} e^{-im\pi/2} e^{-ir} r^{-1/2} + O(r^{-3/2})$$

and

$$I_{a_2}(r) = (2\pi)^{-1/2} e^{i\pi/4} e^{im\pi/2} e^{ir} r^{-1/2} + O(r^{-3/2}).$$

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. In particular,

$$|I_{a_3}(\lambda)| = O(\lambda^{-3/2}).$$

Thus summing up these estimates gives

$$J_m(r) = (2/\pi)^{1/2} \cos(r - \pi/4 - m\pi/2) r^{-1/2} + O(r^{-3/2}).$$

One can carry out the analysis above in more detail to yield a complete asymptotic expansion for this function.

Example. *Consider the Airy function*

$$Ai(x) = \int_{-\infty}^{\infty} e^{i(x\xi + \xi^3/3)} \, d\xi,$$

which arises as a solution to the differential equation $y'' = xy$, which originally was studied by English mathematician George Airy in the analysis of the diffraction pattern of the electromagnetic wave across a rainbow. 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 independent of I. Thus we can interpret the integral as

$$\lim_{n \to \infty} \int_{a_n}^{b_n} e^{i(x\xi + \xi^3/3)} \, d\xi,$$

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 $v = x^{-1/2}\xi$, so that for $x > 0$,

$$Ai(-x) = x^{1/2} \int_{-\infty}^{\infty} e^{ix^{3/2}(v^3/3 - v)} \, dv.$$

If we write $\phi_0(v) = v^3/3 - v$, then ϕ_0 has two stationary points, at $v = \pm 1$. 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

$$I_1 = \pi^{1/2} e^{+i[\pi/4 - (2/3)x^{3/2}]} x^{-1/4} + O(x^{-7/4}),$$

$$I_2 = \pi^{1/2} e^{-i[\pi/4 - (2/3)x^{3/2}]} x^{-1/4} + O(x^{-7/4})$$

Moreover, by the principle of nonstationary phase, $I_3 = O_N(x^{-N})$ for all $N > 0$. It remains to show that $I_4 = O(x^{-1})$. Indeed, an integration by parts shows that

$$I_4 = x^{1/2} \int_{-\infty}^{\infty} e^{ix^{3/2}\phi_0(v)} a_4(v) \, dv$$

$$= \frac{-1}{ix} \int_{-\infty}^{\infty} e^{ix^{3/2}\phi_0(v)} \frac{d}{dv}\left(\frac{a_4(v)}{v^2 - 1}\right) \, dv.$$

Taking in absolute values shows $|I_4| \lesssim 1/x$. Thus as $x \to \infty$,

$$Ai(-x) = (4\pi)^{1/2} \cos(\pi/4 - (2/3)x^{3/2}) x^{-1/4} + O(x^{-1})$$

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

523

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 $\mathbb{R} + ix^{1/2}$ gives

$$Ai(x) = \int_{-\infty}^{\infty} e^{i\phi(\xi + ix^{1/2})} \, d\xi = e^{-(2/3)x^{3/2}} \int_{-\infty}^{\infty} e^{-\xi^2 x^{1/2}} e^{i\xi^3/3} \, d\xi.$$

We have

$$\int_{-\infty}^{\infty} e^{-\xi^2 x^{1/2}} e^{i\xi^3/3} \, d\xi \approx x^{-1/4} \int_{-\infty}^{\infty} e^{-\xi^2} e^{ix^{-3/4}\xi^3/3} \, d\xi.$$

Now a Taylor series shows

$$e^{ix^{-3/4}\xi^3/3} = 1 + O(x^{-3/4}\xi^3/3),$$

so, plugging in, we conclude

$$Ai(x) = \pi^{1/2}x^{-1/4}e^{-(2/3)x^{3/2}} + O(x^{-1}e^{-(2/3)x^{3/2}}).$$

Thus Airy's function decreases exponentially as $x \to \infty$.

Example. *Let us consider the integral quantities*

$$\int_0^1 e^{2\pi i x\xi} e^{2\pi i/x} x^{-\gamma} \, dx$$

where to avoid technicalities we assume $0 \leqslant \gamma < 2$. These integral quantities are not defined absolutely, so we actually interpret this integral as

$$\lim_{\varepsilon \to 0} \int_0^1 e^{2\pi i x\xi} e^{2\pi i/x} x^{-\gamma} \, dx$$

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 $\varepsilon \leqslant \xi^{-1/2}/2$,

$$\begin{aligned}
\int_{\varepsilon_1}^{\varepsilon_2} e^{2\pi i \phi(x)} x^{-\gamma} \, dx &= \frac{1}{2\pi i\xi} \int_{\varepsilon_1}^{\varepsilon_2} \frac{d}{dx}\left(e^{2\pi i \phi(x)}\right) \frac{x^{2-\gamma}}{x^2 - 1/\xi} \, dx \\
&= \frac{-1}{2\pi i\xi} \int_{\varepsilon_1}^{\varepsilon_2} e^{2\pi i \phi(x)} \frac{d}{dx}\left(\frac{x^{2-\gamma}}{x^2 - 1/\xi}\right) + O(\varepsilon^{2-\gamma}) \\
&= O(\varepsilon^{2-\gamma}),
\end{aligned} \tag{29.1}$$

524

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 (29.1) 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 $x \geq 1.5\xi^{-1/2}$. 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.

525

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 $f \in \mathcal{S}(\mathbb{R})$, 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.*

29.2 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

$$I(\lambda) = \int_{\mathbb{R}^d} a(x) e^{i\lambda \phi(x)} \, dx.$$

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 29.9. *Fix a compact set K and a positive integer N > 0, and consider an amplitude* $a \in C^N(\mathbb{R}^d)$ *with* $\mathrm{supp}(a) \subset K$, *and a real-valued phase* $\phi \in C^{N+1}(\mathbb{R}^d)$, *with* $\nabla\phi(x) \neq 0$ *on K. Then*

$$|I(\lambda)| \lesssim_{N,K,\phi} \sum_{|\alpha| \leq N} \left\| \frac{D^\alpha a}{|\nabla\phi|^{2N-|\alpha|}} \right\|_{L^\infty(\mathbb{R}^d)} \lambda^{-N}.$$

where the implicit constant can be made independent of ϕ *for a family of phases if the functions* ϕ *are uniformly bounded in* $C^{N+1}(\mathbb{R})$

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

$$|I(\lambda)| \leqslant \int |(D^*)^N a(x)|\, dx \lesssim_{\phi,a,n} \frac{1}{\lambda^N},$$

and a more careful analysis of this gives the more precise estimates in the lemma.
$\qed$

The quadratic phase argument in the singlevariate case tensorizes, and thus allows us to establish stationary phase asymptotics for a quadratic phase in the multivariate case.

Theorem 29.10. *Let A be an invertible $d \times d$ matrix, fix $x_0 \in \mathbb{R}^d$, 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,

$$c_0 = (2\pi)^{d/2} a(x_0) \prod_{k=1}^{d} \left(\frac{i}{\mu_k} \right)^{1/2},$$

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 $\mathbb{R}$. 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. A density argument then shows the argument generalizes to any smooth a.
$\qed$

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 $a_1, \ldots, a_d \in \{\pm 1\}$ such that, in this coordinate system,

$$\phi(x_0 + t) = a_1 t_1^2 + \cdots + a_d t_d^2.$$

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 29.11. *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*

$$I(\lambda) \sim e^{i\lambda\phi(x_0)} \lambda^{-d/2} \sum_{n=0}^{\infty} c_n \lambda^{-n}.$$

Moreover,

$$c_0 = a(x_0) \cdot \prod_{k=1}^{d} (i/\mu_k)^{1/2},$$

where $\mu_1, \ldots, \mu_d$ are the eigenvalues of the Hessian of ϕ at x_0. For any N, we have

$$\left| I(\lambda) - (2\pi)^{d/2} e^{i\lambda\phi(x_0)} \lambda^{-d/2} \sum_{n=0}^{N} c_n \lambda^n \right| \lesssim_N \lambda^{-d/2-N-1},$$

where the implicit constant here is uniformly bounded provided that the amplitudes form a bounded subset of $C_c^{2N+2}(\mathbb{R}^d)$, the phases ϕ come from a bounded subset of $C^{2N+2+s}(\mathbb{R}^d)$, for some $s > d/2$, and

$$\frac{|\phi'(x)|}{|x - x_0|}$$

is uniformly bounded from below.

Remark. We can obtain a more precise formula for the higher order terms $\{c_k\}$. If we set
$$g_{x_0}(x) = \phi(x) - \phi(x_0) - (x - x_0)^T (H\phi)(x_0)(x - x_0)/2,$$

a function vanishing to third order at x_0. For each j, define a differential operator of order $2j$ by the formula

$$L_j u = - \sum_{\nu - \mu = j} \sum_{2\nu \geqslant 3\mu} \frac{1}{i^j 2^\mu \mu! \nu!} (\nabla_x^T H\phi(x_0)^{-1} \nabla_x)^\nu \{g_{x_0}^\mu u\}(x_0).$$

Then $c_j = L_j a$.

29.3 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 smooth function $a(x, \theta, \lambda)$ on $\mathbb{R}_x^d \times \mathbb{R}_\theta^p \times (0, \infty)$, such that $\operatorname{supp}_\theta(a)$ is compact, and such that

$$\left| \nabla_x^n \nabla_\theta^m \nabla_\lambda^k a(x, \theta, \lambda) \right| \lesssim \langle \lambda \rangle^{t + \delta m}$$

for some $\delta < 1$, and where the implicit constant is locally uniform in x. We can then consider the oscillatory integrals

$$I(x, \lambda) = \int a(x, \theta, \lambda) e^{i\lambda\phi(x, \theta)} \, d\theta.$$

If $\nabla_\theta \phi \neq 0$ on the support of a, then repeated integration by parts in the θ variable gives that

$$\left| \nabla_x^n \nabla_\lambda^k I(x, \lambda) \right| \lesssim_N \langle \lambda \rangle^{-N},$$

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, θ_0) with $\nabla_\theta \phi(x_0, \theta_0) = 0$, but with $H_\theta \phi(x_0, \theta_0)$ invertible. Then by the implicit function theorem, there exists a unique solution (x, θ_x) to $\nabla_\theta \phi(x, \theta) = 0$ in a small neighborhood of θ_0, for x in a small neighborhood of x_0. Intuitively, if a is supported on the small neighborhood of x_0 and θ_0, we should have

$$I(x, \lambda) = \int a(x, \theta, \lambda) e^{i\lambda\phi(x, \theta)} \, d\theta$$

$$\approx (2\pi)^{p/2} \lambda^{-p/2} e^{i\lambda\phi(x, \theta_x)} a(x, \theta_x, \lambda) \prod_{k=1}^p (i/\mu_k)^{1/2},$$

where $\mu_1, \dots, \mu_p$ are the eigenvalues of $H_\theta \phi(x, \theta_x)$. Ignoring the $e^{i\lambda\phi(x, \theta_x)}$ 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

$$\left|\nabla_x^n \nabla_\theta^m \nabla_\lambda^k a(x,\theta,\lambda)\right| \lesssim_{n,m,k} \lambda^{t+\delta m},$$

then, under the support assumptions on a,

$$\nabla_x^n \nabla_\lambda^k \left\{ e^{-i\lambda\phi(x,\theta_x)} I(x,\lambda) \right\} \lesssim_{N,n,m} \lambda^{t-p/2-k},$$

where the implicit constant is locally uniform in x. The result is a simple stationary phase calculation, left as an exercise. Thus we conclude that

$$I(x,\lambda) = e^{i\lambda\phi(x,\theta_x)} \cdot b(x,\lambda),$$

where $b \in \mathcal{S}_{\text{loc}}^{t-p/2}$. A discussion can be found in Corollary 1.1.3 of Sogge's "Fourier Integrals in Classical Analysis". On the other hand, if we have bounds of the form

$$\left|\nabla_x^n \nabla_\theta^m \nabla_\lambda^k a(x,\theta,\lambda)\right| \lesssim_{n,m,k} \lambda^{t+\delta m},$$

for $\delta < 1/2$, then we can still obtain an asymptotic expansion of the form

$$I(x,\lambda) \sim (2\pi)^{d/2} e^{i\lambda\phi(x,\theta_x)} \lambda^{-d/2} \sum_{j=0}^{\infty} c_j(x,\lambda) \lambda^{-j},$$

locally uniform in x, where the coefficients $\{c_j\}$ satisfy bounds of the form

$$\nabla_x^n \nabla_\lambda^k c_j(x,\lambda) \lesssim_{n,k} \langle \lambda \rangle^{t+(2\delta-1)j-d/2},$$

and are equal to a linear differential operator applied to a at (x,θ_x,λ), with coefficients depending solely on the behaviour of the Hessian matrix of ϕ in θ, at (x,θ_x). In particular,

$$c_0 = a(x,\theta_x,\lambda) \prod_{k=1}^{d} (i/\mu_k(x))^{1/2},$$

where $\mu_1(x),\ldots,\mu_d(x)$ are the eigenvalues of the Hessian of ϕ at (x,θ_x). More generally, we have

$$c_j(x,\lambda) = (L_j a)(x,\lambda),$$

where

$$(L_j u)(x,\lambda) = -\sum_{\nu-\mu=j} \sum_{2\nu \geqslant 3\mu} \frac{1}{i^j 2^\mu \mu! \nu!} (\nabla_\theta^T H_\theta \phi(x,\theta_x)^{-1} \nabla_\theta)^\nu \{g_{\theta_0}^\mu u\}(x,\theta_x,\lambda),$$

and where

$$g_{\theta_0}(x,\theta,\lambda) = \phi(x,\theta) - \phi(x,\theta_x) - (\theta-\theta_0)^T \cdot H_\theta \phi(x,\theta_x) \cdot (\theta-\theta_0)/2.$$

The boundedness of the functions $\{c_j\}$ then follows because the derivative bounds on a.

29.4   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 29.12. *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

$$\widehat{\sigma}(\xi) = \int_{S^{d-1}} e^{-2\pi|\xi|x_1} d\sigma(x)$$

$$= |S^{d-2}| \int_{-1}^{1} e^{-2\pi|\xi|t}(1-t^2)^{d/2-1} dt.$$

Setting $r = 2\pi|\xi|$ completes the argument. $\qquad\square$

Since the multivariate stationary phase approach is essentially 'coordinate independent', 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 29.13. *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 $\mu_1, \ldots, \mu_d$ 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 29.14. *If a d dimensional surface Σ is a smooth submanifold of $\mathbb{R}^{d+1}$, 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 $\leq N$ on M are $O_N(1)$, independently of ξ. Similarily, $H_M \phi_\xi$ 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. $\qquad\square$

If Ω is a bounded open subset of $\mathbb{R}^{d+1}$ 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 29.15. *If Ω is a bounded open subset of $\mathbb{R}^d$ 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*

$$|\hat{I}_\Omega(\xi)| \lesssim_\Omega |\xi|^{-(d+1)/2}.$$

Proof. We have

$$\hat{I}_\Omega(\xi) = \int_\Omega e^{-2\pi i \xi \cdot x}\, dx$$

Then we can apply Stoke's theorem for each $1 \leq k \leq d+1$ to conclude

$$\int_\Omega e^{-2\pi i \xi \cdot x}\, dx = \frac{(-1)^k}{2\pi i \xi_k} \int_\Sigma e^{-2\pi i \xi \cdot x}(dx^1 \wedge \cdots \wedge \widehat{dx^k} \wedge \cdots \wedge dx^n).$$

For each k, there is a smooth function ψ_k such that

$$dx^1 \wedge \cdots \wedge \widehat{dx^k} \wedge \cdots \wedge dx^n = \psi_k d\sigma.$$

Thus applying the last case, we find

$$\left| \frac{(-1)^k}{2\pi i \xi_k} \int_\Sigma e^{-2\pi i \xi \cdot x}(dx^1 \wedge \cdots \wedge \widehat{dx^k} \wedge \cdots \wedge dx^n) \right| \lesssim \xi_k^{-1} |\xi|^{-(d-1)/2}.$$

At each point ξ, if we choose ξ_k with the largest value, then $|\xi_k| \sim |\xi|$, so

$$\left| \int_\Omega e^{-2\pi i \xi \cdot x} \, dx \right| \lesssim |\xi|^{-(d+1)/2}. \qquad \square$$

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 29.16. *Let B be the unit ball in $\mathbb{R}^d$, 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

$$N(\lambda) = \hat{f}(0),$$

where $f = \mathbb{I}_{\lambda B} \cdot \sum_{n \in \mathbb{Z}^d} \delta_n$. The Poisson summation formula implies that

$$\hat{f} = \widehat{\mathbb{I}_{\lambda B}} * \left(\sum_{n \in \mathbb{Z}^d} \delta_n \right)$$

Thus

$$\hat{f}(0) = \sum_{n \in \mathbb{Z}^d} \widehat{\mathbb{I}_{\lambda B}}(n) = \lambda^d \sum_{n \in \mathbb{Z}^d} \hat{\mathbb{I}}_B(\lambda n) = |B|\lambda^d + \sum_{n \neq 0} \hat{\mathbb{I}}_B(\lambda n).$$

We want to apply the estimate $\hat{\mathbb{I}}_B(\lambda n) \lesssim (\lambda n)^{-(d+1)/2}$, but the decay here isn't enough for the sum in the Poisson summation formula to converge. To fix this, we mollify $\mathbb{I}_B$ 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

$$\chi_\lambda(x) = \varepsilon^{-d} \mathrm{Dil}_\varepsilon \beta * \mathbb{I}_{\lambda B}.$$

Then set $\tilde{N}(\lambda) = \sum_j \chi_\lambda(j)$. Then $\mathbb{I}_{\lambda B}$ is equal to χ_λ everywhere except on $\lambda - \varepsilon \leqslant |x| \leqslant \lambda + \varepsilon$, and since both functions are positive,

$$\tilde{N}(\lambda - \varepsilon) \leqslant N(\lambda) \leqslant \tilde{N}(\lambda + \varepsilon).$$

We can apply the Poisson summation formula now as above to obtain that

$$\tilde{N}(\lambda) = \lambda^d |B| + \sum_{n \neq 0} \hat{\mathbb{I}}_B(\lambda n) \hat{\beta}(\varepsilon n).$$

535

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

$$\widetilde{N}(\lambda) = \lambda^d |B| + O\big(\lambda^{-(d+1)/2} \varepsilon^{-(d-1)/2}\big).$$

Thus

$$N(\lambda) = \lambda^d |B| + O\big(\varepsilon \lambda^{d-1} + \lambda^{(d-1)/2} \varepsilon^{-(d-1)/2}\big)$$

Choosing $\varepsilon = \lambda^{-(d-1)/(d+1)}$ completes the proof. $\qquad\square$

Remark. One can never have $N(\lambda) = |B|\lambda^d + O(\lambda^{d-2})$, because TODO

Chapter 30

Oscillatory Integral Operators

30.1 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

$$T_\lambda f(x) = \int_{\mathbb{R}^n} a(x, y) e^{2\pi i \lambda \phi(x,y)} f(y) \, dy,$$

where $a \in C_c^\infty(\mathbb{R}^m \times \mathbb{R}^n)$ is real-valued and $\phi \in C^\infty(\mathbb{R}^m \times \mathbb{R}^n)$. 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

$$\mathrm{WF}(K_\lambda) \subset \{(x, y; \lambda \nabla_x \phi(x, y), \lambda \nabla_y \phi(x, y)) : x \in \mathbb{R}^m, y \in \mathbb{R}^n\}.$$

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*

$$C_\phi = \{(x, y; \nabla_x\phi(x, y), -\nabla_y\phi(x, y)) : x \in \mathbb{R}^m, y \in \mathbb{R}^n\}.$$

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 $\|T_\lambda f\|_{L^q(\mathbb{R}^m)} \gtrsim \lambda^{-n/2}\lambda^{-m/2q}$, whereas $\|f\|_{L^p(\mathbb{R}^n)} \sim \lambda^{-n/2p}$, which implies that the operator norm of T_λ from $L^p(\mathbb{R}^n)$ to $L^q(\mathbb{R}^m)$ 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 $\|T_\lambda f\|_{L^\infty(\mathbb{R}^m)} \lesssim \|f\|_{L^1(\mathbb{R}^n)}$ 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 $C_\phi \to T^*\mathbb{R}^n$ is a submersion, and $C_\phi \to T^*\mathbb{R}^m$ is an immersion, with these properties swapped if $m \leqslant n$. Thus, locally speaking, we have a smooth map $T^*\mathbb{R}^m \to C_\phi$ 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 $\alpha = (y, \eta) : T^*\mathbb{R}^m \to T^*\mathbb{R}^n$ such that if

538

$(x_0, y_0; \xi_0, \eta_0) \in C_\phi$, 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 30.1. *Consider a family of oscillatory integral operators*

$$\{T_\lambda : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^m)\}$$

with associated phase ϕ. If $D_y \nabla_x \phi$ is injective, i.e. the map $C_\phi \to T^* \mathbb{R}^m$ is an immersion, then

$$\|T_\lambda f\|_{L^2(\mathbb{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbb{R}^n)}.$$

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 $\|T_\lambda^* T_\lambda f\|_{L^1(\mathbb{R}^n)} \lesssim \lambda^{-n} \|f\|_{L^1(\mathbb{R}^n)}$, and that $\|T_\lambda^* T_\lambda f\|_{L^\infty(\mathbb{R}^n)} \lesssim \lambda^{-n} \|f\|_{L^\infty(\mathbb{R}^n)}$, and interpolation that $\|T_\lambda^* T_\lambda f\|_{L^2(\mathbb{R}^n)} \lesssim \lambda^{-n} \|f\|_{L^2(\mathbb{R}^n)}$, and so $\|T_\lambda f\|_{L^2(\mathbb{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbb{R}^n)}$. $\qquad\square$

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 $\|T_\lambda f\|_{L^\infty(\mathbb{R}^n)} \lesssim \|f\|_{L^1(\mathbb{R}^n)}$ yields that for $2 \leqslant q \leqslant \infty$, if p is the conjugate dual of q, then

$$\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q}\|f\|_{L^p(\mathbb{R}^n)}.$$

More generally, using the fact that $T_\lambda f$ only depends on the behaviour of f on $\operatorname{supp}_y(a)$, we conclude that we have a bound $\|T_\lambda f\|_{L^\infty(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}$ for any $1 \leqslant p \leqslant \infty$. Interpolating this bound with $\|T_\lambda f\|_{L^2(\mathbb{R}^n)} \lesssim \lambda^{-n/2}\|f\|_{L^2(\mathbb{R}^n)}$ then gives $\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q}\|f\|_{L^p(\mathbb{R}^n)}$ 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 $\|\widehat{f}\|_{L^q(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}$, 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 $a \in C_c^\infty(\mathbb{R})$, which equal the normal Fourier transform in the limit as $\lambda \to \infty$ for suitably regular functions f. The operators $T_\lambda = \lambda^{-n}F_\lambda \circ \operatorname{Dil}_\lambda$ 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 $\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q}\|f\|_{L^p(\mathbb{R}^n)}$. Rescaling this inequality shows that $\|F_\lambda f\|_{L^q} \lesssim \|f\|_{L^p(\mathbb{R}^n)}$, 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 $\|T_\lambda f\|_{L^2(\mathbb{R}^n)} \lesssim \lambda^{-n/2} \|f\|_{L^2(\mathbb{R}^n)}$ is impossible for such a phase. More precisely, fix $x_0 \in \operatorname{supp}_x(a)$ and a half-line $l = \{x \in \mathbb{R}^n : (x-x_0)/|x-x_0| = \xi_0\}$ starting at x_0 for some $|\xi_0| = 1$ whose intersection with $\operatorname{supp}_y(a)$ 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 $l \cap \operatorname{supp}_y(a)$, with spatial uncertainty $\lambda^{-1/2}$, and with frequency concentrated near ξ_0. If $f = f_1 + \cdots + f_N$, then the formula above implies that $T_\lambda f$ will have the majority of it's spatial support in the set

$$\{x_0 \in \mathbb{R}^n : |\lambda(x_0 - y_0)/|x_0 - y_0| - \xi_0| \lesssim \lambda^{-1/2}\},$$

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 $\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \gtrsim \lambda^{-(n-1)/2} \lambda^{-(n-1)/2q}$, and so combined with the fact that $\|T_\lambda f\|_{L^p(\mathbb{R}^n)} \sim \lambda^{-(n-1)/2p}$, 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 $\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbb{R}^n)}$ from holding when $q \geq (n+1)/(n-1) \cdot p^*$, which forms a subset of the family of all of the cases we proved above when, in addition, $1 \leq p \leq 2$. We will show, using the fact that the projections of C_ϕ onto $T^* \mathbb{R}^n$ are surfaces of constant curvature, that we can obtain the bound under this assumption when $1 \leq p \leq 2$, and even when $1 \leq p \leq 4$ 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 $\|\tilde{T}_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^p(\mathbb{R}^n)}$, 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 $W_\alpha \subset \mathbb{R}^{n-1}$, 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 $\mathbb{R}^n$. 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

$$\begin{aligned}
T_\lambda^\alpha &= \int a(x,y)\psi_\alpha(x,y)e^{2\pi i\lambda|x-y|}\, dy \\
&= \int \frac{a(x,y_\alpha(v,t))\psi_\alpha(x,y_\alpha(v,t))}{|\det(Dy_\alpha(v,t))|}e^{2\pi i\lambda|x-y_\alpha(v,t)|}f(y_\alpha(v,t))\, dv\, dt \\
&= \int_{I_\alpha}\int_{W_\alpha} a_\alpha(x,v;t)e^{2\pi i\lambda|x-y_\alpha(v,t)|}f(y_\alpha(v,t))\, dv\, dt \\
&= \int_{I_\alpha} T_\lambda^{\alpha,t}f_{t,\alpha}(x),
\end{aligned}$$

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 $\|T^{\alpha,t}f_{t,\alpha}\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q}\|f_{t,\alpha}\|_{L^p(W_\alpha)}$, from which, together with Hölder's inequality, it follows that

$$\begin{aligned}
\|T_\lambda^\alpha f\|_{L^q(\mathbb{R}^n)} &\lesssim \int_{I_\alpha} \|T_\lambda^{\alpha,t}f_{t,\alpha}\|_{L^q(\mathbb{R}^n)} \\
&\lesssim \lambda^{-n/q}\int_{I_\alpha} \|f_{t,\alpha}\|_{L^p(W_\alpha)}\, dt \\
&\lesssim \lambda^{-n/q}|I_\alpha|^{1-1/p}\left(\int \|f_{t,\alpha}\|_{L^p(W_\alpha)}^p\right)^{1/p} \\
&\lesssim \lambda^{-n/q}\|f\|_{L^p(\mathbb{R}^n)}.
\end{aligned}$$

Let us now prove this lemma, we note that the canonical relation associated with the family of operators $\{T_\lambda^{\alpha,t}\}$ is

$$C^{\alpha,t} = \left\{\left(x,v;\frac{|x-y_\alpha(v,t)|}{|x-y_\alpha(v,t)|}, D_v(y_\alpha)(v,t)^T \cdot \frac{x-y_\alpha(v,t)}{|x-y_\alpha(v,t)|}\right)\right\}.$$

542

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 $C^{\alpha,t} \to T^*\mathbb{R}^n$ is an open subset of the submanifold of unit vectors in $T^*\mathbb{R}^n$, 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 30.2. *Suppose $\phi : \mathbb{R}^n \times \mathbb{R}^{n-1}$ is a C^∞ phase, that $\nabla_x D_y\phi$ has rank $n-1$ on the support of $a \in C_c^\infty(\mathbb{R}^n \times \mathbb{R}^{n-1})$, and that the image of the projection $C_\phi \to T * \mathbb{R}^n$ 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*

$$\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q}\|f\|_{L^p(\mathbb{R}^{n-1})}.$$

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 $\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \geqslant \lambda^{(n-2)/2p-(n-2)/2-n/2q}\|f\|_{L^p(\mathbb{R}^{n-1})}$, 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

$$\left\|\int_{\mathbb{R}^{n-1}} e^{2\pi i x\cdot\phi(\xi)} f(\xi)\, d\xi\right\|_{L^q(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^{n-1})}$$

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

$$\|T_\lambda f\|_{L^{(n+1)/2(n-1)}(\mathbb{R}^n)} \lesssim \lambda^{-n(n-1)/2(n+1)}\|f\|_{L^2(\mathbb{R}^{n-1})}.$$

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 $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{n-1}$, 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^t_λ, mapping functions from $\mathbb{R}^{n-1}$ to $\mathbb{R}^{n-1}$ by the formula

$$T^t_\lambda f(z) = \int a(z, y; t) e^{2\pi i \lambda \phi(z, y; t)} f(y) \, dy.$$

Then

$$\|T_\lambda f\|_{L^q(\mathbb{R}^d)} \sim \left(\int \|T^t_\lambda f\|^q_{L^q(\mathbb{R}^{n-1})} \, dt \right)^{1/q},$$

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

$$\|T^t_\lambda f\|_{L^q(\mathbb{R}^{n-1})} \lesssim \lambda^{-(n-1)/q} \|f\|_{L^p(\mathbb{R}^{n-1})}.$$

However, directly plugging this into the inequality only gives a bound $\|T_\lambda f\|_{L^q(\mathbb{R}^n)} \lesssim \lambda^{-n/q} \|f\|_{L^q(\mathbb{R}^{n-1})}$. 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 $b \in C^\infty_c$ 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

$$\begin{aligned}
\nabla_y \Phi(z_1, z_2, y, t_1, t_2) = {} & D_z \nabla_y \phi(z_2, y, t_2) \cdot (z_1 - z_2) \\
& + D_t \nabla_y \phi(z_2, y, t_2) \cdot (t_1 - t_2) \\
& + O(|z_1 - z_2|^2 + |t_1 - t_2|^2).
\end{aligned}$$

544

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

$$\|T_\lambda^{t_1}(T_\lambda^{t_2})^* f\|_{L^\infty(\mathbb{R}^{n-1})} \lesssim \lambda^{-(n-1)/2} |t_1 - t_2|^{-(n-1)/2} \|f\|_{L^1(\mathbb{R}^{n-1})}.$$

This looks like something we could apply the theory of fractional integrals to understand. And indeed we can, as the next result shows: TODO. $\qquad \square$

Chapter 31

Restriction Theorems

If a function f lies in $L^p(\mathbb{R}^d)$, 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

$$\|f\|_{L^q(S,\mu)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in C(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$, and thus there does not exist a bounded operator $T : L^p(\mathbb{R}^d) \to L^q(S,\mu)$ which agrees with the standard restriction of continuous functions to sets of measure zero. We can consider a very similar problem for the Fourier transform; the Fourier transform of a function $f \in L^1(\mathbb{R}^d)$ is continuous, and thus it is meaningful to consider the restriction $\widehat{f}\,|_S$. If there exists a bound of the form

$$\|\widehat{f}\|_{L^q(S,\mu)} \lesssim \|f\|_{L^p(\mathbb{R}^d)} \tag{31.1}$$

for all $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$, then the Hahn-Banach theorem ensures that there exists a bounded operator $R : L^p(\mathbb{R}^d) \to L^q(S,\mu)$, such that $Rf = \widehat{f}\,|_S$ for $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$; for $p < \infty$, R will be unique since $L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ will be dense in $L^\infty(\mathbb{R}^d)$. Thus we have a meaningful way of restricting the Fourier transforms of functions in $L^p(\mathbb{R}^d)$ to S. In light of the failure to restrict elements of $L^p(\mathbb{R}^d)$ to $L^q(S,\mu)$ without taking the Fourier transform, this indicates that the Fourier transform of $L^p(\mathbb{R}^d)$ is a fairly special element of $L^q(\mathbb{R}^d)$, in particular, taking values in a subclass of $L^q(\mathbb{R}^d)$ in which it is meaningful to restrict to the set S. 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 $p \leqslant q$ holds for any meaningful restriction estimate. For

$p = 1$, a restriction bound from $L^1(\mathbb{R}^d)$ to $L^\infty(S, \mu)$ is trivial for any set S and measure μ. On the other hand, for $p = 2$ and $q > 0$, *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, i.e. because the Fourier transform of an element of $L^2(\mathbb{R}^d_x)$ in general behaves like an arbirary element of $L^2(\mathbb{R}^d_\xi)$, and thus cannot be restricted to singular sets. Interpolation shows we also should not expect any bounds when $p > 2$, and indeed such restriction estimates even fail when μ is absolutely continuous with respect to the Lebesgue measure, 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 as possible to $\max(2, q)$.

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 $g \cdot \mu$. Analogous to the restriction operator, it is simple to verify the estimate

$$\|E_S g\|_{L^\infty(\mathbb{R}^d)} \leqslant \|g\|_{L^1(S, \mu)},$$

and that unless μ is absolutely continuous with respect to the Lebesgue measure, no bound of the form

$$\|E_S g\|_{L^2(\mathbb{R}^d)} \lesssim \|g\|_{L^p(S, \mu)}$$

exists. The restriction and extension operators are connected by adjointness for all cases of interest, i.e. using Fubini's theorem that if $f \in L^1(\mathbb{R}^d)$ and $g \in L^1(S, \mu)$, then

$$\int_S (R_S f)(\xi) \overline{g(\xi)} \, d\mu(\xi) = \int_{\mathbb{R}^d} f(x) \overline{(E_S g)(x)} \, dx.$$

A density argument allows us to conclude that for $1 \leqslant p < \infty$ and $1 < q \leqslant \infty$, there exists a bound of the form

$$\|R_S f\|_{L^q(S, \mu)} \leqslant C \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in L^p(\mathbb{R}^d)$ if and only if there exists a bound of the form

$$\|E_S f\|_{L^{p*}(\mathbb{R}^d)} \leqslant C \|f\|_{L^{q*}(\mathbb{R}^d)}.$$

547

But this covers all the cases of interest anyway, since interesting results only hold in the regime $1 < p < 2$ and $p \leqslant q$.

Restriction bounds for product sets reduce to bounds on their projections. Thus the interesting sets S we consider will never be product sets.

Theorem 31.1. *Consider two pairs (S_1, μ_1) and (S_2, μ_2) in $\mathbb{R}^n$ and $\mathbb{R}^m$, 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*

$$\|\widehat{f}\|_{L^q(S,\mu)} \lesssim \|f\|_{L^p(\mathbb{R}^{n+m})}$$

for all $f \in \mathcal{S}(\mathbb{R}^{n+m})$ is equivalent to a pair of bounds of the form

$$\|\widehat{f_1}\|_{L^q(S_1,\mu_1)} \lesssim \|f_1\|_{L^p(\mathbb{R}^n)} \quad and \quad \|\widehat{f_2}\|_{L^q(S_2,\mu_2)} \lesssim \|f_2\|_{L^p(\mathbb{R}^m)}.$$

for all $f_1 \in \mathcal{S}(\mathbb{R}^n)$ and $f_2 \in \mathcal{S}(\mathbb{R}^m)$. Moreover, the operator norm of the restriction operator to S is the product of the operator norms of the restriction operators to S_1 and S_2.

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

$$\|f_1 \otimes f_2\|_{L^p(\mathbb{R}^{n+m})} = \|f_1\|_{L^p(\mathbb{R}^n)}\|f_2\|_{L^p(\mathbb{R}^m)}.$$

If a restriction estimate held on (S, μ), it would then follow that

$$\|\widehat{f_1}\|_{L^q(S_1,\mu_1)} \lesssim \frac{\|f_2\|_{L^p(\mathbb{R}^d)}}{\|\widehat{f_2}\|_{L^q(S_2,\mu_2)}} \|f_1\|_{L^p(\mathbb{R}^n)}.$$

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 $f \in L^1(\mathbb{R}^{n+m}) \cap L^p(\mathbb{R}^{n+m})$. Without loss of generality, we may assume by Littlewood's principle that $q \geqslant p$. If we let $\mathcal{F}_x$ and $\mathcal{F}_y$ denote the Fourier transform in the first and second variables respectively, then by applying Minkowski's

inequality and Fubini's theorem we find that

$$\|\widehat{f}\|_{L^q(S,\mu)} = \left(\int_{\mathbb{R}^m} \int_{\mathbb{R}^n} |\mathcal{F}_x \mathcal{F}_y f(\xi,\eta)|^q \, d\mu_1(\xi) d\mu_2(\eta) \right)^{1/q}$$

$$\lesssim \left(\int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} |(\mathcal{F}_y f)(x,\eta)|^p dx \right)^{q/p} d\mu_2(\eta) \right)^{1/q}$$

$$\leqslant \left(\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} |(\mathcal{F}_y f)(x,\eta)|^q \, d\mu_2(\eta) \right)^{p/q} dx_1 \right)^{1/p}$$

$$\lesssim \left(\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} |f(x,y)|^p dy dx \right)^{1/p} = \|f\|_{L^p(\mathbb{R}^d)}. \qquad \square$$

The main example of current research into restriction theory occurs when S is a smooth surface, and where $\mu = \psi\sigma$, where $\psi \in C_c^\infty(S)$. If $S \subset \mathbb{R}^d$ is the graph of some smooth function $\phi : U \to \mathbb{R}$, where U is an open subset of $\mathbb{R}^{d-1}$, then the restriction estimate is equivalent to an estimate of the form

$$\|\widehat{f}(\eta,\phi(\eta))\|_{L_\eta^q(U,\psi)} \lesssim \|f\|_{L_x^p},$$

and an extension estimate, equivalent to an estimate of the form

$$\left\| \int_U \psi(\eta) g(\eta) e^{2\pi i(\eta \cdot x + \phi(\eta)t)} \, d\eta \right\|_{L_x^{p*} L_t^{p*}} \lesssim \|g\|_{L^{q*}(U,\psi)}.$$

Since one may always localize a restriction estimate to one of this form if one works with a compactly supported measure, such estimates are no more specific than a general restriction estimate to a compact section of a hypersurface. 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. The complete resolution of the restriction problem in this setting is not solved, though much is known.

Restriction estimates often appear in the theory of partial differential equations. For instance, a tempered distribution u on $\mathbb{R}^d$ which solves the free Schrödinger equation $\partial_t u = (i/2\pi)\Delta_x u$ has space-time Fourier transform supported on the parabola defined by the equation $\tau + |\xi|^2 = 0$, and so there exists a tempered distribution v on $\mathbb{R}^d$ such that

$$u(x,t) = \int v(\xi) e^{2\pi i(\xi \cdot x - |\xi|^2 t)} \, d\xi.$$

549

Similarily, a tempered distribution u on $\mathbb{R}^n$ which solves the free wave equation $\partial_t^2 u = \Delta_x u$ has space-time Fourier transform support on the light cone defined by the equation $\tau^2 = |\xi|^2$, and thus there exists two tempered distributions v_1 and v_2 on $\mathbb{R}^n$ such that

$$u(x,t) = \int v_1(\xi)e^{2\pi i(\xi \cdot x + |\xi|t)}\, d\xi + \int v_2(\xi)e^{2\pi i(\xi \cdot x - |\xi|t)}.$$

Thus understanding the regularity of solutions to these equations relates to understanding the extension problem for the parabola and light cone.

Example. *If S is a hyperplane, then* no nontrivial estimates are possible *for the restriction problem. This is because the only possible extension-type estimates of the form*

$$\left\| \int_{\mathbb{R}^d} \psi(\eta)g(\eta)e^{2\pi i(\eta \cdot x)} \right\|_{L_x^{p*} L_t^{p*}} \lesssim \|g\|_{L^{q*}(U,\psi)}$$

must have $p = 1$, since the integral on the left hand side is independent of t. The general heuristic that has been developed in restriction theory is that 'flatness' is the main feature of the hyperplane preventing better estimates. For curved surfaces, one expects to get a much wider range of estimates.

Since flatness prevents getting good restriction estimates, we are lead to begin our study with the study of surfaces with non-vanishing curvature. The classical model 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. There are several simple examples which limit the range of estimates one can get for hypersurfaces of nonvanishing curvature. No improvements to these limitations is known, and it is conjecture that outside of these limitations, one can always get an estimate. This is the *restriction conjecture*.

The first limitation follows from stationary phase. Let S be a hypersurface with nonvanishing curvature containing, without loss of generality, the origin. The standard theory of stationary phase shows that if $\psi \in C_c^\infty(\mathbb{R}^d)$ is supported on a small enough neighborhood of the origin, then for x lying in a cone around the normal vector to the surface at the origin, we have

$$|E_S\psi(x)| \gtrsim |x|^{-(d-1)/2}.$$

Thus $E_S\psi \in L^{p*}(\mathbb{R}^d)$ if and only if $p^* > 2d/(d-1)$, i.e. if $p < 2d/(d+1)$. Thus we can only have a restriction estimate for $p < 2d/(d+1)$.

550

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. It is obtained by exploiting the uncertainty principle. 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 S contains the origin and has normal vector pointing directly upwards at the origin. Consider a bump function $\eta \in C_c^\infty(S)$ supported on a small ball of radius r about the origin, with $\eta(x) = 1$ for $|x| \leqslant 1$, but with $\|\eta\|_{L^\infty(S)} \leqslant 1$. Then

$$\|\psi\|_{L^{q*}(S,\mu)} \sim r^{(d-1)(1-1/q)}$$

On the other hand, $\eta \cdot \mu$ is support on a 'cap'

$$\Theta_r = \{(\xi,\tau) : |\xi| \leqslant r, |\tau| \leqslant cr^2\}$$

for some constant $c > 0$ depending on the surface, but independent of r. The uncertainty principle heuristically implies that $E_S\eta$, which is the Fourier transform of $\eta \cdot \mu$, is locally constant on translations of the 'tube'

$$\Theta_r^* = \{(x,t) : |x| \leqslant 1/r, |t| \leqslant C/r^2\}.$$

Since $\eta \cdot \mu$ has total mass $\sim r^{d-1}$, we conclude that $E_S\eta(x,t) \gtrsim r^{d-1}$ for $(x,t) \in \Theta_r^*$, and thus that

$$\|E_S\eta\|_{L^{p*}(\mathbb{R}^d)} \gtrsim r^{d-1}|\Theta_r^*|^{1/p^*} = r^{(d+1)/p-2}.$$

To be more rigorous, since $E_S\eta(0) \gtrsim r^{d-1}$, it follows from (TODO: record Prop 5.5 of Wolff's notes in these notes) that

$$\int |E_S\eta(x,t)|(1 + rx + r^2t)^{-N}\, dx\, dt \gtrsim_N |\Theta_r^*| r^{d-1} \gtrsim r^{-2}$$

Hölder's inequality implies that if N is chosen larger than d, then

$$\int |E_S(\psi)(x,t)|(1 + rx + r^2t)^{-N}\, dx\, dx_d \lesssim r^{-(d+1)(1/p)}\|E_S\phi\|_{L^{p*}(\mathbb{R}^d)},$$

and so we conclude that $\|E_S\phi\|_{L^{p*}(\mathbb{R}^d)} \gtrsim r^{(d+1)(1/p)-2}$. But this inequality implies that we must have $(d+1)/p - (d-1)/q \leqslant 2$ if we want an estimate of the form $\|E_S\eta\|_{L^{p*}(\mathbb{R}^d)} \lesssim \|\eta\|_{L^{q*}(S,\mu)}$ to hold.

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 $(d+1)/p - (d-1)/q \leqslant 2$, we have restriction (and thus extension) bounds from $L^{q^*}(S, \mu)$ to $L^{p^*}(\mathbb{R}^d)$, and thus a restriction bound from $L^p(\mathbb{R}^d)$ to $L^q(S, \mu)$. This is the *restriction conjecture*, which, to a large extent, still remains an open problem.

31.1 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 $\mathbb{R}^d$ with nonvanishing curvature, and $\mu = \psi\sigma$ for some $\psi \in C_c^\infty(\mathbb{R}^d)$, where σ is the surface measure on S, then for $p \leqslant 2(d+1)/(d+3)$,

$$\|\widehat{f}\|_{L^2(S)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

This result is tight, as shown by the Knapp example. There are many proofs of the Stein-Tomas theorem. The simplest, a proof by Tomas (1975), uses the theory of stationary phase, but has the disadvantage that it does not include the endpoint case where $p = 2(d+1)/(d+3)$.

Theorem 31.2. *Let S be a hypersurface with nonvanishing curvature, and let $\mu = \psi d\sigma$, where σ is the surface measure on S, and $\psi \in C_c^\infty(S)$. Then if $p < 2(d+1)/(d+3)$, then for $f \in L^1(\mathbb{R}^d)$,*

$$\|R_S f\|_{L^2(S, \mu)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

Proof. We apply a TT^* argument, which can be used to study the boundedness of any operator whose codomain is a Hilbert space. For $f \in L^1(\mathbb{R}^d)$,

$$(E_S R_S f)(x) = \int_S \widehat{f}(\xi)e^{2\pi i \xi \cdot x} d\mu(\xi) = \mu(-D)f,$$

The TT^* method implies that the restriction estimate we wish to prove holds if and only if we have a bound of the form

$$\|E_S R_S f\|_{L^{p^*}(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

If k is the convolution kernel associated with $\mu(-D)$, then $E_S R_S f = k * f$, which is a much easier operator to understand than the individual restriction and extension operators. In particular, the fact that S has nonvanishing curvature means that

$$|k(x)| \sim \langle x \rangle^{-(d-1)/2}.$$

552

This implies that $k \in L^r(\mathbb{R}^d)$ for $r > 2d/(d-1)$, and lies in $L^{r,\infty}(\mathbb{R}^d)$ for $r = 2d/(d-1)$. Thus Young's inequality, and it's weak variant, implies that

$$\|k * f\|_{L^{p*}(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

for $p \leqslant 4d/(3d+1)$. To improve this bound to hold for $p < 2(d+1)/(d-1)$, we must exploit the oscillation of k, which is completely discarded by Young's inequality.

To argue this, we consider a dyadic decomposition, writing $k(x) = \sum_{j=0}^{\infty} k_i(x)$, where k_0 is smooth and compactly supported on a ball of radius $O(1)$, and for $j > 0$, $k_j(x) = \mathrm{Dil}_{2^j}\psi \cdot k(x)$ for some $\psi \in C_c^{\infty}(\mathbb{R}^d)$. Since k_0 is smooth and compactly supported, we immediately obtain that

$$\|k_0 * f\|_{L^{p*}(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

For $j > 0$, k_j has magnitude $O(2^{-j(d-1)/2})$ on this set, so we obtain from Young's inequality that

$$\|k_j * f\|_{L^{\infty}(\mathbb{R}^d)} \lesssim 2^{-j(d-1)/2}\|f\|_{L^1(\mathbb{R}^d)}.$$

We wish to obtain an (L^2, L^2) bound using orthogonality. We have

$$\widehat{k_j} = 2^j(\widehat{k} * \mathrm{Dil}_{1/2^j}\widehat{\psi}) = 2^{jd}(\mu * \mathrm{Dil}_{1/2^j}\widehat{\psi}) = 2^{jd}\int \widehat{\psi}(2^j(\cdot - \xi))d\mu(\xi).$$

Now μ is compactly supported on a hypersurface and has total mass $O(1)$ there. Since $\mathrm{Dil}_{1/2^j}\widehat{\psi}$ is supported in a ball of radius $1/2^j$, we should expect $\widehat{k_j}$ to distribute the mass of μ onto a $1/2^j$ thickening of a portion of a hypersurface, which has total mass $O(2^{-j(d-1)})$, so that the expect height of the convolution should be $O(2^{j(d-1)})$. Thus

$$\|\widehat{k_j}\|_{L^{\infty}(\mathbb{R}^d)} \lesssim 2^{jd} \cdot 2^{-j(d-1)} = 2^j.$$

The rigorous argument is not too interesting (just use the fact that $\widehat{\psi}$ is Schwartz) and left to the reader. Thus we conclude that

$$\|k_j * f\|_{L^2(\mathbb{R}^d)} \lesssim 2^j\|f\|_{L^2(\mathbb{R}^d)}.$$

For any $p < 2(d+1)/(d+3)$, interpolation therefore tells us that there is $\varepsilon > 0$ such that

$$\|k_j * f\|_{L^{p*}(\mathbb{R}^d)} \lesssim 2^{-i\varepsilon}\|f\|_{L^p(\mathbb{R}^d)}.$$

Summing up gives the required bounds. □

553

Remark. The only structure we used about S and μ here is that μ has a uniform, Frostman type bound, and that $\hat{\mu}$ has a geometric decay as frequencies become large. For such a pair one has a bound

$$\|R_S f\|_{L^2(S,\mu)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

provided that $1 \leqslant p < 2(d + \beta - \alpha)/(2d + \beta - 2\alpha)$, where $0 < \alpha < n, \beta > 0$, where $|\hat{\mu}(\xi)| \lesssim \langle\xi\rangle^{-\beta}$ and $|\mu(B_r(x))| \lesssim r^\alpha$.

The endpoint proof $p = 2(d + 1)/(d + 3)$ was first given by Stein (1975, unpublished). It involves the same idea, namely, to consider a dyadic decomposition, but to replace the inefficient use of the triangle inequality

$$\|k * f\|_{L^{p*}(\mathbb{R}^d)} \leqslant \sum_{i=0}^{\infty} \|k_i * f\|_{L^{p*}(\mathbb{R}^d)}$$

with a complex interpolation argument. Namely, we prove that for all $t \in \mathbb{R}$,

$$\left\| \sum_{j>0} 2^{[\frac{n-1}{2}+it]j}(f * k_j) \right\|_{L^\infty(\mathbb{R}^d)} \lesssim \|f\|_{L^1(\mathbb{R}^d)}$$

and that

$$\left\| \sum_{j>0} 2^{[-1+it]j}(f * k_j) \right\|_{L^2(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^d)},$$

so that we can apply analytic interpolation. From Young's inequality, the first inequality follows if we can show that

$$\left\| \sum_{j>0} 2^{[\frac{d-1}{2}+it]j} k_j \right\|_{L^\infty(\mathbb{R}^d)} \lesssim 1.$$

This just follows from the fact that all but consecutive kernels in the sequence $\{k_j\}$ have disjoint support (this is much more efficient than the triangle inequality), and the appropriate decay. To establish the second inequality, we must show that

$$\left\| \sum_{j>0} 2^{[(d-1)+it]j}(\mu * \mathrm{Dil}_{1/2^j}\hat{\psi}) \right\|_{L^\infty(\mathbb{R}^d)}.$$

We must be slightly more efficient than we were before. TODO: See Tao's Notes.

Here is an argument that gives the endpoint for the special case of the sphere, relying on an alternate property of the sphere than curvature.

Lemma 31.3. *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 $\sigma_\varepsilon = (1/\varepsilon)\mathbb{I}_{S_\varepsilon}$. 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 $\alpha \in \mathbb{R}, \beta \in \mathbb{R}^{n-1}$, 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 $\qquad \square$

31.2   Bochner-Riesz Multipliers

One consequence of the Stein-Tomas theorem is a characterization of the boundedness for the Bochner-Riesz multipliers

$$m^\delta(\xi) = (1 - |\xi|^2)_+^\delta.$$

Herz showed a necessary condition for boundedness.

Theorem 31.4. *If $m^\delta \in M^p(\mathbb{R}^d)$, then*

$$\delta > n \left| \frac{1}{p} - \frac{1}{2} \right| - 1/2.$$

Proof. An oscillatory integral calculation shows that if $\widehat{K^\delta} = m^\delta$, then

$$K_\delta(x) = \frac{C_d \cos(2\pi|x|)}{|x|^{\frac{d+1}{2}+\delta}} + O(|x|^{-\frac{n+3}{2}+\delta}).$$

555

If χ is the indicator function of the unit ball, then Minkowski's inequality shows that for $1 \leqslant p \leqslant \infty$,

$$\|K^\delta\|_{L^p(\mathbb{R}^d)} \lesssim \|m^\delta(D)\chi\|_{L^p(\mathbb{R}^d)} \lesssim \|m^\delta\|_{M^p(\mathbb{R}^d)}.$$

Without loss of generality, we may assume that $p \geqslant 2$, in which case it suffices to show that one can only have a bound if $d(1/2 - 1/p) - 1/2 < \delta$. But K^δ lies in $L^p(\mathbb{R}^d)$ precisely for this range. $\qquad\square$

The Bochner Riesz conjecture is that this is *precisely* the range for which the operator is bounded.

Theorem 31.5. *Suppose that*
$$\|R$$

Proof. Assume $p \geqslant 2$. We break the kernel K^δ into dyadic space regions, i.e.

$$K^\delta = \sum_{j=0}^\infty K_j^\delta,$$

where K_0^δ is supported in the ball of radius 2, and

$$K_j^\delta(x) = \psi(x/2^j)K^\delta(x) = \psi_j(x)K^\delta(x).$$

Our proof will be complete if we can show that

$$\|K_j^\delta * f\|_{L^p(\mathbb{R}^d)} \lesssim 2^{[n(1/p-1/2)-1/2-\delta]j}\|f\|_{L^p(\mathbb{R}^d)}.$$

Since the support of $K_j^\delta * f$ lies on a 2^j thickening of the support of f, a decomposition argument implies that it suffices to show the result for any function f supported on a ball of radius 2^j at the origin. To make this look more like Tomas-Stein, we note that for such a function f, $K_j^\delta * f$ is supported on a ball of radius $O(2^j)$, and thus

$$\|K_j^\delta * f\|_{L^p(\mathbb{R}^d)} \lesssim 2^{dj(1/p-1/2)}\|K_j^\delta * f\|_{L^2(\mathbb{R}^d)}.$$

Thus it suffices to show that

$$\|K_\delta^j * f\|_{L^2(\mathbb{R}^d)} \lesssim 2^{-(1/2+\delta)j}\|f\|_{L^p(\mathbb{R}^d)}.$$

Plancherel shows that it suffices to prove that

$$\|[\widehat{\psi}_j * m^\delta] \cdot \widehat{f}(\xi)\|_{L_\xi^2} \lesssim 2^{-(1/2+\delta)j}\|f\|_{L_x^p}$$

Now ψ_j is supported away from the origin, and thus it's Fourier transform is highly oscillatory. The multiplier m^δ is also proportionally smooth. Thus we should expect $\widehat{\psi_j} * m^\delta$ to have some decay by virtue of some cancellation. The 'high frequency' terms that make up m^δ are supported near the boundary of the sphere, since the function is smooth everywhere else. We will actually obtain that

$$|(\widehat{\psi_j} * m^\delta)(\xi)| \lesssim_N 2^{-\delta j} \langle 2^j ||\xi| - 1| \rangle^{-N}.$$

Thus we get rapid decay outside of the annulus of thickeness $1/2^j$. It thus suffices to show that if N is suitably large, then

$$\int \frac{|\widehat{f}(\xi)|^2}{\langle 2^j ||\xi| - 1| \rangle^N} \, d\xi \lesssim \|f\|_{L^p(\mathbb{R}^n)}.$$

The fact that f is compactly supported allows us to use the pointwise bound $|\widehat{f}(\xi)| \lesssim 2^{O(j)} \|f\|_{L^p(\mathbb{R}^d)}$, which gives the estimate for $||\xi| - 1| \geqslant 1/2$. The remaining integrand then follows by switching to polar coordinates, and then applying the restriction theorem.

Now let us justify that

$$|(\widehat{\psi_j} * m^\delta)(\xi)| \lesssim_N 2^{-\delta j} \langle 2^j ||\xi| - 1| \rangle^{-N}.$$

This motivates us to perform the decomposition

$$m^\delta(\xi) = m^\delta(\xi)\phi(2^j(|\xi| - 1)) + m^\delta(\xi)[1 - \phi(2^j(|\xi| - 1))] = m_1^\delta(\xi) + m_2^\delta(\xi),$$

where $\phi \in C_c^\infty[0, \infty)$ equals one in a neighborhood of the origin. The symbol $m_1^\delta(\xi)$ is now supported on the annulus of thickness $O(2^{-j})$ and radius 1. The height of the symbol is also $O(2^{-\delta j})$. Thus

$$|(\widehat{\psi_j} * m_1^\delta)(\xi)| \lesssim \int_{|r-1| \sim 2^{-j}} 2^{-\delta j}(|\widehat{\psi_j}| * \sigma_r) \, dr.$$

The fact that $\sigma_r * |\widehat{\psi_j}| \lesssim_N 2^j \langle 2^j ||x| - 1| \rangle^{-N}$ shows this contribution is acceptable. The remaining term is estimated via integration by parts (one may antidifferentiate ψ_j without introducing any problems since it vanishes near the origin). TODO.

$\square$

557

31.3 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

$$S = \{(\xi, \omega) \in \mathbb{R}^{d-1} \times \mathbb{R} : |\xi|^2 = \omega\}.$$

We work with extension estimates, setting, for a function $f : \mathbb{R}^{d-1} \to \mathbb{C}$,

$$Ef(x, t) = \int_{\mathbb{R}^{d-1}} e^{2\pi i(|\xi|^2 t + \xi \cdot x)} f(\xi)\, d\xi.$$

Our goal is to show bounds of the form $\|Ef\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^{d-1})}$. This is possible, provided that $p = 2(d+1)/(d+3)$.

Theorem 31.6. *If $p = 2(d+1)/(d+3)$, then for any $f \in \mathcal{S}(\mathbb{R}^{d-1})$,*

$$\|Ef\|_{L^{p'}(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^{d-1})}.$$

Proof. The Tomas-Stein theorem gives the required bound for any $f \in C_c^\infty(\mathbb{R}^d)$, where the implicit constant depends on the support of f. For such f, define

$$Ef(x, t) = \int_{\mathbb{R}^{d-1}} e^{2\pi i(|\xi|^2 t + \xi \cdot x)} f(\xi)\, d\xi.$$

We note that $E(\mathrm{Dil}_{\xi,a} f) = a^{d-1} \cdot \mathrm{Dil}_{a^{-2},t} \mathrm{Dil}_{a^{-1},x}(Ef)$, in light of the 'parabolic symmetry' of the extension operator. Given a function $f \in C_c^\infty(\mathbb{R}^d)$ supported on a ball of radius R at the origin, $\mathrm{Dil}_{\xi,1/R} f$ is supported on a ball of radius 1 at the origin, and so Tomas-Stein says that

$$\|E(\mathrm{Dil}_{\xi,1/R} f)\|_{L^{p'}(\mathbb{R}^d)} \lesssim \|\mathrm{Dil}_{\xi,1/R} f\|_{L^2(\mathbb{R}^{d-1})} = R^{-(d-1)/2} \cdot \|f\|_{L^2(\mathbb{R}^{d-1})},$$

where the implicit constant is independent of f. But we also know that

$$\begin{aligned}
\|E(\mathrm{Dil}_{\xi,1/R} f)\|_{L^{p'}(\mathbb{R}^d)} &= R^{1-d} \|\mathrm{Dil}_{R^2,t} \mathrm{Dil}_{R,x}(Ef)\|_{L^{p'}(\mathbb{R}^d)} \\
&= R^{1-d} R^{2/p'} R^{(d-1)/p'} \|Ef\|_{L^{p'}(\mathbb{R}^d)},
\end{aligned}$$

and so we conclude that

$$\|Ef\|_{L^{p'}(\mathbb{R}^d)} \lesssim R^{(d-1)/2 - (d+1)/p'} \|f\|_{L^2(\mathbb{R}^{d-1})} = \|f\|_{L^2(\mathbb{R}^{d-1})}.$$

In other words, we have found a bound independent of the support of f. A simple approximation argument extends the bound to general $f \in \mathcal{S}(\mathbb{R}^d)$. $\qquad\square$

Remark. The same scaling symmetries show that we can only have an extension bound

$$\|Ef\|_{L^{p'}(\mathbb{R}^d)} \lesssim \|f\|_{L^{q'}(\mathbb{R}^{d-1})}$$

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

$$\|Ef\|_{L^{p'}(\mathbb{R}^d)} \lesssim \|f\|_{L^{q'}(S,\mu)},$$

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

$$\|Ef\|_{L^{p'}(\mathbb{R}^d)} \lesssim \|f\|_{L^{q'}(\mathbb{R}^{d-1})}$$

for any $f \in \mathcal{S}(\mathbb{R}^{d-1})$, 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.

31.4 Restriction to the Cone

TODO: Conjectured Estimates
    TODO: Lorentz Invariant
    TODO: Application to the wave equation.

31.5 TODO: Hardy Littlewood Majorant Conjecture

31.6 Guth, Wang, and Zhang

The goal of this paper is to obtain a square function estimate for decomposing functions with Fourier support on a neighborhood of the cone into sectors: let

$$\Gamma = \{\xi \in \mathbb{R}^3 : \xi_1^2 + \xi_2^2 = \xi_3^2, 1/2 \leqslant \xi_3 \leqslant 1\}$$

denote the *truncated light cone* in $\mathbb{R}^3$. For any $R > 0$, write Γ as the union of at most $10R^2$ finitely overlapping sectors $\{\theta\}$, where each θ has aperture $R^{-1/2}$. Let Q_θ be a rectangle with dimensions $R^{-1} \times R^{-1/2} \times 1$ containing θ. Our goal is to prove that for any function $f \in \mathcal{S}(\mathbb{R}^3)$ with Fourier support on an $1/R$ neighborhood of the cone, we have a square function identity of the form

$$\|f\|_{L^4(\mathbb{R}^3)}^4 \lesssim_\varepsilon R^\varepsilon \|S_R f\|_{L^4(\mathbb{R}^3)},$$

where

$$S_R f = \left(\sum_\theta |f_\theta|^2 \right)^{1/2},$$

where f_θ is the Fourier multiplier localizing the support of f to Q_θ. This implies the cone-multiplier conjecture, and non endpoint local-smoothing for the wave equation in $2 + 1$ dimensions. To prove this result, we prove a *stronger result*, which behaves better with respect to the technique of induction on scales.

For each sector θ, we can consider the dual rectangle Q_θ^* centered at the origin, with dual dimensions to Q_θ. The uncertainty principle says, heuristically, that each f_θ is locally constant on translates of $Q_{\theta*}$. We call $Q_{\theta*}$ a *plank*, and we call the restriction of f_θ to one plank a *wave packet*. We will prove a more general square function estimate associated with Fourier localization on slightly bigger sectors. For $R^{-1/2} \leqslant s \leqslant 1/100$, we consider a cover of an s-neighborhood of Γ by a family of at most $O(s^{-1})$ sectors $\mathcal{T}_s = \{\tau\}$. For each τ, we now consider a rectangle Q_τ containing τ with dimensions $s^2 \times s \times 1$, and consider the dual rectangle Q_τ^* through the origin. Let $U_{\tau,R} = RQ_\tau^*$ be the rescaling of Q_τ^* to have diameter R. Any translated copy of Q_θ^* which passes through the origin is contained in $10U_{\tau,R}$, and so if we tile $\mathbb{R}^3$ by translations U of $U_{\tau,R}$, then each wave packet is supported on $O(1)$ tiles in the tiling. By the notation

$$\sum_{U \mathbin{/\!/} U_{\tau,R}}$$

we mean the sum of all U in the tiling of $\mathbb{R}^3$ by translates of $U_{\tau,R}$. The improved estimate we will prove is that

$$\|f\|_{L^4(\mathbb{R}^3)}^4 \lesssim_\varepsilon R^\varepsilon \sum_{R^{-1/2} \leqslant s \leqslant 1} \sum_{\tau \in \mathcal{T}_s} \sum_{U \mathbin{/\!/} U_{\tau,R}} |U|^{-1} \|S_\tau f\|_{L^2(U)}^4,$$

where s ranges over *dyadic numbers*, and

$$S_\tau f = \left(\sum_{\theta \subset Q_\tau} |f_\theta|^2 \right)^{1/2}.$$

560

This inequality will scale better under induction on scales.

To get some intuition about this result, suppose f is essentially supported on a ball of radius R, and let's determine when the different values of s in the inequality are tight. If $s = R^{-1/2}$, then τ ranges over the original points θ, and so we get

$$\sum_{\tau \in \mathcal{T}_s} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|S_\tau f\|_{L^2(U)}^4 = \sum_{\theta} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|f_\theta\|_{L^2(U)}^4.$$

Now f_θ is roughly constant on U, which implies that

$$|U|^{-1} \|f_\theta\|_{L^2(U)}^4 \sim \|f_\theta\|_{L^4(U)}^4.$$

This gives that

$$\sum_{\tau \in \mathcal{T}_s} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|S_\tau f\|_{L^2(U)}^4 \sim \sum_{\theta} \sum_{U /\!/ U_{\tau,R}} \|f_\theta\|_{L^4(U)}^4$$

This is proportional to f if the non-zero functions in the collection $\{f_\theta\}$ all have disjoint spatial support. On the other hand, for $s = 1/100$, then we have $O(1)$ sectors $\{\tau_1, \ldots, \tau_m\}$, and $O(1)$ relevant tilings $U /\!/ U\tau_i, R$ on the support of f for each i, and we can label them $\{U_{i,j}\}$. Thus the result says that and the result says that

$$\sum_{\tau \in \mathcal{T}_s} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|S_\tau f\|_{L^2(U)}^4 \sim \max_{i,j} R^{-3} \|S_{\tau_i} f\|_{L^2(U_{ij})}^4.$$

If f is roughly constant on B_R, then

$$\|f\|_{L^4(\mathbb{R}^3)}^4 \sim R^{-3} \|f\|_{L^2(\mathbb{R}^3)}^4 \sim \max_{i,j} R^{-3} \|S_{\tau_i} f\|_{L^2(U_{ij})}^4.$$

Intermediate cases for a particular value of s occur if f has Fourier support on a sector τ, that f has a majority of mass on a translate U of $U_{\tau,R}$, and is roughly constant on this translate.

How does this sum imply the original estimate? If f is a function whose Fourier transform is supported on a $1/R$ neighborhood of Γ, then for any s, any $\tau \in \mathcal{T}_s$, and any $U /\!/ U_{\tau,R}$, Cauchy-Schwartz implies that

$$\|S_\tau f\|_{L^2(U)}^4 = \left(\int_U \sum_{\theta \subset Q_\tau} |f_\theta|^2 \right)^2 \leqslant |U| \int_U \left(\sum_{\theta \subset Q_\tau} |f_\theta|^2 \right)^2.$$

Thus

$$\sum_{\tau \in \mathcal{T}_s} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|S_\tau f\|_{L^2(U)}^4 \leqslant \sum_{\tau \in \mathcal{T}_s} \int_{\mathbb{R}^3} \left(\sum_{\theta \subset Q_\tau} |f_\theta|^2 \right)^2 \lesssim \int_{\mathbb{R}^3} \left(\sum_\theta |f_\theta|^2 \right)^2.$$

561

There are $O(\log R)$ different dyadic values s between $R^{-1/2}$ and 1, and so by summing over s, we conclude that the inequality to be proven implies that

$$\|f\|^4_{L^4(\mathbb{R}^3)} \lesssim_\varepsilon R^\varepsilon \sum_{R^{-1/2} \leqslant s \leqslant 1} \sum_{\tau \in \mathcal{T}_s} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|S_\tau f\|^4_{L^2(U)}$$

$$\lesssim_\varepsilon R^\varepsilon \log R \int_{\mathbb{R}^3} \left(\sum_\theta |f_\theta|^2 \right)^2$$

$$\lesssim R^\varepsilon \log R \cdot \|Sf\|^4_{L^2(\mathbb{R}^3)}.$$

The logarithmic term can then be absorbed into the power of ε.

We require some more notation to be more precise about out specification of the sections $\mathcal{T}_s$. For each $R^{-1/2} \leqslant s \leqslant 1$, we consider $10/s$ equally spaced points $\{\xi\}$ on the unit circle in $\mathbb{R}^2$. We will associate with each ξ a sector τ_ξ of aperture s centered at $(\xi, 1)$. More precisely, we define

$$\mathbf{c}_\xi = (\xi, 1), \quad \mathbf{n}_\xi = (\xi, -1), \quad \text{and} \quad \mathbf{t}_\xi = (\xi^\perp, 0),$$

the *core*, *normal*, and *tangential* directions of the sector, where $(\xi_1, \xi_2)^\perp = (\xi_2, -\xi_1)$. We then have

$$Q_{\tau_\xi} = \{x \in \Gamma : 1 \leqslant \mathbf{c}_\xi \cdot x \leqslant 2 \text{ and } |\mathbf{n}_\xi \cdot x| \leqslant s^2 \text{ and } |\mathbf{t}_\xi \cdot x| \leqslant s\}.$$

For $R \geqslant s^{-2}$, we then define

$$U_{\tau,R} = \{x \in \mathbb{R}^3 : |\mathbf{c} \cdot x| \leqslant Rs^2 \text{ and } |\mathbf{n} \cdot x| \leqslant R \text{ and } |\mathbf{t} \cdot x| \leqslant Rs\}.$$

The reason we use the tiling $U_{\tau,R}$ is that each set roughly contains all possible translates of dual sectors in $\mathcal{T}_{R^{-1/2}}$ intersecting dual sectors of τ.

Our goal is to perform an *induction on scales*. We let $S(r, R)$ be the smallest constant such that for any function f with Fourier support on a $1/R$ neighborhood of Γ, and any cover $\{B_r\}$ of $\mathbb{R}^3$ by radius r balls such that

$$\left\| \sum_{B_r} \mathbb{I}_{B_r} \right\|_{L^\infty} \leqslant 100,$$

we have

$$\sum_{B_r} |B_r|^{-1} \|S_r f\|^4_{L^2(B_r)} \leqslant S(r, R) \sum_{R^{-1/2} \leqslant s \leqslant 1} \sum_{\tau \in \mathcal{T}_s} \sum_{U /\!/ U_{\tau,R}} |U|^{-1} \|S_U f\|^4_{L^2(\mathbb{R}^3)},$$

where

$$S_r f = \left(\sum_{\tau \in \mathcal{T}_{r^{-1/2}}} |f_\tau|^2 \right)^{1/2}$$

Our goal is to obtain from a series of bounds on the function $S(r, R)$ that $S(1, R) \lesssim_\varepsilon R^\varepsilon$, which would imply the required result.

We will prove the inequality required by a Kakeya type estimate that takes these scales into account, namely, that if $\hat{f}$ has Fourier support on a $1/R$ neighborhood of the cone, then

$$\int_{\mathbb{R}^3} \left| \sum_\theta |f_\theta|^2 \right|^2 \lesssim \sum_{R^{-1/2} \leqslant s \leqslant 1} \sum_{\tau \in \mathcal{T}_s} \sum_{U \,/\!/\, U_{\tau,R}} |U|^{-1} \|S_\tau f\|_{L^2(U)}^4.$$

It will follow from this that for any $R \geqslant 10$, and $r \in [R, R^2]$, $S(r, R) \lesssim 1$.

Next, we use a trick to reduce the study of the cone to a subset of the parabola.

TODO: HOW DO THE REQUIRED INEQUALITY FOLLOW FROM THIS.
Question for Dimas: Can you explain why decoupling can't prove local smoothing past the range $p \leqslant 6$? TODO: Go over arguments for Mockenhaupt, Seeger, and Sogge which explain how this result implies local smoothing. TODO: Read Tao and Vargas.

Chapter 32

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 32.1. *Suppose $\{e_n : n \in \mathbb{Z}^d\}$ 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 con-

clude that

$$\|\sum_n a_n e_n\|_H^2 \leqslant C \sum_n |a_n| \sum_m \frac{|a_m|}{\langle n - m \rangle^{d+\delta}}$$

$$\leqslant C \left(\sum_n |a_n|^2 \right)^{1/2} \left(\sum_n \left(\sum_m \frac{|a_m|}{\langle n - m \rangle^{d+\delta}} \right)^2 \right)^{1/2}$$

$$\lesssim_\delta C \left(\sum_n |a_n|^2 \right).$$

We then just take square roots on both sides of the formula. $\qquad \square$

Let us consider some examples of almost orthogonal systems.

Example. *Consider $f \in L^2(\mathbb{R}^d)$ 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 $\langle e_n, e_m \rangle \leqslant \|f\|_{L^2(\mathbb{R}^d)}^2$. Thus we conclude that*

$$\|\sum a_n e_n\|_{L^2(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^d)} \left(\sum |a_n|^2 \right)^{1/2}.$$

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 $\phi \in C_c^\infty(\mathbb{R}^d)$, 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

$$\|\sum a_n e_n\|_{L^2(\mathbb{R}^d)} \gtrsim \left(\sum |a_n|^2 \right)^{1/2}.$$

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 $L^2(\mathbb{R}^d)$ 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). To begin with, let us treat the case where we are dealing with a family of operators which are pairwise orthogonal.

Lemma 32.2. *Let $\{T_n\}$ be a sequence of operators on a Hilbert space H with $\sup_n \|T_n\| < \infty$ for each n, and such that $T_i^* T_j = T_i T_j^* = 0$ for any $i \neq j$. If*

$$S_N = \sum_{n=1}^{N} T_n,$$

then S_N converges in the strong topology to an operator S with $\|S\| \leqslant 1$.

Proof. Since $T_i^* T_j = 0$ for $i \neq j$, it follows that the image of T_j is contained in the kernel of T_i^*, which is the orthogonal complement of the image of T_i. Thus the images $\{V_i\}$ of the operators $\{T_i\}$ are all orthogonal to one another. By symmetry, the images $\{W_i\}$ of the operators $\{T_i^*\}$ are all orthogonal to one another. Now W_i is the orthogonal complement of the kernel of T_i. Thus if $x = x_0 + \sum_i x_i$, with x_0 perpendicular to all W_i, and $x_i \in W_i$, then $\|T_i x\| = \|T_i x_i\|$. But this means that for any $x \in H$,

$$\begin{aligned}
\|S_N x\|^2 &= \sum \|T_i x\|^2 \\
&= \sum \|T_i x_i\|^2 \\
&\leqslant \sum \|x_i\|^2 \\
&\leqslant \|x\|^2.
\end{aligned}$$

Thus $\|S_N x\| \leqslant \|x\|$, and so $\|S_N\| \leqslant 1$. Now since $\sum_{i>N} \|x_i\|^2 \to 0$ as $i \to \infty$ for any $x \in H$, it follows that $\{S_N x\}$ is Cauchy, and so there exists a limit Sx. It is now simple to prove S is a bounded linear operator with $\|S\| \leqslant 1$. $\qquad\square$

Remark. The important point here is that not only are the outputs to all the operators mutually orthogonal, i.e. $T_i^* T_j = 0$, nor are the important imputs to each of the operators mutually orthogonal, i.e. $T_i T_j^* = 0$, but both points are actually true. If only one of these was true, we could only get a bound of the form $\|S_N\| \lesssim \sqrt{N}$.

Cotler-Stein gives an analogous result for almost orthogonal families of operators.

Theorem 32.3 (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*

$$A = \sup_\alpha \sum_\beta \sqrt{a_{\alpha\beta}} \quad and \quad B = \sup_\alpha \sum_\beta \sqrt{b_{\alpha\beta}}$$

are both finite, then $\sum T_\alpha$ converges unconditionally in the strong operator topology, and

$$\left\| \sum T_\alpha \right\| \leq \sqrt{AB}.$$

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 32.4. *Consider a family of functions $\{f_n : n \in \mathbb{Z}^d\}$ in $L^2(\mathbb{R}^d)$, together with a family of sidelength one cubes $\{Q_n\}$ in $\mathbb{R}^d$ such that $\mathrm{supp}(f_n) \subset Q_n$. Suppose*

$$|\langle f_n, f_m \rangle| \leq \frac{1}{\langle n - m \rangle^\beta}.$$

for some $\beta \in (0, d)$. Then for $p < 2d/(2d - \beta)$,

$$\left\| \sum a_n f_n \right\|_{L^p(\mathbb{R}^d)} \lesssim_{d,\beta,p} \left(\sum |a_n|^p \right)^{1/p}$$

Remark. If $\beta > d$, then Young's convolution inequality implies that

$$
\begin{aligned}
\Big\| \sum a_n f_n \Big\|_{L^2(\mathbb{R}^d)} &= \left(\sum_{n,m} a_n \overline{a_m} \langle f_n, f_m \rangle \right)^{1/2} \\
&\leq \left(\sum_{n,m} \frac{a_n \overline{a_m}}{\langle n - m \rangle^\beta} \right)^{1/2} \\
&\leq \left(\sum_n |a_n|^2 \right)^{1/4} \left(\sum_n \Big| \sum_m \frac{\overline{a_m}}{\langle n - m \rangle^\beta} \Big|^2 \right)^{1/2} \\
&\leq \left(\sum_n |a_n|^2 \right)^{1/2} \left(\sum_n \frac{1}{\langle n \rangle^\beta} \right) \\
&\lesssim_{\beta,d} \left(\sum_n |a_n|^2 \right)^{1/2}
\end{aligned}
$$

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 $l^p(\mathbb{Z}^d)$ to $L^p(\mathbb{R}^d)$. We shall prove that for $1 \leq p \leq 2d/(2d - \beta)$, 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 $I \subset \mathbb{Z}^d$,

$$
\Big\| \sum_{n \in I} f_n \Big\|_{L^p(\mathbb{R}^d)} \lesssim \#(I)^{1/p}.
$$

Partition $\mathbb{R}^d$ 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 $x \in \mathbb{R}^d$, there are at most 3^d indices α such that $F_\alpha(x) \neq 0$. Thus

$$
\Big\| \sum_{n \in I} f_n \Big\|_{L^p(\mathbb{R}^d)} = \Big\| \sum_\alpha F_\alpha \Big\|_{L^p(\mathbb{R}^d)} \leq 3^d \left(\sum_\alpha \| F_\alpha \|_{L^p(\mathbb{R}^d)}^p \right)^{1/p}.
$$

568

Applying the almost-orthogonality of the functions $\{f_n\}$,

$$\|F_\alpha\|_{L^2(\mathbb{R}^d)}^2 \lesssim \sum_{n,m\in I_\alpha} \frac{1}{\langle n-m\rangle^\beta}$$

$$\leqslant \sum_{n\in I_\alpha} \sum_{|m|\lesssim \#(I_\alpha)^{1/d}} \frac{1}{\langle n-m\rangle^\beta}$$

$$\lesssim \#(I_\alpha) \cdot \#(I_\alpha)^{1-\beta/d}$$

Thus $\|F_\alpha\|_{L^2(\mathbb{R}^d)} \lesssim \#(I_\alpha)^{1-\beta/2d}$. Combined with the fact that F_α is supported on a sidelength $O(1)$ cube, we conclude that for $0 < p \leqslant 2$, $\|F_\alpha\|_{L^p(\mathbb{R}^d)} \lesssim_p \|F_\alpha\|_{L^2(\mathbb{R}^d)}$. But putting this together means that

$$\left(\sum_\alpha \|F_\alpha\|_{L^p(\mathbb{R}^d)}^p\right)^{1/p} = \left(\sum_\alpha \#(I_\alpha)^{p(1-\beta/2d)}\right)^{1/p}$$

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

$$\left(\sum_\alpha \|F_\alpha\|_{L^p(\mathbb{R}^d)}^p\right)^{1/p} \lesssim \#(I)^{1/p},$$

which completes the proof of the restricted strong type bound. $\qquad\square$

Chapter 33

Weighted Estimates

33.1 Hardy-Littlewood Maximal Function

Let us consider a basic weighted estimate for the Hardy-Littlewood maximal function.

Theorem 33.1. *Suppose $w > 0$ is a measurable function. Then for any $f \in L^1_{loc}(\mathbb{R}^d)$, we have the weak type bound*

$$\int_{\mathbb{R}^d} \mathbb{I}\left(|Mf(x)| > \lambda\right) w(x)\, dx \lesssim_d \frac{1}{\lambda} \int_{\mathbb{R}^d} |f(x)| Mw(x)\, dx,$$

and for all $1 \leqslant p \leqslant \infty$, we have the strong type bound

$$\left(\int_{\mathbb{R}^d} |Mf(x)|^p w(x)\, dx\right)^{1/p} \lesssim_{d,p} \left(\int_{\mathbb{R}^d} |f(x)|^p Mw(x)\, dx\right)^{1/p}.$$

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 $x \in K$,

$$\int_K w(x)\, dx \lesssim_d \int_{\mathbb{R}^d} |f(x)| Mw(x)\, dx.$$

Find a disjoint family of balls $B_1, \ldots, B_N$ such that $3B_1, \ldots, 3B_N$ 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. $\qquad\square$

A simple corollary is a vector-valued generalization of the Hardy-Littlewood inequality.

Theorem 33.2. *If* $1 < p, q < \infty$ *and* $\{f_n\}$ *are any sequence of functions in* $L^1_{loc}(\mathbb{R}^d)$, *then*

$$\left\| \left(\sum_n |Mf_n|^p \right)^{1/p} \right\|_{L^q(\mathbb{R}^d)} \lesssim_{d,p,q} \left\| \left(\sum_n |f_n|^p \right)^{1/p} \right\|_{L^q(\mathbb{R}^d)}$$

and

$$\left| \left\{ x : \left(\sum_n |Mf_n(x)|^p \right)^{1/p} \geq \lambda \right\} \right| \lesssim_{d,p} \frac{1}{\lambda} \cdot \left\| \left(\sum_n |f_n(x)|^p \right)^{1/p} \right\|_{L^1(\mathbb{R}^d)}.$$

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

$$|Mf_n(y)| \leq \frac{1}{|B(y, r_n(y))|} \int_{\mathbb{R}^d} \psi_n(x, y) f_n(y)\, dy,$$

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

$$T(\{f_n\})(y) = \left\{ \int_{\mathbb{R}^d} \frac{1}{|B(y, r_n(y))|} \psi_n(x, y) f_n(x)\, dx \right\}.$$

This is a vector-valued kernel operator with kernel $K(x, y)$ the diagonal matrix with entry. TODO: SEE TAO. $\qquad\square$

571

Chapter 34

Bellman Function Methods

It is interesting to ask whether we can obtain bounds of the form

$$\|Mf\|_{L^p(\mathbb{R}^d)} \lesssim_{d,p} \|f\|_{L^p(\mathbb{R}^d)}$$

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 $f \in L^p(\mathbb{R}^d)$, 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 $x \in \mathbb{R}^d$, $|M_\Delta f(x)| = \lim_{m \to -\infty} |M_{\geqslant m} f(x)|$, 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 $f \in L^p(\mathbb{R}^d)$,

$$\|M_\Delta f\|_{L^p(\mathbb{R}^d)} = \lim_{m \to -\infty} \|M_{\geqslant m} f\|_{L^p(\mathbb{R}^d)}.$$

Thus if we can obtain a bound

$$\|M_{\geqslant m} f\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

with a bound independent of m, we would obtain the required bound on M_Δ. But if we could obtain a bound

$$\|M_{\geqslant 0} f\|_{L^p(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}$$

for all $f \in L^p(\mathbb{R}^d)$, then a rescaling argument, using the fact that

$$M_{\geqslant m} f = \text{Dil}_{1/2^d} M_{\geqslant 0} \text{Dil}_{2^d} f$$

shows that we in fact have

$$\begin{aligned}
\|M_{\geqslant m} f\|_{L^p(\mathbb{R}^d)} &= 2^{d/p} \|M_{\geqslant 0} \text{Dil}_{2^d} f\|_{L^p(\mathbb{R}^d)} \\
&\lesssim 2^{d/p} \|\text{Dil}_{2^d} f\|_{L^p(\mathbb{R}^d)} = \|f\|_{L^p(\mathbb{R}^d)}.
\end{aligned}$$

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 $E_m = \{x \in \mathbb{R}^d : m_x = m\}$, 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

$$\|M_{\geqslant 0} f\|_{L^p(\mathbb{R}^d)} \leqslant \|M_{\geqslant 0} f\|_{L^p(Q)} + \|M_{\geqslant 0} f\|_{L^p(Q^c)} \lesssim \|f\|_{L^p(Q)}.$$

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 $Q_{n,1}, \ldots, Q_{n,2^d}$. For any $f \in L^p(\mathbb{R}^d)$, we consider a pointwise limit $f = \lim_{n \to \infty} f_{n,1} + \cdots + f_{n,2^d}$, where $f_{n,i}$ is equal to f restricted to $Q_{n,i}$, and the limit is monotone. We also have

$$M_{\geqslant 0} f = \lim_{n \to \infty} M_{\geqslant 0} f_{n,1} + \cdots + M_{\geqslant 0} f_{n,2^d}.$$

573

where the limit is pointwise and monotone, so

$$\|M_{\geqslant 0}f\|_{L^p(\mathbb{R}^d)} = \lim_{n\to\infty} \|M_{\geqslant 0}f_{n,1} + \cdots + M_{\geqslant 0}f_{n,2^d}\|_{L^p(\mathbb{R}^d)}$$
$$\lesssim \lim_{n\to\infty} \|f_{n,1}\|_{L^p(\mathbb{R}^d)} + \cdots + \|f_{n,2^d}\|_{L^p(\mathbb{R}^d)} \lesssim 2^d\|f\|_{L^p(\mathbb{R}^d)}.$$

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 independent 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 $f \in L^p(\mathbb{R}^d)$ 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 $Q_1, \ldots, Q_{2^d}$ with sidelength 2^{n-1}. If we decompose $f = f_1 + \cdots + f_{2^d}$ 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

$$\|M_{\geqslant 0}f\|_{L^p(Q)} = \left(\|M_{\geqslant 0}f\|_{L^p(Q_1)}^p + \cdots + \|M_{\geqslant 0}f\|_{L^p(Q_{2^d})}^p \right)^{1/p}$$
$$= \left(\|\max(M_{\geqslant 0}f_1, A)\|_{L^p(Q_1)}^p + \cdots + \|\max(M_{\geqslant 0}f_{2^d}, A)\|_{L^p(Q_{2^d})}^p \right)^{1/p}$$

574

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 35

Brascamp-Lieb Inequalities

Let H, and $H_1, \ldots, H_m$ be Hilbert spaces, and consider linear maps $L_j : H \to H_j$ for $1 \leqslant j \leqslant m$. The Brascamp-Lieb inequality asks whether we have a multilinear inequality of the form

$$\int_H \prod_j (f_j \circ L_j)^{p_j} \lesssim \prod_j \|f_j\|_{L^{p_j}(H_j)}^{p_j}$$

for all non-negative measurable functions $f_j : H_j \to [0, \infty)$, and some fixed exponents $p_1, \ldots, p_j$. The best possible constant in this inequality is denoted $\mathrm{BL}(L, p)$. Here are some examples:

- If we take $H = \mathbb{R}^d \times \mathbb{R}^d$, and $L_j : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ are given by

$$L_1(x, y) = x, \quad L_2(x, y) = y, \quad \text{and} \quad L_3(x, y) = x - y,$$

  then the resulting inequality is

$$\int_{\mathbb{R}^d} f_1(x)^{p_1} f_2(y)^{p_2} f_3(x - y)^{p_3} \, dx \, dy \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^d)}^{p_1} \|f_2\|_{L^{p_2}(\mathbb{R}^d)}^{p_2} \|f_3\|_{L^{p_3}(\mathbb{R}^d)}^{p_3}.$$

  This is a dual form of *Young's inequality*.

- If we take $H = \mathbb{R}^d$, and $L_j : \mathbb{R}^d \to \mathbb{R}^d$ is just the identity for $1 \leqslant j \leqslant m$, the inequality becomes

$$\int_{\mathbb{R}^d} f_1(x)^{p_1} \cdots f_m(x)^{p_m} \, dx \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^d)}^{p_1} \cdots \|f_m\|_{L^{p_m}(\mathbb{R}^d)}^{p_m}.$$

  This is *Holder's inequality*.

- If we take $H = \mathbb{R}^d$, and $L_j : \mathbb{R}^d \to \mathbb{R}^{d-1}$ is defined, for $1 \leqslant j \leqslant d$, to be the orthogonal projection onto the axis-oriented hyperplane orthogonal to e_j. Then the inequality, e.g. in three dimensions, says that

$$\int_{\mathbb{R}^3} f_1(y,z)^{p_1} f_2(x,z)^{p_2} f_3(x,y)^{p_3} \, dx \, dy \, dz \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^3)}^{p_1} \|f_2\|_{L^{p_2}(\mathbb{R}^3)}^{p_2} \|f_3\|_{L^{p_3}(\mathbb{R}^3)}^{p_3}.$$

  This is the *Loomis-Whitney inequality*.

The importance of the study of Brascamp-Lieb inequalities emerges from the fact that we have a very good understanding of such inequalities, and pertubations of these inequalities have proved very useful in modern harmonic analysis.

35.1   Lieb's Theorem

Take a Brascamp-Lieb datum (L, p). If we choose positive-definite linear transformations $A_j : H_j \to H_j$, and define $f_j(x) = \exp(-\pi\langle A_j x, x\rangle)$, then

$$\int_H \prod_j (f_j \circ L_j)(x)^{p_j} \, dx = \int_H \exp\left(-\pi \left\langle \sum_j p_j L_j^T A_j L_j x, x \right\rangle \right) \, dx$$

$$= \det\left(\sum_j p_j L_j^T A_j L_j \right).$$

We also have

$$\|f_j\|_{L^{p_j}(H_j)}^{p_j} = \det(p_j A_j).$$

This motivates us to define

$$\mathrm{BL}_g(L, p) = \sup_{A_1,\ldots,A_m} \frac{\det(\sum_j p_j L_j^T A_j L_j)}{\prod_j \det(p_j A_j)}.$$

This is the best constant in the Brascamp-Lieb inequality, where the inputs are restricted to centered Gaussians. Lieb's theorem says that this constant is really just $\mathrm{BL}(L, p)$, and thus we reduce computing Brascamp-Lieb inequalities to a finite-dimensional optimization problem.

Theorem 35.1 (Lieb). *$BL(L, p) = BL_g(L, p)$.*

35.2 Nonlinear Brascamp-Lieb Inequalities

A natural question is to discuss whether *nonlinear* variants of the Brascamp-Lieb inequality exist, where the surjective linear maps L_j are replaced with smooth submersions $\pi_j : H \to H_j$. It is a result of Bennett, Bez, Buschenhenke, Cowling, and Flock (2020).

Theorem 35.2. *Let $\pi_j : H \to H_j$ be C^2 submersions in a neighborhood of the origin, and let $L_j = D\pi_j(0)$. Then for any $\varepsilon > 0$, there exists a neighborhood U such that*

$$\int_U \prod_j f_j^{p_j}(\pi_j(x)) \, dx \leqslant (1 + \varepsilon) \, BL(L, p) \prod \|f_j\|_{L^1(H_j)}^{p_j}$$

for any non-negative functions $\{f_j\}$.

Key to the analysis of the linear Brascamp-Lieb inequality is Lieb's theorem, which shows $BL(L, p) = BL_g(L, p)$, and thus shows every Brascamp-Lieb functional has Gaussian near extremizers. For a particular Gaussian $g_j(x) = e^{-x^T A_j x}$, the constant in the Brascamp-Lieb inequality is

$$BL_g(L, p; A) = \frac{\prod_j \det(A_j)^{p_j/2}}{\det(\sum_j p_j L_j^* A_j L_j)^{1/2}}.$$

In order to understand the non-linear Brascamp-Lieb inequality, BBBCF must prove a more robust version of Lieb's theorem.

Theorem 35.3. *There is $N > 0$ and $\delta_0 > 0$ such that if $\delta < \delta_0$, and L is any admissable family of maps, then*

$$\inf_{A_j} BL_g(L, p; A)^{-2} \leqslant BL(L, p)^{-2} + \delta \prod_j \det(L_j)^{2p_j},$$

where the A_j range over all matrices such that both $C_j A_j C_j$ and it's inverse have operator norm at most δ^{-N}, where $C_j = |L_j|$.

The moral of this theorem is that Gaussian δ near extremizers for the Brascamp-Lieb inequality can be found.

if $C(L) = \prod_j \det(L_j)^{2p_j}$, then

$$BL(L, p) \leqslant \sup_{A_j} \left(BL_g(L, p; A)^{-2} - \delta \prod_j \det(L_j)^{2p_j} \right)^{-1/2}$$

$$= \sup_{A_j} BL_g(L, p; A) \left(1 - O_L(\delta \cdot C(L) \, BL_g(L, p; A)^2) \right),$$

where the $\{A_j\}$ range over the same values as in the theorem above. If L_0 is a fixed, admissable family of maps, then provided $\mathrm{BL}_g(L, p; A) \lesssim \delta^{-1/2}$, the right hand side is

$$\Omega\left(\mathrm{BL}_g(L, p; A)\right)$$

Chapter 36

Generalized Radon Transforms

In this chapter, we study *generalized Radon transforms* of the form

$$T_\Sigma f(x) = \int_\Sigma f(x + y) \, d\sigma,$$

where Σ^k is a smooth submanifold of $\mathbb{R}^d$, and σ is a measure supported on Σ. We are interested in $L^p \to L^q$ bounds on these operators; by Littlewood's principle, we must assume $q \geqslant p$. By Minkowski's inequality, if σ is a finite measure, we trivially find that

$$\|T_\Sigma f\|_{L^p(\mathbb{R}^d)} \leqslant |\sigma| \|f\|_{L^p(\mathbb{R}^d)}.$$

The interest is therefore on L^p *improving estimates* for the operators T_Σ. We almost mmediately see the necessity of curvature.

Theorem 36.1. *If Σ is contained in a hyperplane, then T_Σ cannot map $L^p(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$ for any $q > p$.*

Proof. Assume Σ is contained in the hyperplane $\{x_d = 0\}$. Then the operator T_Σ tensorizes, i.e. for any function f_1 on $\mathbb{R}$, and any function f_2 on $\mathbb{R}^{d-1}$, we find that

$$T_\Sigma(f_1 \otimes f_2) = f_1 \otimes T_{\Sigma'} f_2,$$

where $\Sigma' = \{x \in \mathbb{R}^{d-1} : (x, 0) \in \Sigma\}$ is the projection of Σ onto $\mathbb{R}^{d-1}$. But this clearly implies T_Σ cannot be L^p improving, since if we fix any function $f_2 \in L^p(\mathbb{R}^{d-1})$ such that $T_{\Sigma'} f_2 \neq 0$, then we find that for any function f_1,

$$\|T_\Sigma(f_1 \otimes f_2)\|_{L^q(\mathbb{R}^d)} = \|f_1\|_{L^q(\mathbb{R}^d)} \|T_{\Sigma'} f_2\|_{L^q(\mathbb{R}^d)} \gtrsim \|f_1\|_{L^q(\mathbb{R}^d)}.$$

Since an inequality of the form $\|f_1\|_{L^q(\mathbb{R}^d)} \lesssim \|f_1\|_{L^p(\mathbb{R}^d)}$ cannot hold for $q > p$, we conclude that an inequality of the form

$$\|T_\Sigma(f_1 \otimes f_2)\|_{L^q(\mathbb{R}^d)} \lesssim \|f_1\|_{L^p(\mathbb{R}^d)} \|f_2\|_{L^p(\mathbb{R}^d)}$$

cannot hold uniformly in f_1. □

We can get a more robust statement.

Theorem 36.2. *If T_Σ is bounded from $L^p(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$ for any $q > p$, then*

$$\sigma\left([-\delta/2, \delta/2] \times B^{d-1}(0)\right) \lesssim \delta^{1/p-1/q}.$$

By applying rotational and translational invariance, we conclude that most of the measure of σ cannot be contained in any hyperplane.

Proof. Set
$$f_\delta = \delta^{-1/p}\left(\mathbb{I}_{B^{d-1}(0,2)} \otimes \mathbb{I}_{[-2\delta,2\delta]}\right).$$

Then $\|f_\delta\|_{L^p(\mathbb{R}^d)} \lesssim 1$, uniformly in δ. Then for $|x'| \leq \delta$, and $|x| \leq 1$, we have

$$|Tf_\delta(x', x_d)| = \left|\int f_\delta(x' + y', x_d + y_d) d\sigma(y)\right|$$
$$\geq \delta^{-1/p} \sigma\left(B^{d-1}(0) \times [-\delta, \delta]\right).$$

Thus
$$1 \gtrsim \|Tf_\delta\|_{L^q(\mathbb{R}^d)} \gtrsim \delta^{1/q-1/p} \sigma\left(B^{d-1}(0) \times [-\delta, \delta]\right). \quad \square$$

Here are some model examples of *curved* surfaces Σ, to which we expect to obtain an L^p improving bound:

- Σ is a hypersurface with non-vanishing Gaussian curvature. Then for any hyperplane H, Σ escapes from H at a rate of

- Let γ be a smooth curve. The *osculating plane H* corresponding to γ is the span of the first $d-1$ linearly independent derivatives of γ, if this is well defined. The escape rate from the osculating plane depends on the first n such that $\gamma^{(n)}(0)$ does not lie in the osculating plane; we then have $d(\gamma(t), H) \gtrsim t^n$. If σ is a smooth measure on γ, and $\gamma(0)$ lies on the support of σ, and we assume that $H = \{x_d = 0\}$, then we find that

$$\sigma([-\delta, \delta] \times B^{d-1}(0)) \gtrsim \delta^{1/n}.$$

Thus we can still have L^p improving estimates for $1/p - 1/q \leq 1/n$.

581

- Let γ parameterize a n dimensional hypersurface. Suppose there is n such that for all $|\omega| = 1$, and $x_0 \in B_1(0)$, there is a multi-index $\alpha \in \mathbb{Z}^k$ with $|\alpha| = n$ such that $D^\alpha(\omega \cdot \gamma)(x_0) \neq 0$. Then the image of γ escapes from any hyperplane at a rate of $|x|^n$, so it is again possible to obtain L^p improving estimates for $1/p - 1/q \leqslant 1/n$.

TODO: This is really a question about convexity, i.e. we constructed a counterexample by finding sets $X \subset \mathbb{R}^d$ and $Y \subset \Sigma$ such that $X + Y \subset X$.

Theorem 36.3. *Let $\gamma : \mathbb{R}^k \to \mathbb{R}^d$ be a smooth parameterization. Assume there is $N > 0$ such that for all $\omega \in S^{d-1}$, and all $|x_0| \leqslant 1$, there exists a multi-index α with $|\alpha| \leqslant N$ such that $\nabla^\alpha(\omega \cdot \gamma)(x_0) \neq 0$. Then T_Σ is bounded from $L^2(\mathbb{R}^d)$ to $L^{\frac{2Nd}{Nd-2}}(\mathbb{R}^d)$.*

Proof. The family of operators covered in this lemma is closed under taking adjoints. Indeed, if γ parameterizes the curve averaged over in T_Σ, then T_Σ^* is obtained by averaging over $-\gamma$. We claim that

$$\mathrm{span}\{\partial^\alpha : |\alpha| \leqslant N\} = \mathrm{span}\{(v \cdot \nabla)^n : v \in S^{k-1}, 1 \leqslant n \leqslant N\}.$$

The advantage of this is that for each $\omega \in S^{d-1}$ and $|x_0| \leqslant 1$, an operator in the basis of the left hand side is non-vanishing, which implies an operator in the basis on the other side must be nonvanishing, i.e. there must exist $v \in S^{k-1}$ and $1 \leqslant n \leqslant N$ such that $(v \cdot \nabla)^n(\gamma \cdot \omega)(x_0) \neq 0$. We prove this in a lemma following this proof.

By compactness of everything, there is $c > 0$ such that we can cover S^{d-1} by balls $W_1 \cup \cdots \cup W_M$, and cover the unit ball B by balls $B_1 \cup \cdots \cup B_L$, such that for all $m \leqslant M$, and $l \leqslant L$, there exists $n_l^m \leqslant N$ and v_l^m such that for all $\omega \in W_m$, and $x \in B_l$,

$$|(\nabla \cdot v_l^m)^{n_l^m}(\omega \cdot \gamma)(x)| \geqslant c.$$

Pick a partition of unity $\psi = \sum \psi_l$ subordinate to the balls $\{B_i\}$.

It suffices to prove the theorem on a single ball $B = B_i$. Without loss of generality assume that $\gamma(0) = 0$. Given a hyperplane $H = \{x \in \mathbb{R}^d : x \cdot \omega = a\}$, we want to bound

$$|\{x \in B : d(\gamma(x), H) < \delta\}|$$

as $\delta \to 0$. We know that v and $n \leqslant N$ exists such that for $x \in B$,

$$|(\nabla \cdot v)^n(\omega \cdot \gamma)(x)| \geqslant c.$$

Rotating, without loss of generality we can assume that $v = e_1$, so that the theorem says that

$$|\partial_1^n (\omega \cdot \gamma)(x)| \geqslant c.$$

It then follows by undergraduate analysis that as $\delta \to 0$,

$$(\omega \cdot \gamma)(x_1, x') = \sum_{k=0}^{n} \frac{\partial_1^k (\omega \cdot \gamma)(0, x')}{k!} x_1^k + O(x_1^{n+1}),$$

which gives that for each $|x'| \leqslant 1$

$$|\{x_1 : |\omega \cdot \gamma(x_1, x') - a| < \delta|\}| \lesssim c\delta^{1/n},$$

and we can then apply Fubini's theorem.

Now let's prove the operator bound. We can write

$$Tf = f * \sigma,$$

where σ is a smooth, compactly supported bump function multiplied by the surface measure on Σ. Thus

$$\widehat{Tf} = \hat{f}\hat{\sigma}.$$

The operator bound will follow from a Sobolev embedding if we can show that

$$|\hat{\sigma}(\xi)| \lesssim \langle \xi \rangle^{-1/N}.$$

But this follows from stationary phase estimates, and the fact that Σ is a curve of finite type N. $\qquad\square$

Lemma 36.4. *For any N,*

$$span\{\partial^\alpha : |\alpha| \leqslant N\} = span\{(v \cdot \nabla)^n : v \in S^k, 1 \leqslant n \leqslant N\}.$$

Proof. Let V denote the left hand side, and W denote the right hand side. V is the space of all constant coefficient partial differential operators of order $\leqslant N$. The right hand side is clearly contained in V. To show the converse, it suffices to show that $\dim(W) \geqslant \dim(V)$. Since $\dim(V)$ is equal to the dimension of the space $\mathcal{P}_N$ of all multivariate polynomials on $\mathbb{R}^d$ with degree at most N, to show $\dim(W) \geqslant \dim(V)$ it suffices to construct a surjective map $T : W \to \mathcal{P}_N^*$. We do this by defining

$$(T\Lambda)(P) = (\Lambda P)(0),$$

i.e. we apply the differential operator to the polynomial, and then evaluate the operator at zero. But one can show that if $(T\Lambda)(P) = 0$ for all $\Lambda \in W$, then P vanishes on every line through the origin, and so $P = 0$. This proves that TW has no annihilator in $\mathcal{P}_N$, and thus T must be surjective. $\qquad\square$

This example is *not sharp* in general, since we only exploited the feature that the curve escaped hyperplanes *in a single direction.* The problem is that the bound

$$|\hat{\sigma}(\xi)| \lesssim \langle \xi \rangle^{-1/N}.$$

is *not* tight for general ξ, i.e. if ξ is never tangent to the given surface. One should therefore expect a much better decay rate in these directions. The problem with exploiting this approach is that this family of directions becomes very complicated, and it's very difficult to perform the resulting analysis. For years, people managed to get some results, exploiting Fourier analysis, but the results were very tied to the given dimension, i.e. most results could only be obtained in two dimensions, with some results in three dimensions. Around 2000, Michael Christ found a result, *not* relying on Fourier analytic decay, which gave a much better theory.

Theorem 36.5. *Let Σ be a graph of a smooth function $\varphi : \mathbb{R}^{d-1} \to \mathbb{R}$ with non-vanishing Hessian (i.e. so Σ has non-vanishing curvature). Then T_Σ maps $L^2(\mathbb{R}^d)$ into $L^{2d}(\mathbb{R}^d)$.*

Proof. If σ denotes the measure on Σ, then curvature implies that

$$|\hat{\sigma}(\xi)| \lesssim \langle \xi \rangle^{-\frac{d-1}{2}}.$$

$\square$

Remark. By applying a rotation, we may assume without loss of generality that $\varphi(0) = 0$ and $\nabla\varphi(0) = 0$. Taking $P_\delta = B_\delta \times [-\delta^2, \delta^2]$ and setting $f \in \mathbb{I}_{P_\delta}$ gives that

$$\frac{1}{q} \geqslant \frac{1}{p} - \frac{d-1}{d+1}.$$

Taking $N_\delta = \{x + (\varphi(t), t) : |x| < \delta\}$ and setting $f \in \mathbb{I}_{N_\delta}$ gives that

$$\frac{d}{q} \geqslant \frac{1}{p}.$$

Example. *Let $\gamma(t) = (t, t^2, \ldots, t^d)$, for $t \in [-1, +1]$. Set*

$$P_\delta = [-\delta, \delta] \times [-\delta^2, \delta^2] \times \cdots \times [-\delta^d, \delta^d].$$

and let $f = \mathbb{I}_{P_\delta}$. Then

$$\|f\|_{L^p(\mathbb{R}^d)} \sim |P_\delta|^{1/p} \sim \delta^{\frac{d(d+1)}{2}\frac{1}{p}}.$$

584

Then for $x \in P_\delta/2$, and $|t| \leq \delta/2$, $x + \gamma(t) \in P_\delta$. Thus we see that for $x \in P_\delta/2$,

$$Tf(x) \gtrsim \delta,$$

and thus

$$\|Tf\|_{L^q(\mathbb{R}^d)} \gtrsim \delta|P_\delta|^{1/q} \sim \delta^{1 + \frac{d(d+1)}{2}\frac{1}{q}}.$$

Thus to get an $L^p(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$ bound, we must have

$$1 + \frac{d(d+1)}{2}\frac{1}{q} \geq \frac{d(d+1)}{2}\frac{1}{p}.$$

Thus we get one condition on the range of p and q to which we can get a result.
   Instead, suppose that $f = \mathbb{I}_{N_\delta}$, where

$$N_\delta = \{x + \gamma(t) : t \in [-1, +1], |x| < \delta\}.$$

Then

$$\|f\|_{L^p(\mathbb{R}^d)} = |N_\delta|^{1/p} \sim \delta^{\frac{d-1}{p}},$$

whereas

$$\|Tf\|_{L^q(\mathbb{R}^d)} \gtrsim |B_\delta|^{1/q} \sim \delta^{d/q}.$$

Thus we conclude that

$$\frac{d}{q} \geq \frac{d-1}{p}.$$

Consider the moment curve in $\mathbb{R}^d$, i.e. given by the parameterization

$$\gamma(t) = (t, t^2, \ldots, t^d).$$

We claim that T is bounded from L^p to L^q. The argument is taken from Christ's 1998 paper "Convolution, Curvature, and Combinatorics: A Case Study".

Theorem 36.6 (Christ (1998)). *For any two sets $E, F \subset \mathbb{R}^d$,*

$$\langle T\, \mathbb{I}_E, \mathbb{I}_F \rangle \lesssim |E|^{\frac{2}{d+1}}|F|^{\frac{d^2+d+2}{d^2+d}}.$$

Proof. We write

$$\alpha = \frac{\langle T\, \mathbb{I}_E, \mathbb{I}_F \rangle}{|F|}$$

585

and
$$\beta = \frac{\langle T\,\mathbb{I}_E, \mathbb{I}_F\rangle}{|E|}.$$

We want to prove that
$$\beta^{2/(d+1)}\alpha^{(d^2+d+2)/(d^2+d)} \leqslant$$
$$\beta^d \alpha^{d(d+1)/2+1} \lesssim \langle T\,\mathbb{I}_E, \mathbb{I}_F\rangle.$$

We calculate that
$$\langle T\,\mathbb{I}_E, \mathbb{I}_F\rangle = \int\int \mathbb{I}_F(x)\,\mathbb{I}_E(x-\gamma(t))\,dt\,dx = |\Omega|,$$

where
$$\Omega = \{(x,t) \in \mathbb{R}^{d+1} : x \in F, x - \gamma(t) \in E\}.$$

We say $x \in F$ is 'good' if
$$|\{t : x - \gamma(t) \in E\}| \geqslant \alpha/2,$$

and bad otherwise. We observe that
$$\langle T\,\mathbb{I}_E, \mathbb{I}_F\rangle = \langle T\,\mathbb{I}_E, \mathbb{I}_{F^{\text{good}}}\rangle + \langle T\,\mathbb{I}_E, \mathbb{I}_{F^{\text{bad}}}\rangle.$$

Now
$$\langle T\,\mathbb{I}_E, \mathbb{I}_{F^{\text{bad}}}\rangle = \int_{F^{\text{bad}}} T\,\mathbb{I}_E \leqslant \int_{F^{\text{bad}}} (\alpha/2) \leqslant (\alpha/2)|F|.$$

On the other hand,
$$\langle T\,\mathbb{I}_E, \mathbb{I}_{F^{\text{good}}}\rangle \geqslant (1/2)\langle T\,\mathbb{I}_E, \mathbb{I}_F\rangle$$

$\square$

TODO: Ask Kaiyi for notes from week of February 7th.

Theorem 36.7. *Let*
$$Tf(x) = \int_{-1}^{+1} f(x - \gamma(t)),$$

where $\gamma(t) = (t, \ldots, t^d)$ is the moment curve. Then this operator, initially defined for continuous functions with compact support, satisfies a bound of the form

$$\|Tf\|_{L^{\frac{d(d+1)}{2(d-1)}}(\mathbb{R}^d)} \lesssim \|f\|_{L^{\frac{d+1}{2}}}.$$

586

Remark. A rescaling argument shows that this theorem implies that the operator

$$Tf(x) = \int_{-\infty}^{\infty} f(x - \gamma(t))$$

is bounded from $L^{\frac{d+1}{2}}(\mathbb{R}^d)$ to $L^{\frac{d(d+1)}{2(d-1)}}(\mathbb{R}^d)$. More generally, this same argument shows that establishing the result by averaging over the moment curve over times in $[-1, +1]$ implies the average over an infinite region provided one is proving estimates from $L^p(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$, where p and q are exponents satisfying

$$1 + \frac{d(d+1)}{2q} - \frac{d(d+1)}{2p} = 0.$$

Proof. We want to prove that the quantity

$$A = \sum_{j,k} 2^{j+k} \langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle$$

satisfies a bound $A \lesssim 1$ whenever

$$\sum 2^{jp}|E_j| \lesssim 1 \quad \text{and} \quad \sum 2^{kr}|F_k| \lesssim 1.$$

Here $p = (d+1)/2$, and r is the dual of the output bound we want to show, and r is the conjugate dual of $2d(d+1)/2(d-1)$. For dyadic ε, η_1, and η_2, let

$$L(\varepsilon, \eta_1, \eta_2) = \left\{ (j,k) : \langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle \sim \varepsilon |E_j|^{1/p}|F_k|^{1/r}, 2^{jp}|E_j| \sim \eta_1, 2^{kr}|F_k| \sim \eta_2 \right\}.$$

We can then write

$$A = \sum_{\varepsilon, \eta_1, \eta_2 \lesssim 1} \sum_{j,k \in L(\varepsilon, \eta_1, \eta_2)} 2^{j+k} \langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle.$$

Let $A(\varepsilon, j, k)$ be the interior sum above. Then we can write

$$A(\varepsilon, j, k) \sim \sum_{j,k \in L(\varepsilon, \eta_1, \eta_2)} 2^{j+k} \varepsilon |E_j|^{1/p}|F_k|^{1/r}$$

$$= \varepsilon \#(L(\varepsilon, \eta_1, \eta_2)) \eta_1^{1/p} \eta_2^{1/r}.$$

Now the assumptions on the choices of $|E_j|$ and $|F_k|$ imply that

$$\#(L(\varepsilon, \eta_1, \eta_2)) \lesssim \eta^{-1} \eta_2^{-1},$$

and so
$$A(\varepsilon, j, k) \lesssim \varepsilon \eta_1^{-1/p^*} \eta_2^{-1/r^*}.$$

Our goal is now to control the size of the projections of $L(\varepsilon, \eta_1, \eta_2)$ onto the two index sets j and k. We begin by stating the useful volume estimate, that

$$\sum_n |G_n| \lesssim |\bigcup_n G_n| + |\bigcup_n G_n|^{1/2} (\sum_{n \neq m} |G_n \cap G_m|)^{1/2}.$$

The result follows by applying Cauchy-Schwartz to the sum of the idnicator functions of G_n. Now set

$$\alpha_{j,k} = \frac{\langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle}{|F_k|} \quad \text{and} \quad \beta_{j,k} = \frac{\langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle}{|E_j|}.$$

There there exists sets $E_j^k \subset E_j$ and $F_k^j \subset F_k$ such that $T \, \mathbb{I}_{E_j}(x) \gtrsim \alpha_{j,k}$ for $x \in F_k^j$, $T^* \, \mathbb{I}_{F_k}(y) \gtrsim \beta_{j,k}$ for $y \in E_j^k$, and

$$\langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle \sim \langle T \, \mathbb{I}_{E_j^k}, \mathbb{I}_{F_k^j} \rangle.$$

Let
$$\mathcal{K}_j = \{k : (j,k) \in L(\varepsilon, \eta_1, \eta_2)\}.$$

Then

$$\#(\mathcal{K}_j) \varepsilon^p |E_j| \lesssim \sum_{k \in \mathcal{K}_j} |E_j^k|$$

$$\lesssim |E_j| + |E_j|^{1/2} \left(\sum_{k \neq k'} |E_j^k \cap E_j^{k'}| \right)^{1/2}.$$

Thus either
$$\#(\mathcal{K}_j) \lesssim \varepsilon^{-p} + \sum_{k \in \mathcal{K}_j} |E_j^k| \lesssim |E_j|,$$

or there is $k \neq k'$ such that

$$|E_j^k \cap E_j^{k'}| \gtrsim \varepsilon^{2p} |E_j|.$$

Similarily, either
$$\#(\mathcal{J}_k) \lesssim \varepsilon^{-r} + \sum_{j \in \mathcal{J}_k} |F_k^j| \lesssim |F_k|,$$

588

or there exists $j \neq j'$ such that $|F_k^j \cap F_k^{j'}| \lesssim \varepsilon^{2r}|F_k|$. We will show that with both the cases $\mathcal{K}_j$ and $\mathcal{J}_k$, the second case *cannot occur*. But this means we can apply Hölder's inequality to conclude that

$$\sum_{(j,k)\in\mathcal{L}_i} \langle T \, \mathbb{I}_{E_j}, \mathbb{I}_{F_k} \rangle$$

$$= \sum_{(j,k)\in\mathcal{L}_i} 2^{j+k} \langle T \, \mathbb{I}_{E_j^k}, \mathbb{I}_{F_k^j} \rangle$$

$$\lesssim \sum_{(j,k)\in\mathcal{L}_i} 2^{j+k} |E_j^k|^{1/p} |F_k^j|^{1/r}$$

$$\lesssim \left(\sum_{(j,k)\in\mathcal{L}_i} \left((2^j)|E_j^k|^{1/p} \right)^{p/\theta} \right)^{\theta/p}$$

Note
$$\varepsilon |E_j|^{1/p} |F_k|^{1/r} \lesssim \langle T \, \mathbb{I}_{E_j^k}, \mathbb{I}_{F_k^j} \rangle \lesssim |E_j^k|^{1/p} |F_k|^{1/r} \lesssim |E_j|^{1/p} |F_k|^{1/p}.$$

To prove this, we may assume that L is $O(\log(1/\varepsilon))$ separated, where the implicit constant is allowed to depend on d. Now (TODO: Think about this) either

$$\sum_k |E_j^k| \lesssim |E_j|$$

or there are k and k', not equal to one another, such that

$$\sum |E_j^k \cap E_j^{k'}| \gtrsim \varepsilon^{2p} |E_j|.$$

589

Or similar statements for $\{F_j^k\}$. We claim this cannot occur. Indeed,

$$\sum_{(j,k)\in L} 2^{j+k}\langle T\,\mathbb{I}_{E_j}, \mathbb{I}_{F_k}\rangle \sim \sum_{(j,k)} 2^{j+k}\langle T\,\mathbb{I}_{E_j^k}, \mathbb{I}_{F_k^j}$$

$$= \sum_{(j,k)\in L} 2^{j+k}|E_j^k|^{1/p}|F_k^j|^{1/r}$$

$$\leqslant \left(\sum_{(j,k)\in L}(2^{jp}|E_j^k|)^{1/\theta}\right)^{\theta/p}\left(\sum_{(j,k)\in L}(2^{kr}|F_k^j|)^{1/\theta}\right)^{\theta/r}$$

$$\leqslant \left(\sum_j 2^{jp/\theta}\left(\sum_j(2^{jp}|E_j|)\right)^{1/\theta}\right)^{\theta/p} \times (ANALOGOUS)$$

$$\leqslant (\eta_1^{1/\theta-1})^{\theta/p}(\eta_2^{1/\theta-1})^{\theta/r}\left(\sum_j 2^{jp}|E_j|\right)^{\theta/p}\left(\sum_k 2^{kr}|F_k|\right)^{\theta/r}$$

(TODO: Think about this) THE IMPORTANT THING HERE IS THAT WE ARE IN AN L^p improving setting, which allows us to pick $\theta < 1$, which gives us more room to play with.

Now set $\Omega_{j,k} = \{(t,x) : x - \gamma(t) \in E_j\}$. Then

$$|\Omega_{jk}| \sim \langle T\,\mathbb{I}_{E_j}, \mathbb{I}_{F_k}\rangle.$$

Set $\alpha_{j,k} = |\Omega_{j,k}|/|F_k|$ and $\beta_{j,k} = |\Omega_{j,k}|/|E_j|$. We setup an iterating scheme (TODO: What is this) like before. We get

$$|E_j| \gtrsim \alpha_{jk}^{\frac{d(d-1)}{2}+1}\beta_{jk}^{d-1}\left(\frac{\alpha_{j'k}}{\alpha_{jk}}\right)^2.$$

On the other hand, because $\langle T\,\mathbb{I}_{E_j}, \mathbb{I}_{F_k}\rangle \gtrsim \varepsilon|E_j|^{1/p}|F_k|^{1/r}$,

$$|E_j| \lesssim \varepsilon^{-\frac{d(d+1)}{2}}\alpha_{jk}^{\frac{d(d-1)}{2}+1}\beta^{d-1}.$$

Thus

$$\alpha_{j'k} \lesssim \varepsilon^{-\frac{d(d+1)}{4}}\alpha_{jk}.$$

By symmetry, the reverse holds. But this means that

$$\frac{\alpha_{jk}}{\alpha_{j'k}} = \frac{\langle T\,\mathbb{I}_{E_j}, \mathbb{I}_{F_k}\rangle}{\langle T\,\mathbb{I}_{E_j}, \mathbb{I}_{F_k}\rangle} \sim \frac{2^{-j}}{2^{-j'}} = 2^{j'-j}.$$

590

So

$$\varepsilon^{\frac{d(d-1)}{4} \lesssim 2^{j'-j}} \lesssim \varepsilon^{-\frac{d(d+1)}{4}}.$$

This gives a contradiction for large values because

$$|j - j'| \gtrsim \log(1/\varepsilon).$$

$\square$

Now we discuss higher dimensional averages.

Theorem 36.8. *Let*

$$Tf(x) = \int_{\mathbb{R}^{d-1}} f(x - |t|^2, x' - t) \, dt$$

Then T is bounded from $L^{(d+1)/2,1}$ to $L^{d+1,\infty}$, i.e.

$$\langle T \, \mathbb{I}_E, \mathbb{I}_F \rangle \lesssim |E|^{\frac{d}{d+1}} |F|^{\frac{d}{d+1}}.$$

Proof. Set

$$\alpha = \frac{\langle T \, \mathbb{I}_E, \mathbb{I}_F \rangle}{|F|} \quad \text{and} \quad \beta = \frac{\langle \mathbb{I}_E, T^* \, \mathbb{I}_F \rangle}{|E|}.$$

We want to convert the bound involving $\langle \cdot, \cdot \rangle$ into more harmless things involving α and β, i.e. we want to prove the inequality

$$\alpha^{\frac{d}{d+1}} \beta^{\frac{1}{d+1}} \lesssim |E|^{\frac{d-1}{d+1}}$$

and

$$\alpha^d \beta \lesssim |E|^{d-1}.$$

We want to set up an iterated TT^* argument as before to generate a map with a nice Jacobian determinant. The problem here is that we have a $d - 1$ dimensional parameter set, and $d - 1$ does not divide d. To solve this issue, the exponent on the right hand side of the inequality above, i.e. E is d dimensional, and we take it to an exponent $d - 1$ to get $d(d - 1)$. Define

$$F^{\text{good}} = \{x \in F : T \, \mathbb{I}_E(x) \geq \alpha/2\}$$

As before, we have

$$\langle T \, \mathbb{I}_E, \mathbb{I}_{F^{\text{good}}} \rangle \geq (1/2) \langle T \, \mathbb{I}_E, \mathbb{I}_F \rangle$$

591

so it is natural to define

$$E^{\text{good}} = \{y \in E : T^* \, \mathbb{I}_{F^{\text{good}}}(y) \geqslant \beta/4\}.$$

Then

$$\langle T \, \mathbb{I}_{E^{\text{good}}}, \mathbb{I}_{F^{\text{good}}} \rangle \geqslant (1/4)\langle T \, \mathbb{I}_E, \mathbb{I}_F \rangle.$$

Let $y_0 \in E^{\text{good}}$ (assuming E^{good} is nonempty). Let

$$\mathcal{S}_{y_0} = \{s \in \mathbb{R}^{d-1} : y_0 + (|s|^2, s) \in F^{\text{good}}\},$$

Then $|\mathcal{S}_{y_0}| \geqslant \beta/4$. If $s \in \mathcal{S}_{y_0}$, then

$$\mathcal{J}_{y_0} = \{t \in \mathbb{R}^{d-1} : x(s) - (|t|^2, t) \in E\}.$$

Then $|\mathcal{J}_{y_0}(s)| \geqslant \alpha/2$. Set

$$\mathcal{U}_{y_0} = \{(s_j, t_1, \ldots, t_{d-1}) : s \in \mathcal{S}_{y_0}, t_j \in \mathcal{J}_{y_0}(s)\}.$$

Define $\Psi : (\mathbb{R}^d)^d \to (\mathbb{R}^d)^{d-1}$ by setting

$$\Psi(s,t)_j = (y_0 + (|s|^2, s) - (|t_j|^2, t)).$$

The Jacobian of Ψ_{y_0} is

$$\det \begin{pmatrix} 2s^T & -2t_1^T & 0 \\ I_{d-1} & I_{d-1} & 0 \\ \ldots & \ddots & \ldots \\ 2s^T & 0 & -2t_{d-1}^T \\ I_{d-1} & \ldots & I_{d-1} \end{pmatrix}.$$

A cofactor expansion gives that, with a constant depending on d, this determinant is proportional to the determinant of the matrix with rows $s - t_i$. Now

$$\begin{aligned}
|E|^{d-1} &\gtrsim |\Psi_{y_0}(U_{y_0})| \\
&\gtrsim \int_{U_{y_0}} |\det D\Psi_{y_0}(s,t)| \, ds \, dt \\
&\gtrsim \int_{\mathcal{S}_{y_0}} \int_{\mathcal{J}_{y_0}(s)^{d-1}} |\det(t_1 - s, \ldots, t_{d-1} - s)| \, dt \, ds \\
&= \int_{\mathcal{S}_{y_0}} \int_{(\mathcal{J}_{y_0}(s) - s)^{d-1}} |\det(u_1, \ldots, u_{d-1})| \, du \, ds.
\end{aligned}$$

Let us consider the special case where $\mathcal{J}_{y_0}(s) - s$ is a ball of volume roughly α centered at the origin. Then it is a ball with radius about $\alpha^{1/(d-1)}$. Then for most points, the determinant is at least α, and so in this special case, we get that

$$|E|^{d-1} \geqslant |\mathcal{S}_{y_0}||\mathcal{J}_{y_0}(s)|^{d-1}\alpha \gtrsim \beta\alpha^d.$$

If $\mathcal{J}_{y_0}$ is more diffuse away from s, the determinant can, unfortunately, be be bigger. We want to be able to measure how eccentric this set is, i.e. to approximate this set by an ellipsoid. $\qquad\square$

Now we complete the proof of the nonlinear Brascamp-Lieb inequality. To do this, we perform an induction on scales, using a special family of functions. For $R > 1$, and $\mu > 0$, we say a function $f \geqslant 0$ is R constant at a scale μ on $\Omega \subset \mathbb{R}^d$ if $f(x) \leqslant Rf(y)$ for $|x - y| \leqslant \mu$. We let $L^1(\Omega; \mu, R)$ be the family of non-negative, integrable such functions. Let $C(x_0, \delta, \mu, R)$ be the best constant in the inequality

$$\int_{B_\delta(x_0)} \prod_j f_j^{q_j} \circ \pi_j(x) \, dx \leqslant C \prod_j \|f_j\|_{L^1}^{q_j}.$$

We wish to get a bound on C that is independent of μ.

We fix parameters $1 < \alpha < 2$ and $0 < \beta < 2 - \alpha$,

$$0 < \gamma < \min\left(\frac{2 - \alpha - \beta}{(d+2)\alpha}, \frac{\beta}{2\alpha}\right)$$

and

$$0 < \tau < \gamma/N,$$

which is the log of the eccentricity of δ near extremizers for the Brascamp-Lieb inequality, and

$$0 < \sigma < \min(\beta - 2\alpha\gamma, \alpha\tau).$$

We now linearly approximate the maps $\{\pi_j\}$. Set

$$L_j^{x_0}(x) = \pi_j(x_0) + d\pi_j(x_0)(x - x_0).$$

Then π is in C^2, so that $|\pi_j(x) - L_j^{x_0}(x)| \lesssim |x - x_0|^2$ for x and x_0 contained in a small enough neighborhood.

We now prove the base case for our induction on scales. The result is what we desire, but the result depends on μ.

593

Theorem 36.9. *There is $\nu > 0$ such that if $x_0 \in B_\nu(0)$ and $R > 1$, $0 < \delta < \nu$, and $\mu > 0$, and $\delta^{\alpha+\beta} \leqslant \mu$, then*

$$C(x_0, \delta, \mu, R) \leqslant BL(d\pi, q).$$

Proof. Let $\rho = \sum_j q_j$. We may assume $|\pi_j(x) - L_j^{x_0}(x)| \lesssim \delta^2$ for all $x \in B_{2\delta}(x_0)$. Thus

$$f_j(\pi_j(x)) \leqslant R f_j(L_j^{x_0}(x)),$$

so

$$\int_{B_\delta(x_0)} f_j^{q_j} \circ \pi_j(x) \leqslant R^\rho \int_{B_\delta(x_0)} \prod_j f_j^{q_j}(L_j^{x_0}(x))\, dx,$$

and then the theorem follows. $\qquad\square$

The inductive step is where the bulk of the proof lies.

Theorem 36.10. *There is $\nu > 0$ such that if $x_0 \in B_\nu(0)$, $R > 1$, $0 < \delta < \nu$, $\mu > 0$, and $\delta 6\alpha + \beta > \mu$,*

$$C(x_0, \delta, \mu, R) \leqslant (1 + \delta^\sigma) \max_{x \in B_{2\delta}(x)} C(x, \delta^\alpha, \mu, R\exp(\delta^\alpha)).$$

The scale δ^α is much smaller than the scale δ, which enables us to apply induction, and get something like a geometric series.

Let us see how this theorem gives us the main result. For any $\varepsilon > 0$, there exists $\delta_0(\varepsilon)$ such that

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leqslant (1 + C\varepsilon)\, BL(d\pi(0), q).$$

Take $\delta_0 < \nu/3$, where ν is sufficiently small that the base case and recursive inequality holds. Let $\mu > 0$. If $\delta_0^{\alpha+\beta} \leqslant \mu$, we win. So assume $\delta_0^{\alpha+\beta} > \mu$. Define $\delta_k = \delta_0^{\alpha^k}$. Then as long as $\delta_k^{\alpha+\beta} > \mu$,

$$C(x_0, \delta_k, \mu, R) \leqslant (1 + \delta_k^\sigma) \max_{x \in B_{2\delta_k}(x_0)} C(x, \delta_{k+1}, \mu, R\exp(\delta_k^\sigma)).$$

Let k_* be the smallest integer such that $\delta_{k_*}^{\alpha+\beta} \leqslant \mu$. Then

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leqslant \prod_{k \leqslant k_* - 1} (1 + \delta_k^\sigma) C(x_0, \delta_*, \mu, (1 + \varepsilon) \prod_{k \leqslant k_* - 1})(1 + \delta_k^\sigma) \exp(\delta_k^\sigma)$$

594

As μ goes to zero, k_* can be made arbirarily large. So in general, this is bounded by the product over all positive integers k, i.e.

$$C(0, \delta_0, \mu, 1 + \varepsilon) \leqslant \prod_k (1 + \delta_0^{\sigma \alpha^k}) \exp(\rho \delta_0^{SOMETHING \cdot \alpha^k}) \leqslant (1 + \varepsilon)$$

provided we make δ_0 sufficiently small.

Now let's prove the inductive claim. Define

$$M_{x), \delta} = \sum_j q_j d\pi_j(x_0)^T A_{x_0, \delta, j} d\pi_j(x_0),$$

which arises when we evaluates $\mathrm{BL}_g(d\pi(x_0), q; A_{x_0, \delta})$. We claim that there is $\nu > 0$ such that

$$\sup_{x_0 \in B_\nu(0)} \|M_{x, \delta}^{-1/2}\| \lesssim \delta^{-\tau_N}$$

for all $\delta \in (0, \nu)$. To prove this, we compute that

$$\sum_j q_j \langle A_j d\pi_j(x_0) v, d\pi_j(x_0) v \rangle \gtrsim \delta^{2\tau_N} \sum_j |d\pi_j(x_0) v|^2 \gtrsim \delta^{2\tau_N}.$$

Now take a Gaussian $g_{x_0, \delta}$. It is not locally constant, so we must truncate it so that it is locally constant. Define

$$g_{x_0, \delta}(x) = (\delta^{1+\gamma})^{-d_j} g_{A_{x_0, \delta, j}}(x/\delta^{1+\gamma}).$$

Then $g_{x_0, \delta}$ is still a near extremizer.

36.1 Non Translation-Invariant Averaging Operators

Based on work by Christ-Nagel-Stein-Wainger, and Tao-Wright, and Gressman.

Consider a function $a : \mathbb{R}^{d+k} \to [0, \infty)$, which is smooth, and has compact support, and let

$$\pi : \mathbb{R}_x^d \times \mathbb{R}_t^k \to \mathbb{R}^d$$

be such that $D_x \pi$ is non-singular, and $D_t \pi$ has rank k. Then π is a diffeomorphism in the x variable, and for each fixed x, $t \mapsto \pi(x, t)$ traces out the image of a k-dimensional submananifold of $\mathbb{R}^d$. We consider an operator of the form

$$Tf(x) = \int_{\mathbb{R}^k} a(x, t) f(\pi(x, t)) \, dt.$$

We have been studying translation-invariant examples where $\pi(x, t) = x - \gamma(t)$. A basic example is given by letting π trace out the solutions to the equation $x \cdot y = 1$.

Any operator T of the form above maps $L^p(\mathbb{R}^d)$ to $L^p(\mathbb{R}^d)$ for any $1 \leqslant p \leqslant \infty$ (just apply Schur's Lemma). The fact that we have introduced a compactly supported cutoff a tells us that $L^p \to L^q$ estimates implies $L^{p'} \to L^{q'}$ estimates for any $p' \geqslant p$ and $q' \leqslant q$. Thus, even though it is now possible to have L^p worsening estimates, we have essentially proved all such estimates from the fact that T is bounded from L^p to L^p. Thus we still are focusing on L^p improving estimates.

There is a simple formula for the adjoint of T. Namely, a change of variables $y = \pi(x, t)$ gives that

$$\langle Tf, g \rangle = \int_{\mathbb{R}^k} \int_{\mathbb{R}^d} a(x, t) f(\pi(x, t)) \overline{g(x)}$$

$$= \int_{\mathbb{R}^k} \int_{\mathbb{R}^d} f(y) \left(a^*(y, t) f(y) g(\pi_t^{-1}(y)) \right)^*.$$

Let us write $\pi^*(y, t) = \pi_t^{-1}(y)$. Then we conclude that

$$T^* g(y) = \int_{\mathbb{R}^k} a^*(y, t) g(\pi^*(y, t)) \, dt.$$

Thus the adjoint of an averaging operator is also an averaging operator.

We will now switch to studying the *double fibration formulation* of the problem. We consider quantities of the form

$$B(f_1, f_2) = \int a(z) f_1(\pi_1(z)) f_2(\pi_2(z)) \, dz,$$

where $z = (y, t)$, and where π_1 and π_2 are submersions from $\mathbb{R}^{d+k}$ to $\mathbb{R}^d$, such that for each $x \in \mathbb{R}^d$, and for any $i \neq j$, π_i, restricted to the k dimensional submanifold $\pi_j^{-1}(x)$, is an immersion (i.e. the level sets of π_1 and π_2 are k-dimensional submanifolds that only intersect at discrete points).

Theorem 36.11. *(Baby version of CNSW / TW) Suppose there is a smooth manifold Σ of positive codimension such that for all $z \in \Sigma$, $\pi_j^{-1}\{\pi_j(z)\} \subset \Sigma$. Then no L^p improving estimates are possible.*

Proof. Suppose Σ is n dimensional. If $\Sigma_j = \pi_j(\Sigma)$, then Σ_j is a smooth manifold of dimension $n - k$. If f_j is a characteristic function of a δ neighborhood of Σ_j, then

$$\|f_j\|_{L^{p_j}(\mathbb{R}^d)} \sim \delta^{\frac{d+k-n}{p_j}}.$$

Then $B(f_1, f_2) \gtrsim \delta^{d+k-n}$ (the defining integrand is $\gtrsim 1$ on a δ neighborhood of Σ). Thus we can only have

$$B(f_1, f_2) \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^d)} \|f_2\|_{L^{p_2}(\mathbb{R}^d)}$$

if $d + k - n \geq (d + k - n)(1/p_1 + 1/p_2)$, i.e. if $1/p_1 + 1/p_2 \leq 1$. But this implies we can have no L^p improving estimates. $\qquad \square$

The same conclusion holds if the π_1 and π_2 fibres are both tangent to Σ to infinite order at some point in Σ. The result of CNSW shows that we *do* get some L^p improvement if we have a quantitative guarantee of the rate of escape from every manifold.

In the translation-invariant setting, where $\pi_1(x, t) = x - \gamma(t)$ and $\pi_2(x, t) = x$, where $\gamma(0) = 0$ and γ is contained in a hyperplane H_0, then $\Sigma = H_0 \times \mathbb{R}_t$ gives a result to which the theorem above applies.

Theorem 36.12. *Our goal is now to show that for $1/p_1 + 1/p_2 > 1$, we have a bound of the form*

$$|B(f_1, f_2)| \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^d)} \|f_2\|_{L^{p_2}(\mathbb{R}^d)}.$$

Proof. We reduce our analysis to a restricted weak type ienquality. Let $f_j = \mathbb{I}_{E_j}$ for some sets E_j. Then

$$|B(f_1, f_2)| = |\Omega|,$$

where

$$\Omega = \pi_1^{-1}(E_1) \cap \pi_2^{-1}(E_2),$$

and our goal is to show that

$$|\Omega| \lesssim |E_1|^{1/p_1} |E_2|^{1/p_2},$$

i.e. that

$$|\Omega| \lesssim |\pi_1(\Omega)|^{1/p_1} |\pi_2(\Omega)|^{1/p_2}.$$

We can rearrange this to say that

$$\alpha_1^{1/p_1} \alpha_2^{1/p_2} = \left(\frac{|\Omega|}{|\pi_1(\Omega)|} \right)^{1/p_1} \left(\frac{|\Omega|}{|\pi_2(\Omega)|} \right)^{1/p_2} \lesssim |\Omega|^{1/p_1 + 1/p_2 - 1}.$$

Raising things to the power of $(1/p_1 + 1/p_2 - 1)^{-1}$, we have

$$\alpha^b = \alpha_1^{b_1} \alpha_2^{b_2} \lesssim |\Omega|.$$

597

As in previous proofs, it is very useful to *parameterize* the fibres.

Consider a vector field

$$
X_j = \det \begin{pmatrix}
\frac{\partial}{\partial x^1} & \cdots & \frac{\partial}{\partial x^d} \\
\frac{\partial \pi_j^1}{\partial x^1} & \cdots & \frac{\partial \pi_j^1}{\partial x^d} \\
\vdots & \ddots & \vdots \\
\frac{\partial \pi_j^{d-1}}{\partial x^1} & \cdots & \frac{\partial \pi_j^{d-1}}{\partial x^d}
\end{pmatrix}.
$$

In the language of differential geometry, we note that if $dy = dy^1 \wedge \cdots \wedge dy^{d-1}$, and $dx = dx^1 \wedge \cdots \wedge dx^d$, then

$$
df \wedge \pi_j^*(dy) = df \wedge d\pi_j^1 \wedge \cdots \wedge d\pi_j^{d-1} = X_j(f) \cdot dx.
$$

We immediately see from this equation that $X_j(\pi_j^k) = 0$ for $1 \leqslant k \leqslant d-1$. Thus the flow of the vector field X_j preserves the level sets of π_j. Moreover, the quantity $X_j(f)$, if nonzero, measures the ability for the system of functions $\{f, \pi_j^1, \ldots, \pi_j^{d-1}\}$, which will then form a coordinate system, to preserve volume – i.e. the volume gets more distorted in the coordinate system the closer $X_j(f)$ gets to zero.

We now introduce the Newton polygon. Define

$$
\Psi_{x_0}(t_0, \ldots, t_d) = e^{t_d X_1} \cdots e^{t_1 X_1}(x_0),
$$

and

$$
\Phi_{x_0}(t_0, \ldots, t_d) = e^{t_d X_{d+1}} \cdots e^{t_1 X_2}(x_0),
$$

where $X_3 = X_1$, $X_4 = X_2$, $X_5 = X_1$, and so on and so forth. The rank of $D\Psi_{x_0}$ and $D\Phi_{x_0}$ is at most two, which is *bad*. So we set

$$
\mathcal{A}_{x_0} = \{\alpha : \partial_t^\alpha \{\det D\Psi_{x_0}\}\}(0) \neq 0\}
$$

and

$$
\mathcal{B}_{x_0} = \{\beta : \partial_t^\alpha \{\det D\Phi_{x_0}\}(0) \neq 0\}.
$$

We define the *degrees*

$$
d_A(\alpha) = (\alpha_1 + \alpha_3 + \cdots + [d/2], \alpha_2 + \alpha_4 + \cdots + [d/2]).
$$

and

$$
d_B(\beta) = .
$$

598

The reason for these degrees is that if $\mathcal{P} = [-\delta_0, \delta_0] \times \cdots \times [-\delta_d, \delta_d]$, for sufficiently small δ, then

$$
\begin{aligned}
|\Omega| &\gtrsim \int_{\mathcal{P}} |\det D\Psi_{x_0}(t)| \, dt \\
&\gtrsim \int_{\mathcal{P}} |t^\alpha| |\partial_t^\alpha D\Psi_{x_0}(t)| \, dt \\
&\gtrsim \delta^\alpha \left(\delta_1^{[d/2]} \delta_2^{[d/2]} \right) \\
&\gtrsim \delta^{d_A(\alpha)}.
\end{aligned}
$$

We now introduce the *Newton polygon* $\mathcal{P}_{x_0}$ to be the convex hull of the set of the union of the quadrants

$$
d_A(\alpha) + [0, \infty)^2 \quad \text{and} \quad d_B(\beta) + [0, \infty)^2
$$

for $\alpha \in \mathcal{A}_{x_0}$ and $\beta \in \mathcal{B}_{x_0}$. Here is an invariant way to define the polygon. For a word $\omega \in \{1, 2\}^n$, define X_ω recursively, such that

$$
X_{i\omega} = [X_i, X_\omega].
$$

Then the Newton polygon is the convex hull of the union of the quadrants $d(\omega_1) + \cdots + d(\omega_d) + [0, \infty)^2$, where $\omega \in \{1, 2\}^d$ ranges over indices such that $\det(X_{\omega_1}, \ldots, X_{\omega_d})$. One can view the polygon as measuring the degree to which the Frobenius theorem fails for the manifold in question.

The theorem of Tao and Wright states that the Newton-Polygon is *necessary* for the averaging operator to be bounded, i.e. in order for a bound of the form $\alpha^b \lesssim |\Omega|$, the vector b *must* lie in the Newton polyhedron.

$\square$

LEt us make the assumption that the exponential maps $(t, x) \mapsto e^{tX_j}(x)$ is a polynomial of degree at most N.

Theorem 36.13. *Let $S \subset \mathbb{R}$ be measurable. Then for all $N > 0$, and all $\varepsilon > 0$, there exists an interval J such that $|J \cap S| \gtrsim |S|$ such that for all polynomials f of degree N,*

$$
\int_S |P(t)| \, dt \gtrsim \sum_{j=1}^N \|P^{(j)}\|_{L^\infty(J)} \left(\frac{|J|}{|S|} \right)^{1-\varepsilon} |S|^{j+1}.
$$

599

Proof. WIthout loss of generality, we can assume that $|S| < \infty$, since otherwise the theory is trivial. Otherwise, let S be contained in an interval I. We claim that that for any S, and any I, there exists subintervals K and J of I, such that $|K| \sim |J| \sim d(K,J)$, $|J \cap S| \gtrsim |S|$, and $|K \cap S| \gtrsim (|S|/|K|)^\varepsilon |S|$. We can iterate this procedure to construct intervals $\{K_i\}$ and $\{J_i\}$ such that K_i and J_i are subintervals of I, K_i and J_i are both contained in J_{i-1}, $|K_i| \sim |J_i| \sim d(K_i, J_i)$, $|J_i \cap S| \gtrsim S|$, and $|K_i \cap S| \gtrsim (|S|/|K_i|)^\varepsilon |S|$.

Write

$$P(t) = A \prod_{j=1}^{N} (t - \xi_j).$$

Suppose that ξ_j is at a distance $\gtrsim |J_N|$ from J_N. Then P is roughly constant on J_N, and so

$$\int_S |P(t)| \, dt \geqslant \int_{S \cap J_N} |P(t)| \, dt$$
$$\sim \frac{|S|}{|J_N|} \int_{J_N} |P(t)| \, dt$$
$$\sim \frac{|S|}{|J_N|} \sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty(J_N)} |J_N|^{j+1}$$
$$= \sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty(J_N)} |J_N|^j |S|.$$

On the other hand, suppose that $d(\xi_j, J_N) \lesssim |J_N|$. Then

$$\|P_j\|_{L^\infty(J_N)} \leqslant |P(\xi_j)| + |\int_{\text{Length } O(|J_N|) \text{ Interval}} P'(z)| \, dz$$
$$\lesssim |J_N| \|P'\|_{L^\infty(J_N)}$$

Now

$$\sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty(I_i)} |I_i|^j \lesssim \sum_{j=1}^{N} \|P^{(j)}\|_{L^\infty(I_i)} |I_i|^j$$

for all $I_i = K_i$ or $I_i = J_i$. Observe that for all $j < d(\xi_j, K_i) \lesssim |K_i|$, we can find i

such that $d(\xi_j, K_i) \gtrsim |K_i|$. Then

$$
\begin{aligned}
\int_S |P(t)|\, dt &\geqslant \int_{S \cap K_i} |P(t)|\, dt \\
&\sim \frac{|S \cap K_i|}{|K_i|} \int_{K_i} |P(t)|\, dt \\
&\gtrsim \frac{|S \cap K_i|}{|K_i|} \sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty(K_i)} |K_i|^{j+1} \\
&\gtrsim \sum_{j=0}^{N} \|P^{(j)}\|_{L^\infty} (|K_i|/|S_i|)^{j+1} (|S|/|K_i|)^{1+\varepsilon} |S|^{j+1}.
\end{aligned}
$$

To prove the existence of K and J as required above, we split I into the unit of four intervals $I_0^1, \ldots, I_0^4$, each with equal lengths. If $|I^k \cap S| \gtrsim_\varepsilon (|S|/|I^k|)^\varepsilon |S|$ for $= 1$ and $k = 4$, then set $J = I^1 \cup I^2$ or $J = I^3 \cup I^4$, chosen so that $|J \cap S| \geqslant |S|/2$, and $K = I^4$ or $K = I^1$, chosen so that K and J are not adjacent. This concludes the proof in this case. Otherwise, in case two we assume that $|I_0^k \cap S| \lesssim_\varepsilon (|S|/|I_0^k|)^\varepsilon |S|$ for $k = 1$ or $k = 4$. We define I_1 by removing this 'bad interval' from I_0. We thus continue iterating this procedure, writing $|I_i| = (3/4)^i |I|$, such that

$$
|I_{i+1} \cap S| \geqslant (1 - C_\varepsilon(|S|/|I_i|)^\varepsilon)|I_i \cap S| \geqslant \prod_{j=0}^{i} (1 - C_\varepsilon(|S|/|I_i|)^\varepsilon)|S|.
$$

We get a contradiction for C_ε sufficiently small. Thus we only need to iterate this procedure $O(1)$ times before we reach case one. $\qquad \square$

Now we can show the sufficiency of the Tao and Wright theorem. We want to show that

$$
|\Omega| \gtrsim \alpha_1^{b_1} \alpha_2^{b_2},
$$

where $\alpha_1 = |\Omega|/|\pi_j(\Omega)|$, and b lies in the Newton polyhedron. We apply the coarea formula, writing

$$
\begin{aligned}
|\Omega| &= \int_{\pi_j(\Omega)} \int_{t=\pi_i^{-1}(y)} |X_j|^{-1}\, dt\, dy \\
&= \int_{\pi_j(\Omega)} \int e^{tX_j}(\sigma_j(y))\, dt\, dy.
\end{aligned}
$$

Here σ_j is a section of π_j. Now α_j is equal to the average length of a fibre. We now apply the method of refinements. To do this, we need to define what it means for a point to be 'good'. We will write

$$\Omega = \Omega_0 \supset \cdots \supset \Omega_{d+1},$$

For a set E, define $E^{\flat_j} = (\sigma_j \circ \pi_j)(E)$. We now 'throw away fibres of length $\ll \alpha_j$', i.e. defining $\Omega^{(i)}$ to be the set of all points $\{e^{tX_j}(x^{\flat_j})\}$, where the set $\{s : e^{sX_j}(x^{\flat_j})\}$ has measure greater than α_j, and t lies in $S^j_{x^{\flat_j}} \cap J^j(S_{x^{\flat_j}})$, where S and J are as in the polynomial Lemma we have just proven. Let $\mathcal{P}_{x_0}$ be the set of $t \in \mathbb{R}^d$ such that t_j lies in $S^j_{x^{\flat_j}} \cap J^j$. Then $\Phi_{x_0}(\mathcal{P}_{x_0}) \subset \Omega$. Thus

$$
\begin{aligned}
|\Omega| &\gtrsim \\
&\int_{\mathcal{P}} |\mathrm{Det}(D\Psi_{x_0})| \\
&\gtrsim \sum_j \int_{\mathcal{P}_{x_0}} \| |\partial^j \mathrm{Det}(D\Phi_{x_0})(t', \cdot)\|_{L^\infty(J(S_j(x_0,t_1,\dots,t_{d-1})))} |S_d|^{j+1} \, dt' \\
&\gtrsim \int_{\mathcal{P}_{x_0}} \sum_j |\partial_{t_d}^j \det(D\psi_{x_0}(t',0))| \alpha_d^{j+1} \, dt'.
\end{aligned}
$$

If we now iterate this procedure, we get

$$\gtrsim_\beta \alpha^{d_A(\beta)} |\det(D\Psi_{x_0})|.$$

Our goal is to show that

$$|\Omega| \gtrsim \min_{x_0 \in \overline{U}} \sum_\beta \alpha^{d(\beta)} |\partial^\beta \mathrm{Det}(D\Phi_{x_0})|.$$

But if $b \in \bigcap_{x_0 \in U} \mathcal{P}_{x_0}$, then we can do some arithmetic-geometric mean magic to get what we want.

Let us briefly talk about how we deal with the C^∞ case of the problem. We consider a large order Taylor approximation of Φ_{x_0}, We want to say that $|\mathrm{Det}(D\Psi_{x_0})|$ is proportional to $|\mathrm{Det}(D\mathcal{P}_{x_0})|$, when $|\mathrm{Det}(D\mathcal{P}_{x_0})| \gtrsim \|\mathcal{P}_{x_0} - \Phi_{x_0}\|$. We want to conclude that this assumption happens most of the time. Now $\|\mathcal{P}_{x_0} - \Phi_{x_0}\| < (\alpha_1 + \alpha_2)^\omega$, while $|\mathrm{Det}(D\mathcal{P}_{x_0})| \gtrsim \alpha^\beta$ by the nondegeneracy hypothesis. But we have a problem, because the inequality $\alpha^\beta \gtrsim (\alpha_1 + \alpha_2)^\omega$ can fail. But we can cut up Ω into $(\alpha_1 \alpha_2)^\varepsilon$ diameter sets where we do get the result. But this ε means the proof of the endpoint result *fails* in the endpoint case.

602

How about the necessity of the Tao-Wright condition. If b does *not* lie in the Newton polygon, then we can find $v \in (0\, infty)$ such that $b^0 v < b^1 v$. We then pick Ω to be an appropriate flow by Φ_{x_0} or Ψ_{x_0} (whichever is bigger), and derive a contradiction to the restricted weak-type inequality as a result.

Let us also discuss the multilinear version. This is the result of Christ, Wainger, Stein, ... But this result is not sharp, and we wish to analyze the multiinear setting.

36.2 Brascamp Lieb Inequality

TODO: Add notes from Kaiyi from Thursday, March the 9th.

Theorem 36.14 (BCCT)**.** *We have a bound of the form*

$$\left| \int_{\mathbb{R}^d} \prod_j f_j(L_j x) a(x)\, dx \right| \lesssim \prod_j \|f_j\|_{L^{p_j}(\mathbb{R}^d)}$$

where $L_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ are linear, surjective maps, if, for $q_j = 1/p_j$, we have, for any subspace $V \subset \mathbb{R}^d$,

$$codim(V) \geqslant \sum_j q_j codim(L_j v).$$

Proof. Let $B_{\mathrm{loc}}(L, q)$ be the optimal implicit constant in the Brascamp-Lieb inequality. Last time (TODO: See Kaiyi's notes), we proved that the condition in the Theorem implies that

$$B_{\mathrm{loc}}(L, q) + d = \sum_j q_j d_j.$$

Let us show that the converse also holds. Consider the two convex sets

$$Q = \{q \in [0, 1]^k : \mathrm{codim}(V) \geqslant \sum_j q_j \mathrm{codim}(L_j v) \text{ for all } V \subset \mathbb{R}^d\}$$

and

$$\mathcal{R} = \{q \in [0, 1]^k : B_{\mathrm{loc}}(L, q) < \infty\}.$$

We have shown that $Q \subset \mathcal{R}$. Both are convex sets, so it suffices to show every extreme point of Q is in $\mathcal{R}$. Let $q \in Q$ be an extreme point. The set Q is *compact* – it is actually a polytope, the intersection of a *finite* family of hyperplanes, since each equation defining Q has bounded, integer coefficients, and there any only finitely many such equations. We split our analysis into two cases:

- $q_i = 0$ for some i. Then we can perform an inductive argument.

- $q_i = 1$ for some i. Suppose without loss of generality that $i = k$. The inequality for $V = \text{Ker}(L_k)$ is

$$d_k \geqslant \sum_{j=1}^{k-1} q_j \text{codim}(L_j V) + d_k.$$

  Thus the equation is an equality, which reduces to the third case.

- q is an extreme point of one of the inequalities defining the convex set, there exists at least two subspaces V_0, V_1 such that the associated equation

$$\text{codim}(V_l) = \sum_j q_j \text{codim}(L_j V_l)$$

  is an equality for $l \in \{0, 1\}$. Let's call these subspaces *critical*. The BCCT condition is then true when we replace $\mathbb{R}^d$ with V. Write $\mathbb{R}^d = V_0 \oplus V_0^\perp$, and $\mathbb{R}^{d_j} = W_j \oplus W_j^\perp$, where $W_j = L_j(V_0)$. Set $l_j : V_0^\perp \to W_j^\perp$. Then the BCCT condition also holds for these maps.

  Consider $f_1, \ldots, f_k$, and assume without loss of generality that these functions are non-negative. Then

$$T(f_1, \ldots, f_k) = \int_{\mathbb{R}^d} \prod_j f_j(L_j x) a(x) \, dx$$

$$\leqslant \int_{V_0^\perp} \int_{V_0} \prod_j f_j(l_j u + L_j v) a_1(v) a_2(u) \, dv \, du.$$

If we define, for $x \in V_0^\perp$,

$$F_j(x) = \left(\int_{V_0} |f_j(x + L_j v)|^{p_j} \right)^{1/p_j},$$

then Brascamp-Lieb on V_0 (by induction) implies that

$$|T(f_1, \ldots, f_k)| \lesssim \int_{V_0^\perp} \prod_j F_j(l_j u) \, a_2(u) \, du \lesssim \prod_j \|f_j\|_{L^{p_j}}.$$

We have proved our argument by induction, i.e. eiver reducing d or reducing k. The base cases that remain to be proved are when $k = 1$, or when $d = 1$. The case $k = 1$ is a linear operator, and if we take the BCCT condition where V to be the kernel of L, we get that

$$d_1 \geqslant q_1 d_1,$$

so that $q_1 \leqslant 1$, so we just use Hölder's inequality. For $d = 1$, take the BCCT condition where $V = \{0\}$, which gives $1 \geqslant \sum q_j$, so we can again just apply Hölder's inequality. $\qquad\square$

There are two natural ways to generalize the Brascamp-Lieb inequality. First, we can consider the nonlinear Brascamp-Lieb inequality, i.e. where we replace linear surjective maps with smooth submersions $\pi_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$. Bennett-Carbery-Wright showed that, if $\mathbb{R}^d = \bigoplus K_j$, where K_j is the kernel of $L_j = D\pi_j(0)$. The proof uses the method of refinements – it should be clear what the iterated map is, which should be flown along fibres of the map. This gives a weak type bound, which can be improved to a strong type bound by using the tensor power trick.

Things should get more transverse if we perturb the maps π_j, since a generic family of maps are transverse to one another. In fact, we should have more inequalities. This is because if

$$\mathrm{Codim}(V) \geqslant \sum_j q_j \mathrm{codim}(L_j V).$$

Most pertubations of V will make the quantities on the right side smaller. That's exactly what happens in the Tao-Wright inequality, since we have new interations between vector fields, that give new inequalities. Thus we're hopeful. But the method of 'slicing subspaces' that we used above completely fails in the nonlinear case. For some time it seemed like one can find a way to use the method of refinements, but there were some obstructions to using this method. We will give the broof by BBBBCF (Bennett, ..., Flock), using induction on scales – the idea is to take an inequality we want to prove, split it into two inequalities, a 'zoomed in version', and a 'zoomed out version'. The zoomed in version is easier since things are more linear, and in the zoomed-out version things get blurred.

The smallest possible value of $\mathrm{codim}(L_j V) = \min(d_j - \dim V, 0)$, so we should actually get

$$\mathrm{Codim}(V) \geqslant \sum_j$$

TODO: Ask Kaiyi for her notes.

605

Theorem 36.15 (Ball's Inequality). *Let $f, g \in L^1(\mathbb{R}^{d_1}) \times \cdots \times L^1(\mathbb{R}^{d_k})$, and let*

$$BL(L, q; f) = \frac{\int \prod_j f_j^{q_j}(L_j x) \, dx}{\prod_j \|f_j\|_{L^1}^{q_j}}.$$

Then

$$\mathcal{BL}(L, q; f)\mathcal{BL}(L, q; g) \leqslant \left(\sup_x \mathcal{BL}(L, q, h^x) \right) \mathcal{BL}(L, q, f * g),$$

where

$$h_j^x(z) = f_j(z_j) g_j(L_j x - z_j).$$

Proof. Without loss of generality, we may normalize our functions so that $\|f_j\|_{L^1} = \|g_j\|_{L^1} = 1$ for all j. Then

$$\begin{aligned}
\mathcal{BL}(L, q; f)\mathcal{BL}(L, q; g) &= \int \prod_j f_j^{q_j}(L_j y) g_j^{q_j}(L_j x - L_j y) \, dy \, dx \\
&= \int \int \prod_j (h_j^x)^{q_j}(L_j y) \, dy \, dx \\
&= \int \mathcal{BL}(L, q; h^x) \prod_j \|h_j^x\|_{L^1}^{q_j} \, dx \\
&\leqslant \sup_x \mathcal{BL}(L, q; h^x) \int \prod_j (f_j * g_j)^{q_j}(L_j x) \, dx \\
&= \sup_x \mathcal{BL}(L, q; h^x)\mathcal{BL}(L, q; f * g).
\end{aligned}$$

$\square$

As a corollary, if g is an L^1 normalized δ near extremizer of a Brascamp-Lieb inequality, then

$$(1 - \delta)\mathcal{BL}(L, q; f) \leqslant \min \left\{ \sup_x \mathcal{BL}(L, q; h^x), \mathcal{BL}(L, q; f * g) \right\},$$

where h^x is defined as above.

As another corollary, the family of extremizers to a Brascamp-Lieb inequality is closed under convolution, and so in particular, if non-empty, includes the Gaussians (by the central limit theorem).

606

We will assume Lieb's theorem , and use it to prove a more sharpened form. Lieb's theorem says that

$$\mathcal{BL}(L,q)^2 = \inf_A \mathcal{BL}_g(L,q;A)^2 = \inf_A \frac{\det(\sum_j q_j L_j^T A_j L_j)}{\prod_j (\det A_j)^{q_j}}.$$

Here A_j are positive definite matrices. To get more intuition out of this theorem, let $A_j = R_j^T X_j R_j$, where X_j is diagonal and positive definite and with entries $\lambda(j,r)$. Set $D = d_1 + \cdots + d_k$, $n(j,l) = d_1 + \cdots + d_{j-1} + l$ for $1 \leqslant j \leqslant k$ and $1 \leqslant l \leqslant d$. Let $v_{n(j,l)} = L_j^T R_j^T e_l^j$, where $e_l^j \in \mathbb{R}^{d_j}$ is a unit vector. Then

$$\sum_j q_j L_j^T A_j L_j = \sum_j q_j \sum_{n=d_{j-1}}^{d_j} e^{\lambda_n} v_n v_n^T = \sum_{n=1}^D q_j e^{\lambda_n} v_n v_n^T = (q_1 e^{\lambda_1} v_1, \ldots, q_D e^{\lambda_D} v_D)(v_1, \ldots, v_D)^T.$$

The Cauchy-Binet formula implies that

$$\det(\sum_j q_j L_j^T A_j L_j) = \sum_{I \in \mathcal{J}} \det(q_1^E e^{\lambda_{i_1}} v_{i_1}, \ldots, q_{i_d}^E e^{\lambda_{i_d}} v_{i_d})(v_{i_1}^T, \ldots, v_{i_d}^T)$$

let $n(j,l) = d_1 + \cdots + d_{l-1+l}$

$$\inf_{\substack{\|C_j^T A_j C_j\| \leqslant \delta^{-N} \\ \|(C_j^T A_j C_j)^{-1}\| \leqslant \delta^{-N}}} \mathcal{BL}(L,q;A) \leqslant \mathcal{BL}(L,q)^2 + \delta \prod_j \det(L_j L_j^T)^{q_j},$$

where $C_j = (L_j L_j^T)^{1/2}$.

36.3 Analyst Friendly Proof of Endpoint Multilinear Kakeya

We consider families of tubes $J_1, \ldots, J_d$ in $\mathbb{R}^n$, where a tube is a 1-neighborhood of a line in a direction $e(T) \in S^{d-1}$.

in a direction e. We assume that for all tubes $T_j \in \mathcal{J}_j$, $|e(T_j) - e_j| < c_0 \ll 1$. Then the endpoint multilinear Kakeya theorem says that for all $q \geqslant 1/(d-1)$, and $2 \leqslant d \leqslant n$,

$$\int_{\mathbb{R}^d} (\prod_j \sum_{T_j \in \mathcal{J}_j} a_j \chi_{T_j})^q \lesssim_{d,q} \prod_j \|a_j\|_{l^1}^q.$$

The original geometric proof given by Guth. This is a proof due to Carbery-Valdimarsson.

Let us consdier a special case where $e(T_j) = e_j$ for all $T_j \in \mathcal{J}_j$. If $q = 1/(d-1)$, and $T_j = \pi_j^{-1}(B_j)$, then the result follows from the Loomis-Whitney inequality. On the other hand, this special case implies Loomis Whitney, so the multilinear Kakeya theorem is stronger than the Loomis-Whitney inequality.

Chapter 37

Maximal Averages Over Curves

37.1 Averages over a Parabola

Given any measurable function $f : \mathbb{R}^2 \to \mathbb{C}$ 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 $\|Mf\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}$ 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 $k \in \mathbb{Z}$, 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

$$\|\tilde{M}_k\|_{L^p(\mathbb{R}^2) \to L^p(\mathbb{R}^2)} = \|\tilde{M}_0\|_{L^p(\mathbb{R}^2) \to L^p(\mathbb{R}^2)}$$

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 $f \in L^2(\mathbb{R}^2)$ and $\hat{f}$ 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

$$\|\tilde{M}_0 f\|_{L^2(\mathbb{R}^2)} = \|\tilde{m}\hat{f}\|_{L^2(\mathbb{R}^2)} \lesssim \|\hat{f}\|_{L^2(\mathbb{R}^2)} = \|f\|_{L^2(\mathbb{R}^2)}.$$

On the other hand, we can decompose $\mathbb{R}^2 - E_0$ into
suppose $\hat{f}$ 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, $\hat{f}$ is supported on the ellipsoid

$$\left\{(\xi, \eta) \in \mathbb{R}^2 : \xi^2/2 + \eta^2/20 \leqslant 1\right\}.$$

Thus the uncertainty principle implies that f is roughly constant on scales $|\Delta x|^2 \leqslant 1/2$ and $|\Delta y|^2 \leqslant 1/20$,

$$s$$

$$\phi(x) = \frac{1}{(1 + 2x^2 + 20y^2)^N}$$

Example. *If M is a hypersurface in $\mathbb{R}^d$, and ψ is a smooth, compactly supported function on M, and f is a smooth, compactly supported function on $\mathbb{R}^d$, we can define a function Af on $\mathbb{R}^d$ 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

$$\widehat{Af}(\xi) = \hat{f}(\xi)\widehat{\psi d\sigma}(\xi).$$

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

$$\widehat{(Af)_\alpha} = (2\pi i\xi)^\alpha \hat{f}(\xi)\widehat{\psi\sigma}(\xi).$$

Since we have shown

$$|\widehat{\psi\sigma}(\xi)| \lesssim |\xi|^{-(d-1)/2},$$

we conclude that if $|\alpha| \leqslant k$, where $k = (d-1)/2$,

$$\|(Af)_\alpha\|_{L^2(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^d)}.$$

In particular, this implies that A extends to a unique bounded operator from $L^2(\mathbb{R}^d)$ to $L_k^2(\mathbb{R}^d)$, i.e. to a map such that for each $f \in L^2(\mathbb{R}^d)$, Af is a square integrable function which has square integrable weak derivatives of all orders less than or equal to k, and moreover, $\|(Af)_\alpha\|_{L^2(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^d)}$ for all $|\alpha| \leqslant k$. Thus the operator A is 'smoothening', in a certain sense.

The operator A is obviously bounded from $L^1(\mathbb{R}^d)$ to $L^1(\mathbb{R}^d)$ and $L^\infty(\mathbb{R}^d)$ to $L^\infty(\mathbb{R}^d)$, 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 $L^p(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$, 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 $\gamma_s = s\ldots(s + N)e^{s^2}$, 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 $\psi_0(x) = \psi(x)(1 + |\nabla_{x'}\phi(x')|^2)^{1/2}$.

Part IV

Pseudodifferential and Fourier Integral Operators

The theory of the Fourier transform enabled the early harmonic analysts to study various linear partial differential operators, like the heat and wave equations. The theory of pseudodifferential and Fourier integral operators allows us to study *variable coefficient* analogues of these operators on the Fourier side of things, just like in the linear case. In particular, we can appropriately *localize* Euclidean Fourier analysis so that it can used in the study of problems on differentiable manifolds.

Pseudodifferential Operators are very useful for the study of *elliptic differential operators*. But it is difficult to apply the theory of such operators to problems that are genuinely non-elliptic, like the Cauchy problem for a hyperbolic equation. This is because pseudodifferential operators are *microlocal*, i.e. the positions of singularities are preserved, whereas the solution operators to the Cauchy problem for hyperbolic are *not*; solutions propogate through space. The theory of Fourier integral operators extends the class of pseudodifferential operators to model such phenomena.

Chapter 38

Symbol Classes

The theory of tempered distributions allows one to build a useful, general theory of expressions of the form

$$\int_{\mathbb{R}^d} a(x)e^{2\pi i \xi \cdot x}\, d\xi,$$

where a is a tempered locally integrable function on the real line. We will define pseudodifferential operators and Fourier integral operators by very similar expressions, i.e. of the form

$$\int_{\mathbb{R}^p} a(x,\theta)e^{2\pi i \phi(x,\theta)}\, d\theta.$$

However, the useful properties of pseudodifferential and Fourier Integral Operators are more sensitive to the regularity of the input function a, i.e. the growth of a and all of it's derivatives. We are thus lead to restrict our analyses where a lies in some given *symbol class*, which quantifies the growth of a and it's derivatives.

Fix $\mu \in \mathbb{R}$, $0 \leqslant \rho \leqslant 1$ and $0 \leqslant \delta \leqslant 1$, and an open subset $\Omega \subset \mathbb{R}^p$, we define the symbol class $\mathcal{S}^\mu_{\rho,\delta}(\Omega \times \mathbb{R}^p)$ to be the class of all smooth functions $a \in C^\infty(\Omega \times \mathbb{R}^p)$ such that

$$\left| D_x^\alpha D_\theta^\beta a(x,\theta) \right| \lesssim_{\alpha,\beta} \langle \theta \rangle^{\mu - \rho|\beta| + \delta|\alpha|}.$$

The optimal choice of constants in these inequalities can be taken as seminorms, which give $\mathcal{S}^\mu_{\rho,\delta}(\Omega \times \mathbb{R}^p)$ the structure of a Fréchet space. The most common spaces that occur in practice are the symbol classes $\mathcal{S}^\mu_{1,0}(\Omega \times \mathbb{R}^p)$, which we simply denote by $\mathcal{S}^\mu(\Omega \times \mathbb{R}^p)$. It is also somewhat common to consider classes where $\rho + \delta = 1$, and so we let $\mathcal{S}^\mu_\rho(\Omega \times \mathbb{R}^p) = \mathcal{S}^\mu_{\rho,1-\rho}(\Omega \times \mathbb{R}^p)$. It is also natural to consider the

classes $\mathcal{S}^{\mu}_{\rho,\delta,\mathrm{loc}}(\Omega \times \mathbb{R}^p)$, where the bounds above hold *locally uniformly* in the x-variable. We set

$$\mathcal{S}^{\infty}_{\rho,\delta} = \bigcup_{\mu} \mathcal{S}^{\mu}_{\rho,\delta} \quad \text{and} \quad \mathcal{S}^{-\infty}_{\rho,\delta} = \bigcap_{\mu} \mathcal{S}^{\mu}_{\rho,\delta},$$

natural because the spaces $\mathcal{S}^{\mu}_{\rho,\delta}$ are nested, increasing in μ. We can then give this space the topology induced by the direct limit. Elements of $\mathcal{S}^{\mu}_{\rho,\delta}$ are called *symbols* of order μ, of type (ρ, δ).

Example. *Consider a function of the form*

$$a(x,\xi) = \sum_{|\alpha| \leqslant n} c_\alpha(x) \cdot \xi^\alpha,$$

where $\{c_\alpha\}$ are a family of functions in $C^{\infty}_{loc}(\mathbb{R}^d_x \times \mathbb{R}^d_\xi)$. Then $a \in S^{\mu}_{loc}(\mathbb{R}^d_x \times \mathbb{R}^d_\xi)$. If the functions $\{c_\alpha\}$ and all their derivatives are bounded functions, then $a \in S^{\mu}(\mathbb{R}^d_x \times \mathbb{R}^d_\xi)$. This is true, in particular, if the functions $\{c_\alpha\}$ all lie in $C^{\infty}_c(\mathbb{R}^d_x)$.

It is also natural for harmonic analysis to study the behaviour of homogeneous functions under the Fourier transform. The next example shows such functions are symbols, and thus the study of such functions is encompassed in the more general theory here.

Example. *Fix a complex number $\mu \in \mathbb{C}$. If $a \in C^{\infty}(\Omega \times \mathbb{R}^p)$, and is homogeneous for large θ, i.e. there exists a constant $R_0 > 0$ such that for $|\theta| \geqslant R_0$, and $\lambda \geqslant 1$, $a(x, \lambda\theta) = \lambda^\mu a(x, \theta)$. Then a is a symbol of order $Re(\mu)$, i.e. $a \in \mathcal{S}^{Re(\mu)}_{loc}(\Omega \times \mathbb{R}^p)$; To see this, assume first that, without loss of generality, $R_0 \geqslant 1$. For any α and β, smoothness implies that for $|\theta| \leqslant 1$, we have*

$$|D^{\alpha}_x D^{\beta}_\theta a(x,\theta)| \lesssim_{\alpha,\beta} 1 \lesssim \langle\theta\rangle^{Re(\mu)},$$

where the implicit constant is locally uniform in the x-variable. Differentiating both sides of the equation $a(x, \lambda\theta_0) = \lambda^\mu a(x, \theta_0)$, we conclude that for $|\theta_0| \geqslant 1$,

$$\lambda^{|\beta|}(D^{\alpha}_x D^{\beta}_\theta a)(x, \lambda\theta_0) = \lambda^\mu (D^{\alpha}_x D^{\beta}_\theta a)(x, \theta_0).$$

If $|\theta| \geqslant 1$, then we can write $\theta = \langle\theta\rangle\theta_0$ for $|\theta_0| = 1$, which implies that

$$|(D^{\alpha}_x D^{\beta}_\theta a)(x,\theta)| = \left|\langle\theta\rangle^{\mu-|\beta|}(D^{\alpha}_x D^{\beta}_\theta a)(x,\theta_0)\right| \lesssim \langle\theta\rangle^{Re(\mu)-|\beta|},$$

where the implicit constant is locally uniform in x.

615

Example. *Suppose a is non-isotropically homogeneous, i.e. we can write $\mathbb{R}^p = \mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p_m}$, such that for $|\theta| \geqslant 1$, and $\lambda \geqslant 1$,*

$$a(x, \lambda^{a_1}\theta_1, \ldots, \lambda^{a_m}\theta_m) = \lambda^a a(x, \theta).$$

If $\mu = \max_j m/m_j$, and $\rho = \min m_j/m_k$, then a is a symbol of order μ, and type $(\rho, 0)$, i.e. $a \in \mathcal{S}^\mu_{\rho,0}(\Omega \times \mathbb{R}^p)$.

Example. *Symbols often arise from the asymptotic study of variable-coefficient stationary phase integrals. The classic example is the Fourier transform of the spherical measure. Since σ is radially symmetric, so is it's Fourier transform. Thus we can write*

$$\hat{\sigma}(\xi) = G(|\xi|),$$

for some function $G : [0, \infty) \to \mathbb{C}$. If we set $\lambda = |\xi|$, and $\xi_0 = \xi/|\xi|$, then

$$G(\lambda) = \int_{|x|=1} e^{-2\pi i \xi \cdot x} \, dx$$
$$= \int_{|x|=1} e^{-2\pi i \lambda \xi_0 \cdot x} \, dx.$$

This is an oscillatory integral, with two stationary points, when $x = \xi_0$, and when $x = -\xi_0$. These stationary points are nondegenerate. Thus, for each ξ_0, we have an asymptotic expansion of the form

$$G(\lambda) \sim \sum_{j=0}^{\infty} \left(c_{j,+} e^{2\pi i\lambda} + c_{j,-} e^{-2\pi i\lambda} \right) \lambda^{-\frac{d-1}{2}-j},$$

in the sense that for $\lambda \geqslant 1$,

$$\left| \left(\frac{d}{d\lambda} \right)^N \left\{ G - \sum_{j<N} (c_{j,+} e^{2\pi i\lambda} + c_{j,-} e^{-2\pi i\lambda}) \lambda^{-\frac{d-1}{2}-j} \right\} \right| \lesssim_N \lambda^{-\frac{d-1}{2}-j-N}.$$

Suppose $\chi \in C_c^\infty(\mathbb{R}^d)$ is radially symmetric, and equal to one in a neighborhood of the origin. Then if we set

$$a_{j,+}(\xi) = c_{j,+}(1 - \chi(\xi)) \cdot |\xi|^{-\frac{d-1}{2}-j},$$

and

$$a_{j,-}(\xi) = c_{j,-}(1 - \chi(\xi)) \cdot |\xi|^{-\frac{d-1}{2}-j},$$

616

then $a_{j,+}$ and $a_{j,-}$ are both radial symbols of order $-(d-1)/2 - j$. The asymptotic formula for G above implies that if we pick radial symbols a_+ and a_- of order $-(d-1)/2$ such that

$$a_+ \sim \sum_j a_{j,+} \quad and \quad a_- \sim \sum_j a_{j,-},$$

then

$$\hat{\sigma}(\xi) \sim a_+(\xi)e^{2\pi i|\xi|} + a_-(\xi)e^{-2\pi i|\xi|},$$

in the sense that the difference between the left and right hand side is a symbol of order $-\infty$.

It is simple to see that a differential operator $D_x^\alpha D_\theta^\beta$ is a continuous, bounded linear operator from $\mathcal{S}_{\rho,\delta}^\mu$ to $\mathcal{S}_{\rho,\delta}^{\mu-\rho|\beta|+\delta|\alpha|}$.

Example. *If $\chi \in \mathcal{S}(\mathbb{R}^d)$, and a is smooth and for $|\theta| \geqslant 1$,*

$$a(x,\theta) = \chi\left(|\theta|^\varepsilon x\right).$$

Then a is a symbol of order zero, and type $(1,\varepsilon)$. Indeed, it can be verified by induction that $D_\theta^\beta a$ is a finite sum of functions of the form

$$x^n s(x,\theta)\phi(|\theta|^\varepsilon x)$$

where s is a symbol of order at most $-n(1-\varepsilon) - m$, and $n + m = |\beta|$. It can then be checked that $D_x^\alpha D_\theta^\beta a$ is a finite sum of terms of the form

$$P(x)s(x,\theta)\phi(|\theta|^\varepsilon x)$$

where P is a polynomial vanishing to order k at the origin, s is a symbol of order at most $-n(1-\varepsilon) - m + \varepsilon(|\alpha| - n + k)$, and ϕ is Schwartz, where $n + m = |\beta|$ and $n - |\alpha| \leqslant k \leqslant n$. We thus have

$$\left|D_x^\alpha D_\theta^\beta a(x,\theta)\right| \lesssim_N \max_{n,m,k} |x|^k \langle\theta\rangle^{-n(1-\varepsilon)-m+\varepsilon(|\alpha|-n+k)} \langle|\theta|^\varepsilon x\rangle^{-N}$$

$$= \max_k \left(|x|^k \langle\theta\rangle^{\varepsilon k} \langle|\theta|^\varepsilon x\rangle^{-N}\right) \langle\theta\rangle^{\varepsilon|\alpha|-|\beta|}.$$

For $|\theta| \geqslant 1$, this quantity is

$$\lesssim \max_k \langle|\theta|^\varepsilon x\rangle^{k-N} \langle\theta\rangle^{\varepsilon|\alpha|-|\beta|}.$$

Choosing $N = n$, we conclude that

$$\left| D_x^\alpha D_\theta^\beta a(x,\theta) \right| \lesssim_{\alpha,\beta} \langle \theta \rangle^{\varepsilon |\alpha| - |\beta|}.$$

Note that for $|\theta| \geqslant 1$, the function a concentrates itself in the region $|x| \lesssim |\theta|^{-\varepsilon}$, which, by the mean value theorem, is only possible for a symbol of type (ρ, ε) if $\delta \geqslant \varepsilon$.

Example. *If a and ϕ are smooth, and for $|\theta| \geqslant 1$, the function*

$$a(x,\theta) = \exp\left(2\pi i \phi(x) |\theta|^{1-t}\right)$$

is a symbol of order zero, and type $(t, 1-t)$. This makes sense since, taking Taylor expansions, near a point (x_0, θ_0), the function a oscillates in the x variable at a frequency at most $|\theta_0|^{1-t}$, and the function a oscillates in the θ variable at a frequency $|\theta_0|^{-t}$. This makes us expect the symbol to be of type $(t, 1-t)$. We can check this fairly simply; by induction, $D_x^\alpha D_\theta^\beta a(x,\theta)$ is a finite sum of terms of the form

$$s(x,\theta) \exp\left(2\pi i \phi(x) |\theta|^{1-t}\right),$$

where s is a symbol of order at most $|\alpha|(1-t) - |\beta| t$.

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 $t \in \mathbb{R}$, there exists N_0 such that for $N \geqslant N_0$, $a - \sum_{n=0}^N a_n$ is a symbol of order t. If a_n is a symbol of order t_n, and $\lim_{n \to \infty} t_n = -\infty$, then a symbol a always exists satisfying these asymptotics.

Theorem 38.1. *Consider a sequence of symbols $\{a_n\}$, with $a_n \in \mathcal{S}^{t_n}(\Omega \times \mathbb{R}^p)$, where $\lim_{n \to \infty} a_n = -\infty$, and let $t = \max t_n$. Then there exists a symbol $a \in \mathcal{S}^t(\Omega \times \mathbb{R}^p)$ such that $a \sim \sum a_n$.*

Proof. Fix a bump function $\phi \in \mathcal{D}(\mathbb{R}^p)$ 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

$$\left| \nabla_x^j \nabla_\lambda^k \{ \phi(\theta/r_n) a_n(x,\theta) \} \right| \leqslant 2^{-n} \langle \theta \rangle^{t_n + 1 - k}$$

618

for $x \in \Omega$, 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 $x \in \Omega$, we find that

$$|\nabla_x^j \nabla_\lambda^k R_N(x, \theta)| \lesssim_{N,i,j} \langle \theta \rangle^{\max_{n \geqslant N} t_n + 1 - k}.$$

Thus $R_N \in \mathcal{S}^{\beta_N}(\Omega \times \mathbb{R}^p)$, where $\beta_N = \max_{n \geqslant N} t_n + 1$. 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 $E_N \in \mathcal{S}^{-\infty}(\Omega \times \mathbb{R}^p)$. $\qquad\square$

Remark. A similar result holds for local families of symbols.

A symbol

$$a \sim \sum_{n=0}^{\infty} a_n$$

is *not* uniquely defined by it's asymptotic expansion; for any $r \in \mathcal{S}^{-\infty}(\Omega \times \mathbb{R}^p)$, we have

$$a + \psi \sim \sum_{n=0}^{\infty} a_n.$$

Nonetheless, this is most general way that uniqueness fails to hold – any two symbols satisfying the same asymptotic expansion differ by an element of $\mathcal{S}^{-\infty}(\Omega \times \mathbb{R}^p)$. For this reason, it is often natural in many contexts to work with the space $\dot{\mathcal{S}}^t(\Omega \times \mathbb{R}^p)$, the quotient space of $\mathcal{S}^t(\Omega \times \mathbb{R}^p)$ by $\mathcal{S}^{-\infty}(\Omega \times \mathbb{R}^p)$. In this space of symbols, asymptotic expansions are unique.

To verify asymptotic formulae, the following Lemma is often helpful.

Lemma 38.2. *Suppose $a \in C^\infty(\Omega \times \mathbb{R}^p)$, and for any $n, m > 0$, there exists t_{nm} such that*

$$|\nabla_x^n \nabla_\theta^m a(x, \theta)| \lesssim_{n,m} \langle\theta\rangle^{t_{nm}}.$$

If, for any $t \in \mathbb{R}$,

$$|a(x, \theta)| \lesssim_t \langle\theta\rangle^t,$$

then $a \in \mathcal{S}^{-\infty}(\Omega \times \mathbb{R}^p)$.

Proof. We begin by showing that for $f \in C^2(\mathbb{R})$,

$$\|f'\|_{L^\infty(\mathbb{R})} \leqslant \sqrt{2\|f\|_{L^\infty(\mathbb{R})}\|f''\|_{L^\infty(\mathbb{R})}}.$$

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

$$|f'(x)| \leqslant \varepsilon^{-1}\|f\|_{L^\infty(\mathbb{R})} + (\varepsilon^{-1}/2)\|f''\|_{L^\infty(\mathbb{R})}.$$

Optimizing our choice of ε 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. $\qquad\square$

Corollary 38.3. *Suppose $\{a_n\}$ are a family of symbols, with $a_n \in \mathcal{S}^{t_n}(\Omega \times \mathbb{R}^p)$ for each n, and $\lim_{n\to\infty} t_n = -\infty$. Then if $a \in C^\infty(\Omega \times \mathbb{R}^p)$, and for each N and M, there exists t_{NM} such that*

$$|\nabla_x^N \nabla_\theta^M a(x, \theta)| \lesssim \langle\theta\rangle^{t_{NM}}.$$

If for each n, there exists β_n such that

$$\left|a(x,\theta) - \sum_{k=0}^{n} a_n(x,\theta)\right| \lesssim_n \langle\theta\rangle^{\beta_n},$$

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

$$a(x,\theta) \sim \sum_{n=0}^{\infty} a_{t-n}(x,\theta),$$

where $a_{t-n} \in S^{\mathrm{Re}(t)-n}(\Omega \times \mathbb{R}^p)$ is homogeneous of degree $t-n$ in the θ variables for suitably large θ (the symbol must be smooth, and so cannot be homogeneous for small θ unless it is a polynomial). We denote the class of such symbols of order t by $\mathcal{S}^t_{\mathrm{cl}}(\Omega \times \mathbb{R}^p)$. Sometimes this class is also called the class of *polyhomogeneous symbols*, and denoted $\mathcal{S}^t_{\mathrm{phg}(\Omega \times \mathbb{R}^p)}$. We can also consider polyhomogeneous symbols with non integer step sizes, i.e. the class $\mathcal{S}^{t,h}_{\mathrm{phg}}(\Omega \times \mathbb{R}^p)$, i.e. those symbols that satisfy an asymptotic expansion of the form

$$a(x,\theta) \sim \sum_{n=0}^{\infty} a_{t-hn}(x,\theta)$$

where $a_{t-hn} \in S^{\mathrm{Re}(t)-hn}(\Omega \times \mathbb{R}^p)$ is homogeneous of degree $t - hn$. But the case $h = 1$ is often of the most interest.

Remark. Let M be a manifold, and let E be a vector bundle over M. We can define the space $\mathcal{S}^t_{\mathrm{loc}}(E)$ of symbols of order t on E to be the family of all scalar functions a on E which are symbols of order t in local coordinates. These have a very similar theory of the theory we have expounded above. In particular, one can consider asymptotic developments of symbols.

Chapter 39

Oscillatory Integral Distributions

The operators we will discuss in these notes will be Schwartz operators, whose kernels are defined by certain kinds of formal expressions of the form

$$I_{a,\phi}(x) = \int a(x,\theta) e^{2\pi i \phi(x,\theta)} \, d\theta.$$

which we will call an oscillatory integral. Here Ω is some open subset of $\mathbb{R}^d$, a is a *symbol* lying in some class $\mathcal{S}^\mu_{\mathrm{loc}}(\Omega \times \mathbb{R}^p)$, and $\phi : \mathbb{R}^d \times \mathbb{R}^p - \{0\} \to \mathbb{R}$ is some smooth function, which we call the *phase*. If $\mu \leqslant -p$, the integrand is Lebesgue integral, and so we can interpret the definition of $I_{a,\phi}$ pointwise as a Lebesgue integrable function. For $\mu > -p$, one cannot interpret the integral above as a Lebesgue sense, and we must find another way to formally define the integral. In this chapter, we describe the basic theory through which we can rigorously interpret such formulas.

We will make several assumptions on ϕ so that as θ increases in magnitude, the phase behaves in a way to make the integrand more and more oscillatory. By making sense of the cancellation of this oscillation when θ has large magnitude, we will be able to make sense of the expression $I_{a,\phi}$, as a distribution. In order to guarantee this kind of oscillation, we do need to make some assumptions on ϕ. The canonical assumptions to make are the following:

- ϕ is *homogeneous* of degree one in the θ variable.

- The functions $\nabla_x \phi$ and $\nabla_\theta \phi$ have no common zeroes on μ-supp(a).

In this chapter, by a *phase function*, we will mean a function with these properties.

These are the easiest phases to handle. In practice, there are other distributions which arise by oscillatory integrals not satisfying these assumptions; for instance, the fundamental solution Φ of the Schrödinger operator

$$\partial_t - (i/2\pi)\Delta$$

on $\mathbb{R}^d$ is naturally expressed by the oscillatory integral

$$\Phi(x,t) = \int e^{2\pi i(\xi \cdot x - |\xi|^2)} \, d\xi,$$

which has a phase with a quadratic component. It is possible to extend some of the results here to such other phase, but the degree one case is simplest to analyze, and still features much of the complexity dealt with in the study of other operators.

Theorem 39.1. *Fix $\chi \in C_c^\infty(\mathbb{R}^p)$, equal to one in a neighborhood of the origin. The distributions*

$$I_{a,\phi,R} = \int a(x,\theta)\chi(\theta/R)e^{2\pi i\phi(x,\theta)} \, d\theta$$

converge in the weak $*$ topology as $R \to \infty$ to a distribution $I_{a,\phi}$ on Ω, independent of the choice of χ. For $\mu \geqslant -p$, the distribution $I_{a,\phi}$ has order at most the smallest integer larger than $p + \mu$. For a fixed phase function ϕ, the map $a \mapsto I_{a,\phi}$ is continuous from $\mathcal{S}^\infty(\Omega \times \mathbb{R}^p)$ to $\mathcal{D}(\Omega)^*$, equipped with the weak $*$ topology.

Proof. If $\mu < -p$, the result follows from the dominated convergence theorem, and moreover, the distribution is continuous the result gives a distribution of order zero. Now fix $f \in \mathcal{D}(\Omega)$ with support in a compact set $K \subset \Omega$. Our goal is to show that

$$\langle I_{a,\theta,R}, f \rangle = \int a(x,\theta)\chi(\theta/R)e^{2\pi i\phi(x,\theta)} \, d\theta$$

converges as $R \to \infty$ to a quantity independent of the choice of χ, and that moreover, this quantity is upper bounded by a constant depending on only finitely many derivatives of f. Let us write

$$a_R(x,\theta) = a(x,\theta)\chi(\theta/R).$$

We will assume without loss of generality that $a(x,\theta) = 0$ for $|\theta| \leqslant 1$, since the resulting weak $*$ convergence obtained by taking a cutoff function supported away from the origin is not affected, and the quantity

$$\int a(x,\theta)\chi(\theta)e^{2\pi i\phi(x,\theta)} \, d\theta$$

623

defines a function in $C^\infty(\Omega)$. For similar reasons, we will assume without loss of generality that the functions $\nabla_x\phi$ and $\nabla_\theta\phi$ have no common zeroes on the *support* of a, rather than just the microsupport. We now apply the principle of nonstationary phase, i.e. integrating by parts. Consider the *homogenized gradient*

$$(\nabla^H\phi)(x,\theta) = (\nabla_x\phi, |\theta|\nabla_\theta\phi(x,\theta)).$$

Then $\nabla^H_{x,\theta}\phi$ is also homogeneous of degree one in the θ variable, and $|\nabla^H_{x,\theta}\phi| > 0$ on the support of a. We note that

$$(\nabla^H_{x,\theta}e^{2\pi i\phi}) = (2\pi i)e^{2\pi i\phi}\nabla^H_{x,\theta}\phi.$$

Note that the formal transpose of $\nabla^H_{x,\theta}$ is the differential operator L given for pairs of functions $F_1 : \Omega \times \mathbb{R}^p \to \mathbb{R}^d$ and $F_2 : \Omega \times \mathbb{R}^p \to \mathbb{R}^p$ by the formula

$$\begin{aligned}
L(F_1, F_2) &= -(\nabla_x \cdot F_1 + \nabla_\theta \cdot (|\theta|F_2) \\
&= -\left(\nabla_x \cdot F_1 + |\theta|^{-1}(\theta \cdot F_2) + |\theta|(\nabla_\theta \cdot F_2)\right).
\end{aligned}$$

Thus we conclude that

$$\begin{aligned}
\langle I_{a,\theta,R}, f\rangle &= \frac{1}{2\pi i}\int \frac{a\cdot f}{|\nabla^H_{x,\theta}\phi|^2}\left(\nabla^H_{x,\theta}e^{2\pi i\phi}\right)\cdot(\nabla^H_{x,\theta}\phi) \\
&= \frac{1}{2\pi i}\int L\left\{\frac{a\cdot f}{|\nabla^H_{x,\theta}\phi|^2}\nabla^H_{x,\theta}\phi\right\}e^{2\pi i\phi}.
\end{aligned}$$

Expanding out the differential operator L, we see we can write

$$\langle I_{a,\theta,R}, f\rangle = \sum_{|\alpha|\leqslant 1}\langle I_{1,a_{1,\alpha},\theta,R}, \partial^\alpha f\rangle,$$

where $a_{1,\alpha}$ is a symbol of order at most $\mu - 1$. Iterating this argument, we find that for any $N > 0$, we can write

$$\langle I_{a,\theta,R}, f\rangle = \sum_{|\alpha|\leqslant N}\langle I_{N,a_{N,\alpha},\theta,R}, \partial^\alpha f\rangle,$$

where $a_{N,\alpha}$ is a symbol of order at most $\mu - N$. If $N > \mu + p$, then we conclude the result follows for each term on the right hand side, and these quantities are each distributions of order zero. But this means that the result holds on the left hand side, and the resulting distribution $I_{a,\theta}$ has order at most N. $\qquad\square$

Remark. Suppose a is a symbol of order μ. Using the fact that if $\mu < -p$, then $I_{a,\phi}$ is continuous, and the fact that $\partial^\alpha I_{a,\phi} = I_{b,\phi}$, where b is a symbol of order at most $\mu + |\alpha|$. It thus follows that if $\mu < -p - n$, then $I_{a,\phi} \in C^n(\Omega)$. In particular, if $\mu = -\infty$, then $I_{a,\phi} \in C^\infty(\Omega)$.

Remark. If M is a manifold, and E is a vector bundle over M, then for any homogeneous phase $\phi \in C^\infty(E - \{0\})$ and any symbol $a \in \mathcal{S}^t(E^*)$, we can consider the oscillatory integral distribution formally defined by the integral

$$I_{a,\phi}(x) = \int_{E_x} a(x,\theta) e^{2\pi i \phi(x,\theta)} \, d\theta,$$

which is a distribution on M. One can develop a theory of such distributions, but we mainly focus on the simpler case defined here.

The wavefront set of $I_{a,\phi}$ has a simple formula. If ψ is a bump function supported in a neighbourhood of some point x_0, then rescaling gives

$$\widehat{I_{a,\phi}\psi}(\lambda\xi_0) = \lambda^d \int \int e^{2\pi i \lambda(\phi(x,\theta) - x \cdot \xi_0)} a(x,\lambda\theta)\psi(x) \, dx \, d\theta.$$

This is an oscillatory integral, and the phase is non-stationary in the x and θ variables provided that either:

- $\nabla_\theta\phi(x_0,\theta_0) \neq 0$, and the support of ψ is appropriately small.

- $\nabla_x\phi(x_0,\theta_0) \neq \xi_0$, and the support of ψ is appropriately small.

Thus we are lead to conclude that if we define the set

$$\Lambda_\phi = \{(x_0, \nabla_x\phi(x_0,\theta_0)) : \nabla_\theta\phi(x_0,\theta_0) = 0\},$$

then $\mathrm{WF}(I_{a,\phi}) \subset \mu\text{-supp}(a) \cap \Lambda_\phi$.

Theorem 39.2. *If $a \in \mathcal{S}^t(U \times \mathbb{R}^p)$, and $\phi \in C^\infty(U \times \mathbb{R}^p)$ is a phase satisfying the standard conditions to define the oscillatory integral distribution $I_{a,\phi}$, then*

$$WF(I_{a,\phi}) \subset \Lambda_\phi.$$

Proof. Fix $(x_0,\xi_0) \in T^*U$, and suppose that for all $(x_0,\theta_0) \in U \times \mathbb{R}^p$, either $\nabla_\theta\phi(x_0,\theta_0) \neq 0$, $\nabla_x\phi(x_0,\theta_0) \neq \xi_0$, or $(x_0,\theta_0) \notin \mu\text{-supp}(a)$. Without loss of generality, we can assume that this third condition never holds, by decomposing a as the sum of a symbol in $\mathcal{S}^{-\infty}$ and a symbol where a vanishes in a neighborhood of

625

any point (x_0, θ_0) with $\nabla_\theta \phi(x_0, \theta_0) = 0$ and $\nabla_x \phi(x_0, \theta_0) = \xi_0$, and using the fact that if $a \in \mathcal{S}^{-\infty}$, then $I_{a,\phi} \in C^\infty(U)$. Write

$$\widehat{I_{a,\phi}\psi}(\lambda\xi_0) = \lambda^d \int \int e^{2\pi i \lambda \tilde{\phi}(x,\theta;\xi_0)} \tilde{a}_\lambda(x,\theta) \, dx \, d\theta$$

where $\tilde{a}_\lambda(x,\theta) = a(x, \lambda\theta)\psi(x)$, and write $\tilde{\phi}(x, \theta; \xi) = \phi(x, \theta) - x \cdot \xi$. If Z is the set of all $\theta \in \mathbb{R}^p$ with $|\theta| = 1$ such that $\nabla_\theta \phi(x_0, \theta) = 0$, then Z is closed, and thus compact. Since $\nabla_x \phi(x_0, \theta) \neq \xi_0$ for all $\theta \in Z$, it follows by compactness that

$$|\nabla_x \phi(x_0, \theta) - t\xi_0| \gtrsim 1$$

for all $\theta \in Z$ and all $t > 0$. By homogeneity, for any $\theta \in \mathbb{R}^p - \{0\}$ such that $\nabla_\theta \phi(x_0, \theta) = 0$,

$$|\nabla_x \phi(x_0, \theta) - \xi_0| \gtrsim |\theta|.$$

By reducing the implicit constant slightly, we may assume that there is $\varepsilon > 0$ such that for any $x \in \mathbb{R}^d$ and $\xi \in \mathbb{R}^d$ with $|x - x_0| \leqslant \varepsilon$ and $|\xi - \xi_0| \leqslant \varepsilon$, and any $\theta \in \mathbb{R}^p - \{0\}$ such that $|\nabla_\theta \phi(x_0, \theta)| \leqslant \varepsilon$, then

$$|\nabla_x \phi(x, \theta) - \xi| \gtrsim |\theta|.$$

Now if ψ has support in a ε neighborhood of x_0, it follows that for all $x \in \mathrm{supp}(\psi)$ and any $\theta \in \mathbb{R}^p$,

$$|\nabla_{x,\theta}\tilde{\phi}(x, \theta, \xi)| \gtrsim 1.$$

The principle of nonstationary phase thus guarantees that for any $N > 0$,

$$|\widehat{I_{a,\phi}\psi}(\lambda\xi_0)| \lesssim_{\phi,N} \lambda^{-N} \sup_{|\alpha| \leqslant N} \|D^\alpha_{x,\theta}\tilde{a}_\lambda(x,\theta)|\nabla_{x,\theta}\tilde{\phi}(x,\theta;\xi)|^{|\alpha|-2N}\|_{L^\infty(U \times \mathbb{R}^p)}.$$

Now

$$|\nabla_{x,\theta}\tilde{\phi}(x, \theta; \xi)| = |\nabla_x \phi(x, \theta) - \xi| \lesssim 1,$$

and if $D^\alpha_{x,\theta} = D^{\alpha_1}_x D^{\alpha_2}_\theta$, then $D^\alpha_{x,\theta}\tilde{a}_\lambda(x,\theta)$ is a finite sum of $O(\alpha)$ terms, each of the form

$$\lambda^{|\alpha_2|} D^{\beta_1}_x D^{\alpha_2}_\theta a(x, \lambda\theta) D^{\beta_2}_x \psi(x),$$

and

$$|\lambda^{|\alpha_2|} D^{\beta_1}_x D^{\alpha_2}_\theta a(x, \lambda\theta) D^{\beta_2}_x \psi(x)| \lesssim_{\beta_1,\beta_2} \lambda^{|\alpha_2|}\langle\lambda\rangle^{t-|\alpha_2|} \lesssim \langle\lambda\rangle^t$$

Thus we conclude that

$$|\widehat{I_{a,\phi}\psi}(\lambda\xi)| \lesssim_{\phi,\psi,N} \langle\lambda\rangle^{t-N}.$$

Since N was arbitrary, we thus find that $\xi_0 \notin \mathrm{WF}(I_{a,\phi})$. $\qquad\square$

Example. *If we set $\phi(x, \xi) = -x \cdot \xi$, and $a(x, \xi) = a(\xi)$ is a symbol depending only on the ξ variable, then*

$$I_{a,\phi}(x) = \int a(\xi) e^{-2\pi i x \cdot \xi} \, d\xi$$

is the Fourier transform of a. We calculate that

$$\Lambda_\phi = \{0\} \times \mathbb{R}^d_\xi.$$

Thus we conclude that

$$WF(I_{a,\phi}) \subset \{0\} \times \mathbb{R}^n.$$

Thus the Fourier transform of any symbol is smooth away from the origin. This result should be compared to the standard result that the Fourier transform of a homogeneous distribution which is C^∞ away from the origin is also homogeneous, and C^∞ away from the origin.

As in the theory of pseudodifferential operators, it is often natural to work with oscillatory integral distributions modulo elements of $C^\infty(\Omega)$. Modulo C^∞, the result above implies that we can assume without loss of generality that any symbol a is supported on an arbitrarily small neighborhood of it's microsupport, and an arbitrarily small neighborhood of Λ_ϕ.

Near the wavefront set of an oscillatory integral distribution, we can compute an asymptotic formula which characterizes the behaviour of the distribution near the wavefront set, via integrating against a test function.

Theorem 39.3. *Consider a phase function $\phi : U \times \mathbb{R}^p \to \mathbb{R}$. Fix $(x_0, \theta_0) \in U \times \mathbb{R}^p - \{0\}$, such that*

$$\nabla_\theta \phi(x_0, \theta_0) = 0.$$

Let $\xi_0 = \nabla_x \phi(x_0, \theta_0)$. Consider any phase function $\psi : U \times \mathbb{R}^q \to \mathbb{R}$, and $\sigma_0 \in \mathbb{R}^q$ such that

$$\nabla_x \phi(x_0, \theta_0) = \nabla_x \psi(x_0, \sigma_0).$$

Furthermore, assume that the Hessian $H_{x,\theta}(\phi - \psi)$ of the function

$$\phi - \psi : U \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$$

is non-degenerate at $(x_0, \theta_0, \sigma_0)$, i.e. if the symmetric $(p + n) \times (p + n)$ matrix

$$\begin{pmatrix} H_x \phi & D_x(\nabla_\theta \phi) \\ D_\theta(\nabla_x \phi) & H_\theta \phi \end{pmatrix} - \begin{pmatrix} H_x \psi & 0 \\ 0 & 0 \end{pmatrix}$$

627

is invertible. Then there exists a conical neighborhood Γ of (x_0, θ_0), an open neighborhood V of x_0, and an open neighborhood W of σ_0, such that if $f \in \mathcal{D}(V)$, and a is a symbol on $U \times \mathbb{R}^p$ with $\mu\text{-supp}(a) \subset \Gamma$, then there exists a sequence of smooth, bounded functions c_k such that as $\lambda \to \infty$, for $\sigma \in W$,

$$\int I_{a,\phi_1}(x) f(x) e^{-2\pi i \lambda \psi(x,\sigma)} \, dx$$

$$\sim e^{-2\pi i \lambda \psi(x_\sigma, \sigma)} \frac{e^{(i\pi/4) sgn\{Q(\sigma)\}}}{|\det Q(\lambda)|^{1/2}} \lambda^{\frac{p-d}{2}} \sum_{k=0}^{\infty} c_k(\sigma, \lambda) \cdot \lambda^{-k},$$

Here $x_\sigma = x(\sigma)$ and $\theta_\sigma = \theta(\sigma)$ are defined locally in a neighborhood of σ_0, such that

$$\nabla_x \phi(x_\sigma, \theta_\sigma) = \nabla_x \phi(x_\sigma, \sigma) \quad \text{and} \quad \nabla_\theta \phi(x_\sigma, \theta_\sigma) = 0,$$

and such that $\theta_{\sigma_0} = \theta_0$, and $x_{\sigma_0} = x_0$. These functions exist precisely because the assumptions of the implicit function theorem are given by the Hessian conditions. The functions $\{c_k\}$ are defined as follows, though explicitly computing them is often more trouble than what it's worth. Consider the matrix

$$Q(\sigma) = H_{x,\theta}(\phi - \psi)(x_\sigma, \theta_\sigma, \sigma),$$

which we can expand as

$$\begin{pmatrix} H_x \phi(x_\sigma, \theta_\sigma) - H_x \psi(x_\sigma, \sigma) & D_x(\nabla_\theta \phi)(x_\sigma, \theta_\sigma) \\ D_\theta(\nabla_x \phi)(x_\sigma, \theta_\sigma) & H_\theta \phi(x_\sigma, \theta) \end{pmatrix}.$$

In a small neighborhood of σ_0, our assumption guarantees that $Q(\sigma)$ is an invertible matrix. For each σ, we now choose a diffeomorphism $(x, \theta) \mapsto y(x, \theta, \sigma)$, such that

$$y(x_\sigma, \theta_\sigma, \sigma) = 0, \qquad (D_{x,\theta} y)(x_\sigma, \theta_\sigma, \sigma) = I,$$

and,

$$(\phi - \psi)(x, \theta, \sigma) = (\phi - \psi)(x_\sigma, \theta_\sigma, \sigma) + y^T Q(\sigma) y / 2.$$

The function y exists, and is unique in a neighborhood of $(x_0, \theta_0, \sigma_0)$, by a result which is essentially the Morse lemma (see Lemma 1.2.2 of Duistermaat). If we now define

$$R(\sigma) = \frac{i}{4\pi} \nabla_y^T Q(\sigma)^{-1} \nabla_y,$$

and consider a function g defined for $y = y(x, \theta, \sigma)$ by the relation

$$g(y, \sigma, \lambda) = \frac{a(x, \lambda\theta) f(x)}{|\det(D_{x,\theta} y(x, \theta, \sigma))|}.$$

628

We then have $c_k = R(\sigma)^k g(0, \sigma, \lambda)/k!$.

As special cases, we have

$$c_0(\sigma, \lambda) = a(x_\sigma, \lambda\theta_\sigma) f(x_\sigma)$$

A derivative in y corresponds to a linear combination of derivatives in the x and θ variables. This shows each function $\{c_k\}$ is smooth in the σ and λ variables. Since $\theta_\sigma \neq 0$, we have

$$\nabla_\theta^N \{a(x, \lambda\theta)\}\big|_{\theta=\theta_\sigma} \lesssim_N 1,$$

uniformly in the conic set of λ under consideration. Thus we have $|c_N| \lesssim_N 1$, and so the asymptotic formula under study does actually decay in λ as $\lambda \to \infty$.

The phase ϕ of an oscillatory integral distribution is called *nondegenerate* if whenever $\nabla_\theta\phi(x, \theta) = 0$, the matrix $D_{x,\theta}(\nabla_\theta\phi)(x, \theta)$ has full rank p. It follows that

$$\Sigma_\phi = \{(x, \theta) : \nabla_\theta\phi(x, \theta) = 0\}$$

is a d dimensional submanifold of $U \times \mathbb{R}^p$. Moreover, the map f from Σ_ϕ to $U \times \mathbb{R}^d$ given by $(x, \theta) \mapsto (x, \nabla_x\phi(x, \theta))$ is an immersion, the immersed submanifold in the image being denoted by $\Lambda_\phi \subset T^*U$. To verify the map is an immersion, we note that at a point $(x, \theta) \in U \times \mathbb{R}^p$, the tangent space of Σ_ϕ consists of vectors $(v, w) \in \mathbb{R}^d \times \mathbb{R}^p$ such that

$$D_x\nabla_\theta\phi(x, \theta) \cdot v + D_\theta\nabla_\theta\phi(x, \theta) \cdot w = 0.$$

Now
$$D_{x,\theta}f(x, \theta)(v, w) = (v, D_x\nabla_x\phi(x, \theta) \cdot v + D_\theta\nabla_x\phi(x, \theta) \cdot w).$$

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 set 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 $(x_0, \theta_0, \sigma_0)$ if and only if ϕ is nondegenerate phase function in a neighborhood of (x_0, θ_0), and the covector field $d_x\psi : U \times \mathbb{R}^q \to T^*U$ 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' in the class of all phase functions.

The assumptions of the asymptotic theorem above can be specified geometrically in terms of Λ_ϕ. More precisely, the assumptions of the theorem above hold at a point $(x_0, \theta_0, \sigma_0)$ if and only if ϕ is a non-degenerate phase function, and *also* that the section $d_x \psi : U \times \mathbb{R}^q \to T^* U$ intersects Λ_ϕ transversally at the point $(x_0, \theta_0, \sigma_0)$. In particular, given that ϕ is non-degenerate, and stationary at (x_0, θ_0), a generic ψ satisfying $\nabla_x \phi(x_0, \theta_0) = \nabla_x \psi(x_0, \sigma_0)$ will satisfy the assumptions of the theorem.

The manifold Λ_ϕ of $U \times \mathbb{R}^d$ actually has special geometric structure. Consider the two form

$$\sigma = d\xi^1 \wedge dx^1 + \cdots + d\xi^d \wedge dx^d.$$

The $\sigma = d\omega$, where $\omega = \xi^1 dx^1 + \cdots + \xi^d dx^d$. 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 $T^* \mathbb{R}^d$ with respect to the two form σ.

For any phase function ϕ (possibly degenerate), we can define

$$\Sigma_\phi = \{(x, \theta) : \nabla_\theta \phi(x, \theta) = 0\}$$

$$\Lambda_\phi = \{(x, \nabla_x \phi(x, \theta)) : \nabla_\theta \phi(x, \theta) = 0\}.$$

If Λ_ϕ is an immersed Lagrangian manifold (though not necessarily an immersion through the map $f : \Sigma_\phi \to \Lambda_\phi$), we say $I_{a,\phi}$ is a *Lagrangian distribution*. The phase ϕ might be degenerate in this case.

Example. *A degenerate example of a Lagrangian distribution is given for $p = d + 1$, $\theta = (\theta_0, \theta_1)$ with $\theta_0 \in \mathbb{R}^d$ and $\theta_1 \in \mathbb{R}$, and*

$$\phi(x, \theta) = x \cdot \theta_0,$$

then ϕ defines a Lagrangian distribution $I_{a,\phi}$ for any symbol a, provided that $a(x, \theta)$ vanishes for any θ in a cone containing the θ_1 axis. Now $\Sigma_\phi = \{0\} \times \mathbb{R}^{d+1}$, and $\Lambda_\phi = \{0\} \times \mathbb{R}^d$, which is a Lagrangian manifold. Thus the distributions

$$I_{a,\phi}(x) = \int a(x, \theta_0, \theta_1) e^{2\pi i x \cdot \theta_0} \, d\theta_0 \, d\theta_1$$

630

are Lagrangian.

If $f : U \to V$ is a diffeomorphism between open subsets of $\mathbb{R}^d$, 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 Lagrangian submanifolds of T^*M for any manifold M, and Lagrangian distributions on manifolds.

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.

Chapter 40

Pseudodifferential Operators

The goal of this chapter is to define the calculus of pseudodifferential operators, a general family of operators which allows us to manipulate the spatial and frequential properties of functions simultaneously. It is impossible to do this *locally*, because of the uncertainty principle, which prevents us from locally isolating the spatial and frequency support to an arbitrary precision, 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. Roughly speaking, we will define a family of operators $a(x, D)$, associated with functions $a(x, \xi)$, such that if the support of a function f is concentrated near a point x_0, and the support of $\hat{f}$ is concentrated near ξ_0, then $a(x, D)f \approx a(x_0, \xi_0)f$. 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 most basic spatial multipliers in analysis are the *position operators*, which are the family of operators $X^\alpha : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$, defined by setting

$$X^\alpha f(x) = x^\alpha f(x).$$

The *momentum operators* $D^\alpha : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ provide the most basic frequency multipliers, given by the relationship

$$\widehat{D^\alpha f}(\xi) = \xi^\alpha \hat{f}(\xi).$$

Note that in this chapter, the operators $\{D^\alpha\}$ will be normalized as such, and thus differ from the usual differential operators, which we will here denote by ∂^α, by the constant $(2\pi i)^{-|\alpha|}$. For each $m \in C^\infty(\mathbb{R}^d)$ such that m and all of it's derivatives

are slowly increasing, we can define a bounded operator $m(X) : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ by setting

$$m(X)f(x) = m(x)f(x).$$

We can also define an operator $m(D) : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^d)$ by setting

$$\widehat{m(D)f}(\xi) = m(\xi)\hat{f}(\xi).$$

Thus we have found two homomorphisms from a ring of smooth functions on $\mathbb{R}^d$ to the ring of bounded operators on $\mathcal{S}(\mathbb{R}^d)$.

The family of such operators is very useful in analysis, since families of functions are more amenable to intuition than families of operators, and so we can try and understand what features of the function m tell us about the resulting operators $m(X)$ and $m(D)$. For instance, an analysis of operators of the form $m(D)$ is very important to the study of *elliptic* differential operators with constant coefficients. Recall that a partial differential operator $L = \sum c_\alpha D^\alpha$ of degree k is *elliptic* if the homogeneous polynomial $\sum_{|\alpha|=k} c_\alpha \xi^\alpha$ is non-vanishing away from the origin. Let us consider an example.

Theorem 40.1. *If L is an elliptic partial differential operator on $\mathbb{R}^d$, with constant coefficients, then L has a fundamental solution, i.e. there exists a distribution $\Phi \in \mathcal{D}(\mathbb{R}^d)^*$ such that $L(\Phi) = \delta$.*

Proof. Suppose L has degree k, and write $L = \sum c_\alpha D^\alpha$ for some constants $\{c_\alpha\}$. Since L is elliptic, there exists $R > 0$ such that the polynomial $P(\xi) = \sum_\alpha c_\alpha \xi^\alpha$ satisfies $|P(\xi)| \sim |\xi|^k$ for $|\xi| \geqslant R$. If $\chi \in \mathcal{D}(\mathbb{R}^d)$ is chosen such that $\chi(\xi) = 1$ for $|\xi| \leqslant R$, and we define a distribution Φ_0 such that

$$\widehat{\Phi_0}(\xi) = \frac{(1 - \chi(\xi))}{P(\xi)}.$$

Then $\widehat{\Phi_0}$ is a smooth function with $|\widehat{\Phi_0}(\xi)| \lesssim \langle \xi \rangle^{-k}$ for all $\xi \in \mathbb{R}^d$, which means that $\widehat{\Phi_0}$ is a well defined tempered distribution, and thus Φ_0 is also a well defined tempered distribution. But then

$$\widehat{L\Phi_0} = 1 - \chi(\xi).$$

Since $\chi \in \mathcal{D}(\mathbb{R}^d)$, it follows by the Paley-Wiener theorem, taking the inverse Fourier transform that $L\Phi_0 = \delta - w$, where w is an entire analytic function of at most polynomial increase. The Cauchy-Kovalevskaya theorem (i.e. solving

633

the equation by expanding out power series) allows us to find an entire analytic function u of at most exponential increase such that $Lu = w$. Then $\Phi = \Phi_0 + u$ is a fundamental solution for L. $\qquad\square$

The theory of pseudodifferential operators was introduced primarily to generalize these kinds of constructions to elliptic linear partial differential equations with *non constant* coefficients. A *(left) parametrix* for a linear differential operator L with smooth coefficients on a domain Ω is an operator $S : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)^*$ such that $1 - S \circ L$ is a *smoothing operator*. We think of S as given an 'approximate inverse' for the operator T. The existence of a regular parametrix for an elliptic linear differential operator, which will be justified by the theory of pseudodifferential operators, is quite important in the theory of differential equations. In particular, it proves that certain differential operators are *hypoelliptic*, i.e. that if $u \in \mathcal{D}(\Omega)$, then sing-supp$(Lu) = $ sing-supp(u); it suffices to show sing-supp$(u) \subset$ sing-supp(Lu), since the inclusion sing-supp$(Lu) \subset$ sing-supp(u) is true for any differential operator L with smooth coefficients. We recall that a Schwartz operator S is called *very regular* if it is regular in each variable (a smooth, distribution valued function in each variable) and agrees with integration against a smooth function away from the diagonal.

Theorem 40.2. *Let L be a differential operator with smooth coefficients. If L has a very regular left parametrix S, then L is hypoelliptic.*

Proof. For any very regular operator S, sing-supp$(Su) \subset$ sing-supp(u). It suffices to prove that for any *compactly supported* distribution u, we have sing-supp$(u) \subset$ sing-supp(Lu), since the general case follows by localization. Since $1 - S \circ L$ is a smoothing operator, we have

$$
\begin{aligned}
\text{sing-supp}(u) &\subset \text{sing-supp}((1 - S \circ L)u) \cup \text{sing-supp}((S \circ L)u) \\
&\subset \text{sing-supp}((S \circ L)u) \\
&\subset \text{sing-supp}(Lu). \qquad\qquad\qquad\square
\end{aligned}
$$

The family of pseudodifferential operators we will study are *microlocal*, i.e. not only do they not expand the singular support of distributions, but they also do not expand the wavefront set of distributions. It will therefore follow from the theory that any elliptic differential operator has a pseudodifferential parametrix, and an analogous argument to that given above gives the stronger equation

$$
\text{WF}(Lu) = \text{WF}(u)
$$

634

for any distribution u.

Returning to the general question of constructing a functional calculus which includes both the position and momentum operators, we recall the *spectral calculus*, whose goal, for a suitable algebra of normal operators A, is to produce an isomorphism of A with an algebra of functions on some space X, called the *spectrum* of A. A natural hope would be to find such a calculus for an algebra A of operators which includes the position and momentum operators. This would, in particular, enable us to analyze linear differential operators with non-constant coefficients. However, we quickly see that such a theory would not quite work in as standard a way as the spectral calculus provides, because the families of operators $\{X^\alpha\}$ and $\{D^\alpha\}$ do *not* commute with one another, i.e. the chain rule implies that

$$[D^i, X^i] = D^i X^i - X^i D^i = 1.$$

The key thing we should notice from this equation, however, is that this equation indicates that position and momentum operators commute 'up to lower order terms'. In other words, if we think of X^α and D^α as being operators of *order* $|\alpha|$, then $[D^\alpha, X^\beta]$ is equal to zero, *modulo terms of order* $|\alpha| + |\beta| - 1$. This fact will enable us to obtain an 'approximate' functional calculus for the desired algebra of operators. This is precisely the *calculus of pseudodifferential operators*.

We will associate, with each suitably regular function $a(x, \xi)$, an operator $a(x, D)$. The association will then be a homomorphism 'modulo lower order terms'. This association will have the property that if $a(x, \xi) = \sum c_\alpha(x) \xi^\alpha$, then $a(x, D)$ will be the differential operator $\sum c_\alpha(x) D^\alpha$. Indeed, this is where the notation $a(x, D)$ comes from. The association will also generalize the two families of multiplier operators; if $a(x, \xi) = m(x)$, then $a(x, D)$ is equal to $m(X)$, and if $a(x, \xi) = m(\xi)$, then $a(x, D)$ is equal to $m(D)$. 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

$$\begin{aligned} a(x, D)f &= \sum c_\alpha(x) D^\alpha f(x) \\ &= \int_{\mathbb{R}^d} \sum_\alpha c_\alpha(x) \xi^\alpha \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi \\ &= \int_{\mathbb{R}^d} a(x, \xi) e^{2\pi i \xi \cdot x} \widehat{f}(\xi) \, d\xi. \end{aligned}$$

We use this integral formula to define $a(x, D)$ for a much more general family of functions $a(x, \xi)$.

Fix an open set $\Omega \subset \mathbb{R}^d$. Given any distribution $a(x, \xi)$ which is tempered in the ξ variable, i.e. any continuous, bilinear map $a : \mathcal{D}(\Omega_x) \times \mathcal{S}(\mathbb{R}^d_\xi) \to \mathbb{C}$, or more technically, any element of the tensor product $\mathcal{D}(\Omega_x)^* \otimes^\wedge \mathcal{S}(\mathbb{R}^d_\xi)^*$, we can associate an operator $a(x, D) : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)^*$, such that for any $f, g \in \mathcal{D}(\Omega)$,

$$\langle a(x, D) f, g \rangle = \int a(x, \xi) \widehat{f}(\xi) e^{2\pi i \xi \cdot x} g(x) \, dx \, d\xi.$$

We call any operator T which can be given in the form $a(x, D)$ a *pseudodifferential operator*. The symbol a is uniquely determined by the operator T, since the action of $a(x, D)$ on $\mathcal{D}(\Omega)$ determines the behaviour of a, viewed as a bilinear map, on an arbitrary element of $\mathcal{D}(\Omega_x) \times \mathcal{S}(\mathbb{R}^d_\xi)$. For any set $S \subset \mathcal{D}(\Omega_x)^* \otimes^\wedge \mathcal{S}(\mathbb{R}^d_\xi)^*$, the notation $\text{Op}(S)$ is often used to denote the family of all pseudodifferential operators defined by an element of S.

Example. *Consider the Laplacian Δ on $\mathbb{R}^n$. Then Δ is the pseudodifferential operator on $\mathbb{R}^n$ of order two, corresponding to the symbol $a(x, \xi) = -4\pi^2 |\xi|^2$. Since Δ is a constant coefficient operator, it just acts as a Fourier multiplier. If $\Delta u = f$, then, modulo a harmonic function, which is arbitrarily smooth, $u = \Phi * f$, where Φ is a fundamental solution to the Laplacian. The operator $f \mapsto \Phi * f$ is a pseudodifferential operator with symbol $b(x, \xi) = f.p(1/4\pi^2 |\xi|^2)$.*

Example. *The Cauchy-Riemann operator on $\mathbb{R}^2$ given by*

$$\frac{\partial}{\partial \overline{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right)$$

is a pseudodifferential operator of order one with symbol $i\pi(\xi + i\eta)$.

Any pseudodifferential operator $T = a(x, D)$ is continuous from $\mathcal{D}(\Omega)$ to $\mathcal{D}(\Omega)^*$, and has Schwartz kernel

$$K_a(x, y) = \int a(x, \xi) e^{2\pi i \xi \cdot (x-y)} \, d\xi,$$

where in general the oscillatory integral must be interpreted formally, i.e. as an oscillatory integral distribution. It is also useful to write this kernel in the convolution form $k_a(x, z) = K_a(x, x - z)$, because we then have

$$T\phi(x) = \int k_a(x, z) f(x - z) \, dz,$$

which reflects the fact that when $a(x, \xi)$ is independant of x, k_a is a function of z, and then $Tf = k_a * f$ is a convolution operator. In fact, all pseudodifferential operators are infinite sums of convolution operators, in the following sense: if $a \in \mathcal{D}(\mathbb{R}^d)^* \otimes^\wedge \mathcal{S}(\mathbb{R}^d)^*$, then we can write a as a sum of the form

$$\sum_{i=1}^{\infty} u_i \otimes v_i,$$

where $\{u_i\}$ are in $\mathcal{D}(\mathbb{R}^d)^*$, and $\{v_i\} \in \mathcal{S}(\mathbb{R}^d)^*$, and the convergence occurs unconditionally, in the distributional topology. It then follows that for $\phi \in \mathcal{D}(\Omega)$,

$$a(x, D)\phi = \sum_{i=1}^{\infty} u_i \cdot (v_i * \phi).$$

Thus $a(x, D)$ can be interpreted as a non-constant coefficient sum of convolution operators.

As we increase the regularity of a, we no longer need to treat the definition of a pseudodifferential operator quite as formally, and so we can define the operator on a more general family of functions. Here are some non-comprehensive examples of this phenomenon:

- If $a \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)^*$, then $a(x, D)$ extends to a continuous linear operator from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}(\mathbb{R}^d)^*$. The Schwartz kernel theorem implies that any continuous linear operator from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}(\mathbb{R}^d)^*$ is of this form, which probably indicates that the family of such operators is too general to obtain interesting results.

- If $a \in \mathcal{S}_{\text{loc}}^t(\Omega \times \mathbb{R}^d)$, then we will see later on that $a(x, D)$ extends to a continuous operator from $\mathcal{E}(\Omega)^*$ to $\mathcal{D}(\Omega)^*$ and from $\mathcal{D}(\Omega)$ to $\mathcal{E}(\Omega)$.

- If $a \in \mathcal{S}^t(\mathbb{R}^d \times \mathbb{R}^d)$, then $a(x, D)$ extends to a continuous operator from $\mathcal{S}(\mathbb{R}^d)$ to itself.

- If $a \in \mathcal{S}_{\text{loc}}^{-\infty}(\Omega \times \mathbb{R}^d)$, then we will see later in this chapter that $a(x, D)$ has a kernel in $C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, and is therefore a smoothing operator, thus extending to a continuous operator from $\mathcal{E}(\mathbb{R}^d)^*$ to $\mathcal{E}(\mathbb{R}^d)$.

  Conversely, we will also see that if T is *any* 'pseudolocal' smoothing operator, in the sense that it has a kernel $K \in C^\infty(\Omega \times \Omega)$ satisfying bounds of the form

  $$|\nabla_x^n \nabla_y^m K(x, y)| \lesssim_N \langle x - y \rangle^{-N},$$

then $T \in \text{Op}(\mathcal{S}_{\text{loc}}^{-\infty}(\Omega \times \mathbb{R}^d))$. In particular, any proper smoothing operator is of this form.

- If $a(x, D)$ is a proper operator, then it maps $\mathcal{D}(\Omega)$ into $\mathcal{E}(\Omega)^*$ and from $\mathcal{E}(\Omega)$ into $\mathcal{D}(\Omega)^*$. In combination with the previous results, we conclude that if $a \in \mathcal{S}_{\text{loc}}^t(\Omega \times \mathbb{R}^d)$, then $a(x, D)$ is an operator from $\mathcal{D}(\Omega)^*$ to itself, $\mathcal{E}(\Omega)^*$ into itself, and from $\mathcal{D}(\Omega)$ into itself. If $a \in \mathcal{S}_{\text{loc}}^{-\infty}(\Omega \times \mathbb{R}^d)$, then $a(x, D)$ maps $\mathcal{D}(\Omega)$ to itself and maps $\mathcal{D}(\Omega)^*$ into $\mathcal{E}(\Omega)$.

  Conversely, let $T : \mathcal{D}(\Omega) \to \mathcal{E}(\Omega)^*$ be *any* proper operator, and let K be it's kernel. Then K lies in $\mathcal{D}(\Omega_x)^* \otimes^\wedge \mathcal{E}(\Omega_y)^*$. Thus we can define a symbol $a \in \mathcal{D}(\Omega_x)^* \otimes^\wedge \mathcal{S}(\mathbb{R}_\xi^d)^*$ by setting

  $$a(x, \xi) = \int K(x, y) e^{2\pi i \xi \cdot (x - y)} \, dy.$$

  We verify using the Fourier multiplication formula that

  $$T_a \phi(x) = \int a(x, \xi) \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi = \int K(x, y) f(y) \, dy = T\phi(x).$$

  Thus *any* proper operator is a pseudodifferential operator.

That every proper operator, and that every operator on Schwartz space, is a pseudodifferential operator, indicates that the theory of pseudodifferential operators is too general to study more detailed in the form above. We will mostly focus on pseudodifferential operators defined by various symbol classes, since most practical operator occuring in PDE and analysis are of this form, and we can build up a sophisticated calculus from tis family.

40.1   Basic Definitions

There are two varieties of the theory of pseudodifferential operators, whose basic results are roughly analogous to one another. The first, which works best when considering pseudodifferential operators on $\mathbb{R}^d$, works with operators specified by symbols $a : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ in $\mathcal{S}^t(\mathbb{R}^d \times \mathbb{R}^d)$, i.e. satisfying estimates of the form

$$|\nabla_x^n \nabla_\xi^m a(x, \xi)| \lesssim_{n,m} \langle \xi \rangle^{t - m}$$

where the bound holds uniformly in both x and ξ, for any integers n and m. As mentioned above, $a(x, D)$ then extends to a continuous operator from $\mathcal{S}(\mathbb{R}^d)$ to

itself, which leads to an elegant theory. However, this theory is less easy to work with locally, e.g. working in various different coordinate systems, or obtaining a definition of pseudodifferential operator that applies to operators on manifolds. The approach here is best taken by describing a theory described by symbols $a \in \mathcal{S}_{\mathrm{loc}}^t(\Omega \times \mathbb{R}^d)$, i.e. satisfying an inequality of the form above uniformly in ξ, but only *locally uniformly* in x.

Fix an open set $\Omega \subset \mathbb{R}^d$, and consider a symbol $a \in \mathcal{S}_{\mathrm{loc}}^t(\Omega \times \mathbb{R}^d)$. From this symbol, we can define a continuous operator $T_a : \mathcal{D}(\Omega) \to \mathcal{E}(\Omega)$ by setting

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

where the integral can now be interpreted in the usual Riemann / Lebesgue sense. The function $T_a f$ is smooth since $a \in C^\infty(\Omega_x, \mathcal{S}(\mathbb{R}_\xi^d)^*)$. We then say T_a is a *pseudodifferential operator of order t*. The family of all such operators is denoted $\Psi_{\mathrm{loc}}^t(\Omega)$; we reserve the notation $\Psi^t(\Omega)$ to denote those operators defined by elements of $\mathcal{S}^t(\Omega \times \mathbb{R}^d)$ rather than simply elements of $\mathcal{S}_{\mathrm{loc}}^t(\Omega \times \mathbb{R}^d)$.

The kernel of a pseudodifferential operator given by a symbol $a(x, \xi)$ of the class above is of the form

$$K_a(x, y) = \int a(x, \xi) e^{2\pi i \xi \cdot (x-y)} \, d\xi,$$

where the integral is now an oscillatory integral distribution. In particular, we know from our discussion of oscillatory integral distributions that

$$\mathrm{WF}(K_a) \subset \{(x, x; \xi, \xi) : x, \xi \in \mathbb{R}^d\}.$$

The microlocal analysis of distributions implies the existence of a continuous extension $T_a : \mathcal{E}(\Omega) \to \mathcal{D}(\Omega)^*$, and we find that $\mathrm{WF}(T_a u) \subset \mathrm{WF}(u)$. Thus the operator T_a preseres the location of singularities of a distribution in both position and frequency. This is the first instance of the *microlocal nature* of pseudodifferential operators; these operators roughly preserve the location of the mass and frequency support of a function, but with some additional 'fuzz' that is usually neglible to the problem, but must be managed.

The symbol a is uniquely determined from the operator T_a. To actually recover the symbol from an operator, we have several methods. Formally, we can calculate that

$$a(x, \xi) = e^{-2\pi i \xi \cdot x} T_a(e^{2\pi i \xi \cdot y}).$$

The wavefront calculation above shows that the convolution kernel $k_a(x,z)$ of a pseudodifferential operator agrees with a smooth function away from the line $z = 0$. We will see very shortly that it decays rapidly away from this line, and therefore k_a is tempered in the z-variable. The above formal equation then implies the less formal equation

$$a(x,\xi) = \int k_a(x,z)e^{2\pi i \xi \cdot z} \, dy.$$

Thus the symbol a is obtained by taking the inverse Fourier transform of the convolution kernel k_a in the z variable.

Here is a quantitative estimate on the kernel of a pseudodifferential operator, which show another instance of it's pseudolocal nature, i.e. localization on the spatial side of things. In particular, the result implies that if $\Omega = \mathbb{R}^d$, then T_a extends to a continuous operator from $\mathcal{S}(\mathbb{R}^d)$ to itself. As mentioned above, $K_a \in \mathcal{D}(\Omega \times \Omega)^*$ agrees with a C^∞ function away from the diagonal Δ_Ω. Thus if $f \in \mathcal{D}(\mathbb{R}^d)$, and $x \notin \operatorname{supp}(f)$, then the multiplication formula for tempered distributions implies that

$$T_a f(x) = \int a(x,\xi)\hat{f}(\xi)e^{2\pi i \xi \cdot x} \, d\xi = \int K(x,y)f(y) \, dy,$$

where, since the integral on the right hand side vanishes in a neighborhood of x, we can actually interpret the right hand integral as a Lebesgue integral, rather than a formal integral. Moreover, we have even better estimates for the behaviour of K_a away from the origin, which reflects the pseudolocal behaviour of the operator. To discuss these estimates, we introduce the differential operators $\partial_z^i = \partial_x^i - \partial_y^i$, and the induced operators ∇_z^m, which measures the derivatives measured in the direction normal to the diagonal in $\Omega \times \Omega$.

Theorem 40.3. *Let $a \in \mathcal{S}_{loc}^t(\Omega)$. Then for any pair of integers $n, m \geqslant 0$, and any $N \geqslant 0$ such that $t + d + m + N \geqslant 0$,*

$$|\nabla_x^n \nabla_z^m K_a(x,y)| \lesssim_{n,m,N} |x - y|^{-t-d-m-N},$$

where the implicit constant is locally uniform in x. If $a \in \mathcal{S}^t(\Omega)$, then we can choose the implicit constant to be uniform in x.

Proof. If $\operatorname{supp}_\xi(a)$ is compact, then $a \in \mathcal{S}_{loc}^{-\infty}(\Omega)$, and by the compactness, we conclude that for any $N \geqslant 0$,

$$|\nabla_x^n \nabla_z^m K_a(x,y)| = |\mathcal{F}_\xi\{a\}(x, x - y)| \lesssim_N 1/|x - y|^N,$$

640

where the uniform estimate happens because we a is compactly supported in the ξ variable, uniformly in x, and smooth, locally uniformly in x. 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

$$a(x,\xi) = \sum_{i=0}^{\infty} a_i(x,y,\xi),$$

where $a_i(x,\xi) = \rho(\xi/2^i)a(x,\xi)$ is supported on $|\xi| \sim 2^n$. Let K_i be the kernel of the pseudodifferential operator $a_i(x,D)$. Then

$$K(x,z) = \sum_{i=0}^{\infty} K_i(x,z),$$

where the convergence is distributional. We claim that for any i, and any $n, m, N \geqslant 0$,

$$|\nabla_x^n \nabla_z^m K_i(x,y)| \lesssim_{n,m,N} |x-y|^{-N} 2^{i(t+d+m-N)},$$

where the implicit constant is locally uniform in x. This follows from a simple integration by parts, applied to the integral

$$K_i(x,y) = \int \rho(\xi/2^i)a(x,\xi)e^{2\pi i\xi\cdot(x-y)} \, d\xi.$$

These bounds, if N is taken large enough, imply that the sum $K = \sum K_i$ converges uniformly on any set of the form

$$\{(x,y) \in \Omega \times \Omega : x \in K, |y-x| > \varepsilon\},$$

where $K \subset \Omega$ is compact. Summing up these bounds for sufficiently large N gives the required inequality for $|x-y| \geqslant 1$. For $0 < |x-y| \leqslant 1$, we break the sum into two parts, i.e. writing

$$K(x,y) = \sum_{2^i \leqslant 1/|x-y|} K_i(x,y) + \sum_{2^i > 1/|x-y|} K_i(x,y).$$

For the first sum, we take $N = 0$, and for the second sum, we take $N > t+d+m$, which gives the required bounds. $\qquad\square$

These singularity conditions characterize those kernels K_a induced by pseudodifferential operators of some order $t < 0$. For $t \geqslant 0$, we must also assume an additional cancellation condition.

Lemma 40.4. *Let $K \in \mathcal{D}(\Omega \times \Omega)^*$ be a Schwartz kernel with*

$$sing\text{-}supp(K) \subset \{(x, x) : x \in \Omega\}.$$

Suppose that for some t, and any non-negative integers n, m, and N, the kernel K satisfies the growth condition

$$|\nabla_x^n \nabla_z^m K(x, y)| \lesssim |x - y|^{-t-d-m-N}$$

locally uniformly in x, for all n, m, and N for which $t + d + m + N \geqslant 0$. Then:

- *If $t < 0$, and if, for each $x \in \Omega$, and any $\phi \in \mathcal{D}(\Omega)$,*

$$\int K(x, y)\phi(y) \, dy = \lim_{\varepsilon \to 0} \int_{|x-y| > \varepsilon} K(x, y)\phi(y) \, dy,$$

  where the right hand side exists, and defines a distribution induced by a locally integrable function on Ω by virtue of the growth condition with m and N equal to zero, then K is the Schwartz kernel of a pseudodifferential operator given by a local symbol of order t.

- *If $t \geqslant 0$, and any $\phi \in \mathcal{D}(\Omega)$, the kernel K satisfies the* cancellation conditions

$$\left| \int D_x^\alpha K(x, y)\phi((x - y)/R) \, dy \right| \lesssim_{\phi, \alpha} R^t,$$

  where the implicit constant is independent of R, locally uniform in x, and a continuous seminorm on $\mathcal{D}(\Omega)$ for each multi-index α, then K is the kernel of some pseudodifferential operator given by a local symbol of order t.

If we replace the inequalities that locally uniformly depend on x with inequalities that are uniform in x, then the symbols we find can also be chosen to be uniform.

Proof. TODO: Brian's Book on Multiparameter Singular Integrals has a good discussion on this. □

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, which is related to it's microlocal nature. Consider a symbol $a(x, \xi)$, a smooth function $f(y)$, and a smooth phase $\phi(y)$ with $\nabla\phi(y)$ nonvanishing on $supp_x(a)$. 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 up to first order in λ, and in fact, we can obtain a complete asymptotic development as $\lambda \to \infty$. For simplicity, we assume $\mathrm{supp}_x(a)$ is compact.

Theorem 40.5. *Fix a symbol $a \in \mathcal{S}^t(\Omega \times \mathbb{R}^d)$, compactly supported in the x-variable, a smooth function $f \in \mathcal{D}(\Omega)$, and a smooth, real-valued function $\phi \in C^\infty(\Omega)$ with $\nabla\phi$ nonvanishing on $\mathrm{supp}_x(a)$. Let*

$$r_x(y) = \nabla\phi(x) \cdot (x - y) - (\phi(x) - \phi(y)).$$

Then for any $N > 0$ and $\lambda > 0$, we can write $e^{-2\pi i\lambda\phi(x)}T_a\{fe^{2\pi i\lambda\phi}\}(x)$ as

$$\sum_{|\beta|<N} \frac{1}{\beta! \cdot (2\pi i)^\beta} \cdot \partial_\xi^\beta a(x, \lambda\nabla\phi(x)) \, \partial_y^\beta\{e^{2\pi i\lambda r_x}f\}\big|_{y=x} + R_N(x, \lambda),$$

where $\lambda^{t-\lceil N/2\rceil}R_N \in L^\infty(\mathbb{R}^d \times (0, \infty))$. In particular, for $N = 3$, we find that $e^{-2\pi i\lambda\phi(x)}T_a\{fe^{2\pi i\lambda\phi}\}(x)$ is equal to

$$a(x, \lambda\nabla\phi(x))f(x)$$

$$+ \frac{1}{2\pi i}\sum_{k=1}^d (\partial_\xi^k a)(x, \lambda\nabla\phi(x)) \cdot (\partial_x^k f)(x)$$

$$- \frac{i\lambda}{4\pi}\sum_{|\beta|=2} (\partial_\xi^\beta a)(x, \lambda\nabla\phi(x))(\partial_x^\beta\phi)(x)f(x)$$

$$+ O(\lambda^{t-2}).$$

If $\phi(x) = \xi \cdot x$ for some ξ, then we find

$$T_a\{fe^{2\pi i\lambda\phi}\}(x) \sim \sum_\beta \frac{1}{\beta!(2\pi i)^\beta}\partial_\xi^\beta a(x, \lambda\xi) \cdot \partial_x^\beta f(x).$$

Proof. We write

$$e^{-2\pi i\lambda\phi(x)}T_a\{fe^{2\pi i\lambda\phi}\}(x) = \lambda^d \int a(x, \lambda\xi)f(y)e^{2\pi i\lambda[(\xi-\nabla\phi(x))\cdot(x-y)+r_x(y)]} \, dy \, d\xi.$$

This integral is oscillatory, with phase

$$\Phi_x(\xi, y) = (\xi - \nabla\phi(x)) \cdot (x - y) + r_x(y).$$

Now $\nabla_\xi \Phi_x(\xi, y) = 0$ precisely when $y = x$, and $\nabla_y \Phi_x(\xi, x) = 0$ precisely when $\xi = \nabla\phi(x)$. If we consider a smooth cutoff $\psi \in C_c^\infty(\mathbb{R}_\xi^d)$ supported on

$$(1/2) \cdot \inf_{x \in \text{supp}_x(a)} |\nabla\phi(x)| \leq |\xi| \leq 2 \cdot \sup_{x \in \text{supp}_x(a)} |\nabla\phi(x)|$$

then we can write $e^{-2\pi i \lambda \phi(x)} T_a \{ f e^{2\pi i \lambda \phi} \}(x) = I_1 + I_2$, where I_1 is obtained by substituting ψ into the integrand, and I_2 is obtained by substituting $1 - \psi$ into the integrand. Nonstationary phase tells us that

$$|I_2| \lesssim_N \lambda^{-N}$$

for all $N > 0$, where the implicit constant is independent of λ. Thus it suffices to concentrate on I_1. In the sequel, we will therefore assume without loss of generality that $a(x, \lambda\xi) = a(x, \lambda\xi)\psi(\xi)$, i.e. $a(x, \lambda\xi)$ is supported near $|\xi| \sim 1$. A change of variables $\xi \mapsto \xi + \nabla\phi(x)$ allows us to rewrite I_1 as

$$\lambda^d \int e^{2\pi i \lambda(\xi \cdot (x-y) + r_x(y))} a(x, \lambda[\nabla\phi(x) + \xi]) f(y) \, d\xi \, dy.$$

Using Taylor's formula, we write

$$a(x, \lambda\nabla\phi(x) + \lambda\xi) = \sum_{|\beta| < N} \frac{\lambda^\beta}{\beta!} \partial_\xi^\beta a(x, \lambda\nabla\phi(x)) \xi^\beta + R_{N,\lambda}(x, \xi),$$

where

$$R_{N,\lambda}(x, \xi) = \sum_{|\beta| = N} \xi^\beta \frac{N}{\beta!} \int_0^1 (1 - t)^{N-1} \lambda^N \partial_\xi^\beta a(x, \lambda\nabla\phi(x) + t\lambda\xi) \, dt.$$

We only care about this formula when $|\xi| \sim 1$, and differentiating this formula gives that, for such ξ,

$$|\partial_\xi^\alpha R_{N,\lambda}(x, \xi)| \lesssim_\alpha \lambda^{t - |\alpha|},$$

and $\partial_\xi^\alpha R_{N,\lambda}(x, 0) = 0$ for all $|\alpha| < N$. Stationary phase thus implies that

$$\left| \lambda^d \int \int e^{2\pi i \lambda((x-y) \cdot \xi + r_x(y))} R_{N,\lambda}(x, \xi) f(y) \, d\xi \, dy \right| \lesssim \lambda^{t - [N/2]}.$$

644

Finally, we note that via an integration by parts,

$$\int e^{2\pi i \lambda(\xi \cdot (x-y) + r_x(y))} \xi^\beta f(y) \, d\xi \, dy$$
$$= \lambda^{-(d+\beta)} (2\pi i)^{-\beta} \, \partial_y^\beta \{ e^{2\pi i \lambda r_x(y)} f(y) \} \big|_{y=x}$$

and substituting them into the formula completes the proof. □

Remark. One application of this result will occur when we analyze how pseudodifferential operators behave under a change of variables. If T is a pseudodifferential operator on $\Omega \subset \mathbb{R}^d$ defined by a symbol a with $\operatorname{supp}_x(a)$ compact, and $\kappa : \Psi \to \Omega$ is a diffeomorphism, then we can consider the operator $S : \mathcal{D}(\Psi) \to \mathcal{D}(\Psi)^*$ given by 'changing the coordinates of T', i.e.

$$S\phi(x) = T(\phi \circ \kappa^{-1})(\kappa(x)).$$

We can thus write, after a change of coordinates, with $\tilde{a}(x,\xi) = a(\kappa(x),\xi)$,

$$S\phi(x) = \int \tilde{a}(x,\xi) e^{2\pi i \xi \cdot (\kappa(x) - \kappa(y))} |\det(\kappa(y))| \phi(y) \, dy \, d\xi.$$

This leads to an analysis of quantities similar to that obtained in the above proof, which will lead us to conclude that S itself is a pseudodifferential operator. Note, however, that purely from the spectral anlaysis of singularities, we see immediately from the integral kernel that S is a microlocal operator, i.e. it preserves the wavefront set of distributional inputs.

Another application occurs much later on, when we discuss the composition of a pseudodifferential operator with a *Fourier integral operator*.

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, for $t < -d$, the symbol is actually *integrable* in ξ. A ΨDO of order $-\infty$ has a kernel K lying in $C^\infty(\Omega \times \Omega)$, and satisfying estimates of the form

$$|\nabla_x^n \nabla_z^m K(x,y)| \lesssim_{n,m,N} \langle x - y \rangle^{-N}$$

for any $N \geq 0$. Thus a ΨDO of order $-\infty$ is smoothing. Conversely, if $K \in C^\infty(\Omega \times \Omega)$ satisfies estimates of the form above, then it is the ΨDO corresponding to the symbol

$$a(x,\xi) = \int K(x,y) e^{2\pi i \xi \cdot (x-y)} \, dy,$$

which is a symbol of order $-\infty$. In particular, all properly supported smoothing operators are ΨDOs of order $-\infty$.

Much of the theory of pseudodifferential operators to come is most elegantly explained *modulo smoothing operators*. Working modulo smoothing operators is usually fine in harmonic analysis provided that we are trying to establish localized estimates for certain quantities, since smoothing operators, once localized, satisfy all the inequalities we might be interested in. In particular, working modulo smoothing operators often makes the theory more flexible. For instance, one interesting operator to study is the square root of the Laplacian, which is the pseudodifferential operator with symbol $a(x, \xi) = 2\pi|\xi|$. The fact that a is not smooth near the origin means that $a(x, D)$ does not quite fit the theory of pseudodifferential operators given above. Nonetheless, for any bump function $\rho \in \mathcal{D}(\mathbb{R}_\xi^d)$ equal to one in a neighborhood of the origin, the symbol $\tilde{a}(x, \xi) = 2\pi|\xi|(1 - \rho(\xi))$ lies in $\mathcal{S}^1(\mathbb{R}^d \times \mathbb{R}^d)$. Since $\text{supp}_\xi(a - \tilde{a})$ is compact, it follows that $a(x, D) - \tilde{a}(x, D)$ is a smoothing operator. Thus, modulo smoothing operators, it makes sense to identify a with a symbol of order one. Thus we define the symbol class $\dot{\mathcal{S}}^t(\Omega \times \mathbb{R}^d)$ to be the space of all distributional symbols which differ from an element of $S_{\text{loc}}^t(\Omega \times \mathbb{R}^d)$ by a distributional symbol which induces a smoothing operator. It then follows that $a \in \dot{\mathcal{S}}^1$, so we can study the theory of the square root of the Laplacian as a pseudodifferential operator, modulo a smoothing operator.

One can already see from this example, that this theory works well with pseudodifferential operators defined in terms of homogeneous functions, of which the square root of the Laplacian is a special example, since homogeneous functions often fail to be smooth at the origin. Another useful idea when working with symbols modulo smoothing operators is that we only need to specify pseudodifferential operators by symbols $a(x, \xi)$ defined for suitably large ξ, since any two extensions of these functions to symbols defined near the origin will differ by a symbol compactly supported in the ξ variable, and thus these operators will differ by a smoothing operator.

Another virtue of working modulo smoothing operators is that asymptotics can be used to specify symbols. In other words, for any family of symbols $a_k \in \mathcal{S}_{\text{loc}}^{t_k}(\Omega \times \mathbb{R}^d)$, provided that $t_k \to -\infty$ and with $t = \max(t_k)$, we can define a unique symbol $a \in \dot{\mathcal{S}}^t(\Omega \times \mathbb{R}^d)$ by an asymptotic formula of the form

$$a(x, \xi) \sim \sum_{k=0}^{\infty} a_k(x, \xi),$$

since the difference of any two symbols satisfying this asymptotic formula lies in

$\mathcal{S}_{\text{loc}}^{-\infty}(\Omega \times \mathbb{R}^d)$, and thus induces a smoothing operator.

40.2 Compound Operators and Quantization

It is natural to 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, and we can calculate an explicit asymptotic expansion for the symbol of this operator.

Lemma 40.6. *For any $a \in \mathcal{S}_{\text{loc}}^t(\Omega_x \times \Omega_y \times \mathbb{R}^d)$, the operator T_a lies in $\Psi_{\text{loc}}^t(\Omega)$, and the symbol of T_a has the asymptotic expansion*

$$\sum_\beta \frac{1}{\beta!} \frac{1}{(2\pi i)^{|\beta|}} \partial_\xi^\beta \partial_y^\beta a(x, x, \xi).$$

Proof. We begin by noting that, by very similar techniques to those above, the kernel K of an operator defined by a compound symbol of order $-\infty$ is smooth, and satisfies estimates of the form

$$|\nabla_x^n \nabla_z^m K(x, y)| \lesssim_{n,m,N} \langle x - y \rangle^{-N}$$

locally uniformly in x and y, which means that K is the kernel of an operator in $\Psi_{\text{loc}}^{-\infty}(\Omega)$. Thus we obtain the result for $t = -\infty$. In general, we perform a Taylor expansion, writing

$$a(x, y, \xi) = \sum_{|\beta| \leqslant N} \frac{1}{\beta!} \partial_y^\beta a(x, x, \xi) \cdot (y - x)^\beta + R_N(x, y, \xi),$$

where $\partial_y^\beta R_N(x, x, \xi) = 0$ for all $|\beta| \leqslant N$. We can find C^∞ functions $b_\beta(x, y, \xi)$, for $|\beta| = N + 1$, such that

$$R_N(x, y, \xi) = \sum_{|\beta| = N+1} (2\pi i)^{N+1} (y - x)^\beta b_\beta(x, y, \xi).$$

647

Now integration by parts shows that

$$\int R_N(x, y, \xi) e^{2\pi i \xi \cdot (x-y)} \, d\xi$$

$$= (-1)^{N+1} \sum_{|\alpha|=N+1} \int (D_\xi^\alpha b_\alpha)(x, y, \xi) e^{2\pi i \xi \cdot (x-y)} \, d\xi.$$

The functions b_β are symbols of order t, so $D_\xi^\alpha b_\beta$ are symbols of order $t - (N + 1)$. Thus the operators specified by the compound symbol R_N have order at most $t - (N + 1)$. On the other hand, another integration by parts again shows that

$$\int \partial_y^\beta a(x, x, \xi) \cdot (y - x)^\beta \cdot e^{2\pi i \xi \cdot (x-y)} \, d\xi$$

$$= \frac{1}{(2\pi i)^{|\beta|}} \int \partial_\xi^\beta \partial_y^\beta a(x, x, \xi) e^{2\pi i \xi \cdot (x-y)} \, d\xi.$$

Thus the operator corresponding with symbol $\partial_y^\beta a(x, x, \xi) \cdot (y - x)^\beta$ also corresponds to the symbol $1/(2\pi i)^{|\beta|} \cdot \partial_\xi^\beta \partial_y^\beta a(x, x, \xi)$. Thus if we consider a symbol $\tilde{a}(x, \xi)$ satisfying the asymptotic equation defined in the theorem, then we see that $\tilde{a}(x, D)$ differs from T_a by a compound symbol of order $-\infty$, which, as previously discussed, is an element of $\Psi_{\text{loc}}^{-\infty}(\Omega)$. $\qquad\square$

Remark. If $b(x, \xi) = a(x, x, \xi)$, then the formula above can be written more formally as $a(x, y, \xi) \sim e^{2\pi i D_x \cdot D_\xi} b$. This makes sense, since if K is the kernel of T_a, and T_a corresponds to a pseudodifferential operator, then it's symbol would correspond precisely to

$$a(x, \xi) = \int K(x, x - y) e^{-2\pi i \xi \cdot y} \, dy$$

$$= \int \int a(x, x - y, \xi - \eta) e^{-2\pi i \eta \cdot y} \, d\eta \, dy.$$

If we define $a_x(y, \xi) = a(x, y, \xi)$, and $c(x, \xi) = e^{-2\pi i \xi \cdot x}$, then $a(x, \xi) = (a_x * c)(x, \xi)$. But c is a Gaussian, and thus if (ξ', x') are the dual variables to (x, ξ), then $\hat{c}(\xi', x') = e^{2\pi i \xi' \cdot x'}$. Thus

$$a(x, \xi) = e^{2\pi i D_x \cdot D_\xi} a_x(x, \xi),$$

where $e^{2\pi i D_x \cdot D_\xi}$ is the Fourier multiplier operator with symbol $e^{2\pi i \xi' \cdot x'}$. Taking the power series expansion of the exponential gives the expansion above.

648

For any pseudodifferential operator T with kernel $K_a(x,y)$, we can consider it's formal adjoint T^* with kernel $K(x,y) = \overline{K_a(y,x)}$. We now use the calculation above to show the adjoint of a pseudodifferential operator $a(x,\xi)$ is a pseudodifferential operator; it is simple to calculate that the adjoint of any ΨDO $a(x,D)$ 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 $a^*(x,\xi) = e^{2\pi i D_x \cdot D_\xi} \overline{a}(x,\xi)$, which we can write explicitly as an asymptotic expansion as

$$a^*(x,\xi) \sim \sum_\beta \frac{1}{\beta!} \frac{1}{(2\pi i)^{|\beta|}} \overline{\partial_\xi^\beta \partial_x^\beta a(x,\xi)}.$$

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 such that $a(x,D)$ is a *symmetric* pseudodifferential operator, i.e. such that for all $f, g \in \mathcal{D}(\Omega)$,

$$\langle a(x,D)f, g \rangle = \langle f, a(x,D)g \rangle,$$

then $a \approx \mathrm{Re}(a)$, in the sense that $a - \mathrm{Re}(a)$ is a symbol of order $t-1$.

The choice of (x,ξ) variables for pseudodifferential operators is common, but certainly not standard. The association of the pseudodifferential operator $a(x,D)$ with any symbol $a(x,\xi)$ is called the *Kohn-Nirenberg quantization*. We could also use the *adjoint Kohn-Nirenberg 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,

$$a(y,\xi) \sim \sum_\beta \frac{1}{\beta!} \frac{1}{(2\pi i)^{|\beta|}} \partial_\xi^\beta \partial_x^\beta a(x,\xi).$$

In particular, the symbol $(x,y,\xi) \mapsto a(y,\xi) - a(x,\xi)$ is a pseudodifferential operator of order $t-1$. Thus the two quantizations describe exactly the same family of operators, and the association of an operator with a symbol only matters up to lower order terms.

The family of operators one can describe via the adjoint Kohn-Nireberg quantization is the same as the Kohn-Niernberg quantization. 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.

We thus see that the differences in quantization arise because differential operators and multiplier operators *do not commute*. The operators above 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 $a(x,\xi)$ gives the Pseudodifferential operator T with compound symbol $(x,y,\xi) \mapsto a((x+y)/2,\xi)$. This approach has the advantage that T will be self-adjoint if and only if a is real-valued, rather than just this only being true up to lower order terms. The Weyl quantization 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).

40.3   Compositions of ΨDOs

Let $a(x,D)$ and $b(x,D)$ be pseudodifferential operators defined by local symbols of order t and s. It is not always possible to define the composition $a(x,D) \circ b(x,D)$. This is because the image of an element of $\mathcal{D}(\Omega)$ under the operator $b(x,D)$ in general lies in $\mathcal{E}(\Omega)$, and one cannot necessarily apply $a(x,D)$ to elements of $\mathcal{E}(\Omega)$. If we were working on $\mathbb{R}^d$ and working with uniform symbols this wouldn't be a problem, since if $a \in \mathcal{S}^t(\mathbb{R}^d \times \mathbb{R}^d)$ and $b \in \mathcal{S}^t(\mathbb{R}^d \times \mathbb{R}^d)$, then $a(x,D)$ and $b(x,D)$ are both continuous operators from $\mathcal{S}(\mathbb{R}^d)$ to itself, and so the composition is well defined. In our approach, we must make a slightly technical assumption: we assume that either $a(x,D)$ or $b(x,D)$ maps $\mathcal{D}(\mathbb{R}^d)$ to itself, which is true in particular if either $a(x,D)$ or $b(x,D)$ is a *proper* pseudodifferential operator.

Regardless of which method we use, the composition of a ΨDO of order t and a ΨDO of order s will then be 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 40.7. *Let $a(x, \xi)$ and $b(x, \xi)$ be symbols of order t and s, corresponding to operators T_a and T_b. Then $T_a \circ T_b$ is a ΨDO of order $t + s$, and has symbol*

$$(a \circ b)(x, \xi) = e^{2\pi i D_\xi \cdot D_y} \{a(x, \xi) b(y, \eta)\}\big|_{y=x, \eta=\xi}.$$

In particular, we have the asymptotic expansion

$$(a \circ b)(x, \xi) \sim \sum_\alpha \frac{1}{\alpha!} \frac{1}{(2\pi i)^{|\alpha|}} \partial_\xi^\alpha a(x, \xi) \cdot \partial_x^\alpha b(x, \xi).$$

Thus $(a \circ b)(x, \xi) - a(x, \xi) b(x, \xi)$ is a symbol of order $t + s - 1$.

Remark. If we consider the two form

$$\omega = dx \wedge d\xi - d\xi \wedge dx$$

on $T^* \mathbb{R}^n$, then we can define the *Poisson bracket* of two functions on $T^* \mathbb{R}^n$ by setting

$$\{a, b\} = \omega(\nabla a, \nabla b) = \sum_{i=1}^d \frac{\partial a}{\partial \xi^i} \frac{\partial b}{\partial x^i} - \frac{\partial a}{\partial x^i} \frac{\partial b}{\partial \xi^i}.$$

For any two symbols a and b, the result above implies that the commutator $[a(x, D), b(x, D)]$ of the two operators is a pseudodifferential operator of order $t + s$ with some symbol $[a, b]$, such that the symbol $[a, b] - (4\pi i)^{-1} \{a, b\}$ has order $t + s - 2$.

Proof. We can write

$$(T_a \circ T_b) f(x) = \int a(x, \eta) e^{2\pi i \eta \cdot (x-z)} T_b f(z) \, dz \, d\eta$$

$$= \int a(x, \eta) b(z, \xi) e^{2\pi i (\eta - \xi) \cdot (x-z)} e^{2\pi i \xi \cdot (x-y)} f(y) \, dy \, dz \, d\xi \, d\eta.$$

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

$$a(x, \eta) = \sum_\alpha \frac{1}{\alpha!} \partial_\xi^\alpha a(x, \xi) (\eta - \xi)^\alpha$$

651

and

$$b(z, \xi) = \sum_\beta \frac{1}{\beta!} \partial_x^\beta b(x, \xi)(z - x)^\beta.$$

Using the Fourier inversion formula, we calculate that

$$\int (\eta - \xi)^\alpha (z - x)^\beta e^{2\pi i (\eta - \xi) \cdot (x - z)} \, d\eta \, dz$$

$$= \int \tau^\alpha y^\beta e^{-2\pi i \tau \cdot y} \, d\tau \, dy$$

$$= \begin{cases} 0 & \alpha \neq \beta, \\ \alpha!/(2\pi i)^\alpha & \alpha = \beta. \end{cases}$$

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 $t + s$. $\qquad\qquad\square$

Remark. This result shows that $\Phi = \bigcup_t \Phi^t$ and the subfamily $\Phi_{\mathrm{loc,prop}}$ of proper pseudodifferential operators in $\Phi_{\mathrm{loc}} = \bigcup_t \Phi^t_{\mathrm{loc}}$ form graded algebras. We note for any pseudodifferential operator $T \in \Phi^t_{\mathrm{loc}}$, there is a properly supported pseudodifferential operator $\tilde{T} \in \Phi^t_{\mathrm{loc,prop}}$ such that $T - \tilde{T}$ is a smoothing operator. If we define $\dot{\Phi}^t$ to be the space of all operators which differ from an element of Φ^t_{loc} by a pseudodifferential smoothing operator, then $\dot{\Phi} = \bigcup_t \dot{\Phi}^t$ is a graded algebra under addition and composition, which induces, by bijection, an alternate graded algebra structure on $\dot{S}$, agreeing up to first order with the standard algebra structure on this space given by multiplication.

40.4   Parametrices for Elliptic Operators

A *pseudodifferential parametrix* for a pseudodifferential operator T is a pseudodifferential operator S such that $S \circ T$ and $T \circ S$ are the identity operator, modulo smoothing operators. One useful result of our calculations is that we can easily construct *parametrices* for suitable pseudodifferential operators. A symbol $a \in \dot{S}^t(\Omega \times \mathbb{R}^d)$ is called *elliptic* if, locally in the x variable, we can find $R > 0$ such that for $|\xi| \geqslant R$,

$$|a(x,\xi)| \sim \langle\xi\rangle^t,$$

where the implicit constant is also locally uniform. In this case, we can interpret $b_0(x,\xi) = 1/a(x,\xi)$ as an element of $\dot{S}^{-t}$, since the reciprocal is well defined for large ξ, and satisfies the required symbol estimates. By the composition calculus, $1 - (a \circ b_0)(x,\xi)$ is a symbol $r_1 \in \dot{S}^{-1}$. For $i \geqslant 1$, given $r_i \in \dot{S}^{-i}$, if we define $b_i = r_i/a$ in $\dot{S}^{t-i}$, then the composition calculus tells us that $r_{i+1} = 1 - (a \circ (b_0 + \cdots + b_i))$ lies in $\dot{S}^{-i-1}$, and we can continue the iteration. If we consider $b \in \dot{S}^{-t}$ defined by the asymptotic development

$$b \sim \sum_{i=0}^{\infty} b_i$$

then $b(x,D)$ is a right pseudodifferential parametrix for $a(x,D)$. Similarily, we can construct $c \in \dot{S}^{-t}$ such that $c(x,D)$ is a left pseudodifferential parametrix for

$a(x, D)$. But this means that, in $\dot{\mathcal{S}}^{-t}$,

$$b = b \circ (a \circ c) = (b \circ a) \circ c = c$$

so $b = c$, modulo smoothing operators. Thus a left parameterix for an elliptic operator is automatically a right parameterix, and we have constructed such a parametrix.

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, it suffices to analyze the case $t = 0$, since in general we can replace $a(x, \xi)$ with $a(x, \xi)\langle \xi \rangle^{-t}$ and $b(x, \xi)$ with $b(x, \xi)\langle \xi \rangle^t$. If there is $b \in \dot{\mathcal{S}}^0$ such that $b(x, D)$ is a parametrix for $a(x, D)$, then the composition calculus tells us that $1 - a(x, \xi)b(x, \xi)$ is a local symbol of order -1. Thus we find that $|a(x, \xi)b(x, \xi) - 1| \lesssim \langle \xi \rangle^{-1}$ locally uniformly in x, which implies that, locally in x, there exists $R > 0$ such that for $|\xi| \geq R$,

$$|a(x, \xi)b(x, \xi) - 1| \leq 1/2.$$

Thus

$$|a(x, \xi)| \geq \frac{0.5}{|b(x, \xi)|},$$

and combined with the fact that $|b(x, \xi)| \lesssim 1$, we conclude that $|a(x, \xi)| \gtrsim 1$.

It follows from our theory that if T is an elliptic pseudodifferential operator on Ω, then for any $u \in \mathcal{E}(\Omega)^*$, sing-supp$(Tu) =$ sing-supp(u). This is very similar to hypoellipticity, except that for differential operators, this condition can be applied to arbitrary distributions, not just the compactly supported distributions. More generally, we actually find that $\mathrm{WF}(Tu) = \mathrm{WF}(u)$ for $u \in \mathcal{E}(\Omega)^*$ in virtue of the microlocal nature of pseudodifferential operators.

We can use similar asymptotic tools to construct formal fractional powers of an elliptic pseudodifferential operator. For simplicity, we assume we are constructing the fractional powers of an elliptic symbol $a(x, \xi)$ such that $a(x, \xi)$ is never a negative, real number, so that $a(x, \xi)^{p/q}$ is well defined as a principal branch of $z^{p/q}$. Of course, one can consider a similar development for any other choice of branch, assuming an appropriate constraint on the range of a.

Theorem 40.8. *Let $a \in \dot{\mathcal{S}}^t$ be an elliptic symbol such that $a(x, \xi)$ is negative a negative, real number, and suppose that*

$$|arg(a(x, \xi)) - \pi| \gtrsim 1$$

for $\xi \gtrsim 1$, where the implicit constants are locally uniform in x. Then for any pair of positive integers p and q, there exists a unique $b \in \dot{S}^{t(p/q)}$ with principal symbol $a(x, \xi)^{p/q}$, such that $a(x, D)^p = b(x, D)^q$, modulo smoothing operators.

Proof. Let $b_0(x, \xi) = a(x, \xi)^{p/q}$. Then b_0 is a symbol of order $t(p/q)$, and the composition calculus implies that $a(x, D)^p - b_0(x, D)^q$ is a pseudodifferential operator of order $tp - 1$, say, with symbol r_1. Given that we have chosen $b_0, \ldots, b_N$ such that $a(x, D)^p - (b_0 + \cdots + b_N)(x, D)^q$ is a pseudodifferential operator of order $tp - (N + 1)$, say, with symbol r_N. The composition calculus means if we choose $b_{N+1} \in \dot{S}^{t(p/q)-(N+1)}$ by setting

$$b_{N+1} = q^{-1} r_N / (b_0 + \cdots + b_N)^{q-1}.$$

Here we rely on the fact that a is elliptic, so that the denominator is non-vanishing for large ξ. Then

$$a(x, D)^p - (b_0 + \cdots + b_N + b_{N+1})^q$$

is a pseudodifferential operator of order $tp - (N + 2)$, allowing us to continue the construction. If we pick $b \sim \sum_{n=0}^{\infty} b_n$, then $a(x, D)^p - b(x, D)^q$ will be a smoothing operator. Moreover, it is clear that the choice of such symbols at each step is essentially unique, which shows that b is unique up to a smoothing operator. $\quad\square$

Given the assumptions of the theorem, we denote the unique operator b by $a^{p/q}$. Using a^{-1} instead of a, one can also construct negative fractional powers of the symbol a. It is simple to see that for $r_1, r_2 \in \mathbb{Q}$, we have $a^{r_1} \circ a^{r_2} = a^{r_1 + r_2}$.

40.5 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, maps $H^s(\mathbb{R}^d)$ to $H^{s-m}(\mathbb{R}^d)$ for each s. This feature remains true for a general pseudodifferential operator.

Theorem 40.9. *For any $a \in \mathcal{S}^t_{loc}(\mathbb{R}^d)$, $a(x, D)$ extends uniquely to a continuous operator from $H^s_c(\mathbb{R}^d)$ to $H^{s-t}_{loc}(\mathbb{R}^d)$.*

Proof. Let T have symbol $a(x, \xi)$. Without loss of generality, since we need only prove local estimates in the output we may assume that a is compactly supported in the x-variable. Then a is uniformly integrable in the x-variable, and we let

$$A(\lambda, \xi) = \int a(x, \xi) e^{-2\pi i \lambda \cdot x} \, dx$$

denote the Fourier transform of a in the x-variable. For any input $\phi \in \mathcal{S}(\mathbb{R}^d)$, $T\phi \in \mathcal{S}(\mathbb{R}^d)$, and we may calculate that

$$\widehat{T\phi}(\lambda) = \int A(\lambda - \xi, \xi) \hat{\phi}(\xi) \, d\xi.$$

The assumptions on the symbol a imply that

$$|A(\lambda, \xi)| \lesssim_N \langle \xi \rangle^t \langle \lambda \rangle^{-N}.$$

Without loss of generality, assume that $s = t$. Applying Schur's lemma, the operator

$$Sf(\lambda) = \int \langle \lambda - \xi \rangle^{-N} f(\xi) \, d\xi,$$

somewhat analogous to the Hardy-Littlewood-Sobolev fractional integration operator, is bounded from $L^2(\mathbb{R}^d)$ to itself provided that $N > d$. But this means that if we pick $N > d$, then

$$\begin{aligned}
\|T\phi\|_{L^2(\mathbb{R}^d)} &\lesssim_N \left\| \int \langle \lambda - \xi \rangle^{-N} \langle \xi \rangle^t \hat{\phi}(\xi) \, d\xi \right\|_{L^2(\mathbb{R}^d)} \\
&\lesssim \|\langle \xi \rangle^t \hat{\phi}\|_{L^2(\mathbb{R}^d)} \\
&\lesssim \|\phi\|_{H^t(\mathbb{R}^d)}.
\end{aligned}$$

Thus T is bounded from $H^t(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$. But for general $s \in \mathbb{R}^d$, T will be bounded from $H^s(\mathbb{R}^d)$ to $H^{s-t}(\mathbb{R}^d)$ if and only if $(1 - \Delta)^{s-t} T (1 - \Delta)^{t-s}$ is bounded from $H^t(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$, and this follows because $(1 - \Delta)^{s-t} T (1 - \Delta)^{t-s}$ is also a pseudodifferential operator of order t. The bounds we have proven show that there is a unique extension of T to $H^s(\mathbb{R}^d)$ for all $s \in \mathbb{R}^d$ such that $\|Tf\|_{H^{s-t}(\mathbb{R}^d)} \lesssim \|f\|_{H^s(\mathbb{R}^d)}$ for all $f \in H^s(\mathbb{R}^d)$. The closed graph theorem shows that this extension agrees with the definition of Tf given for any compactly supported f, which we can view as an element of $\mathcal{E}(\mathbb{R}^d)^*$. $\qquad\square$

Remark. For symbols of order zero, a simpler proof follows by writing

$$T^\lambda \phi(x) = \int A(\lambda, \xi) \hat{\phi}(\xi) e^{2\pi i (\lambda + \xi) \cdot x} \, d\xi.$$

then the Fourier inversion formula shows that

$$T\phi(x) = \int T^\lambda \phi(x) \, d\lambda.$$

Now the operators $\{T^\lambda\}$ are just Fourier multiplier operators with symbols $\{m_\lambda\}$, where $m_\lambda(\xi) = A(\lambda, \xi) e^{2\pi i \lambda \cdot x}$, the bounds on A imply that $\|m_\lambda\|_{L^\infty} \lesssim \langle \lambda \rangle^{-N}$, and this immediately gives boundedness from $L^2(\mathbb{R}^d)$ to itself. The general result follows from the same trick as in the end of the last proof using the composition calculus.

Remark. If a is *properly supported* pseudodifferential operator of order t, then $a(x, D)$ extends to a continuous operator from $H^s_{\text{loc}}(\Omega)$ to $H^{s-t}_{\text{loc}}(\Omega)$.

Using Calderon-Zygmund theory, we can obtain better estimates. Let us restrict ourselves at first to pseudodifferential operators of order 0. The kernel of a ΨDO of order zero satisfies estimates of the form

$$|K(x, y)| \lesssim \frac{1}{|x - y|^d} \quad \text{and} \quad |\nabla_y K(x, y)| \lesssim \frac{1}{|x - y|^{d+1}}.$$

Thus K is a singular kernel. We therefore 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 40.10. *If $a \in S^0(\mathbb{R}^d \times \mathbb{R}^d)$, then for any $f \in S$,*

$$\|T_a f\|_{L^2(\mathbb{R}^d)} \lesssim \|f\|_{L^2(\mathbb{R}^d)}.$$

Consequently, for any $1 < p < \infty$,

$$\|T_a f\|_{L^p(\mathbb{R}^d)} \lesssim_p \|f\|_{L^p(\mathbb{R}^d)}.$$

Proof. If $\text{supp}_x(a)$ is compact, then we have already proven this result. To prove the result for more general symbols, we work with a kernel representation of T_a. Thus we write

$$T_a f(x) = \int_{\mathbb{R}^d} K(x, y) f(x) \, dx,$$

657

where

$$K(x, y) = \int_{\mathbb{R}^d} a(x, \xi) e^{2\pi i \xi \cdot (x-y)} \, d\xi.$$

We have already shown that the kernel K is C^∞ away from the diagonal, and decays rapidly away from that 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 $x_0 \in \mathbb{R}^d$,

$$\int_{|x-x_0| \leqslant 1} |T_a f(x)|^2 \, dx \lesssim_N \int_{\mathbb{R}^d} \frac{|f(x)|^2}{\langle x - x_0 \rangle^N} \, dx.$$

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 cutoff supported on $|x| \leqslant 3$, then the symbol $\eta(x) a(x, \xi)$ is compactly supported, and so

$$\int_{|x| \leqslant 1} |T_a f_1(x)|^2 \, dx = \int_{|x| \leqslant 1} |T_{\eta a} f_1(x)|^2 \lesssim \|f_1\|_{L^2(\mathbb{R}^d)}^2$$

$$\lesssim_N \int_{\mathbb{R}^d} \frac{|f_1(x)|^2}{\langle x \rangle^N} \, dx \leqslant \int_{\mathbb{R}^d} \frac{|f(x)|^2}{\langle x \rangle^N} \, dx.$$

On the other hand, since $f_2(x)$ is supported on $|x| \geqslant 2$, we find that

$$\int_{|x| \leqslant 1} |T_a f_2(x)|^2 \, dx = \int_{|x| \leqslant 1} \left| \int K(x, y) f_2(y) \, dy \right|^2 dx$$

$$\leqslant \int \int_{|x| \leqslant 1} |K(x, y)|^2 |f_2(y)|^2 \, dy \, dx$$

$$\lesssim_N \int \int_{|x| \leqslant 1} \frac{|f_2(y)|^2}{|x - y|^N} \, dy \, dx$$

$$\lesssim \int \int_{|x| \leqslant 1} \frac{|f_2(y)|^2}{\langle y \rangle^N} \, dy$$

$$\lesssim \int \frac{|f_2(y)|^2}{\langle y \rangle^N} \, dy.$$

658

But we now find that if $N > d$, then

$$\int |T_a f(x)|^2 \, dx \lesssim \int \int_{|x-y|\leqslant 1} |T_a f(y)|^2 \, dy \, dx$$

$$\lesssim_N \int \int \frac{|f(y)|^2}{\langle x-y \rangle^N} \, dy \, dx$$

$$\lesssim \int |f(y)|^2 \, dy,$$

which gives L^2 boundedness. $\qquad\square$

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

$$\|T_a f\|_{L^p_s(\mathbb{R}^d)} \lesssim_{p,s} \|f\|_{L^p_{t+s}(\mathbb{R}^d)}.$$

If T_a is an elliptic pseudodifferential operator of order t, then it has a parametrix S of order $-t$, and so for any $r > 0$

$$\|f\|_{L^p_{t+s}(\mathbb{R}^d)} = \|S T_a f + (I - S T_a)f\|_{L^p_{t+s}(\mathbb{R}^d)}$$

$$\leqslant \|S T_a f\|_{L^p_{t+s}(\mathbb{R}^d)} + \|(I - S T_a)f\|_{L^p_{t+s}(\mathbb{R}^d)}$$

$$\lesssim_{p,s,r} \|T_a f\|_{L^p_s(\mathbb{R}^d)} + \|f\|_{L^p_{-r}(\mathbb{R}^d)}$$

Thus T is *almost* invertible as a map from $L^p_{t+s}(\mathbb{R}^d)$ to $L^p_t(\mathbb{R}^d)$, except that we cannot quite obtain the bound $\|Tf\|_{L^p_s(\mathbb{R}^d)} \sim_{p,s} \|f\|_{L^p_{t+s}(\mathbb{R}^d)}$, for instance, by virtue of the fact that T might not even be invertible. But we can obtain a less quantitative result.

Theorem 40.11. *Let T be an elliptic pseudodifferential operator given by a symbol uniformly of order t. If $f \in \mathcal{E}(\mathbb{R}^d)^*$, and Tf lies in $L^p_s(\mathbb{R}^d)$, then f lies in $L^p_{s+t}(\mathbb{R}^d)$.*

Proof. Since T is elliptic, we can find a parametrix S, which is a pseudodifferential operator of order $-t$. Thus there exists a smoothing operator U such that $1 = ST + U$. Since $Tf \in L^p_s(\mathbb{R}^d)$, $STf \in L^p_{s+t}(\mathbb{R}^d)$. Since f is compactly supported, $Uf \in C^\infty(\mathbb{R}^d)$, and thus in $L^p_{s+t,\mathrm{loc}}(\mathbb{R}^d)$. This means $f \in L^p_{s+t,\mathrm{loc}}(\mathbb{R}^d)$, and since f is compactly supported, this means $f \in L^p_{s+t}(\mathbb{R}^d)$. $\qquad\square$

40.6 Pseudodifferential Operators on Manifolds

It is an important fact that the class of pseudodifferential operators whose kernels are compactly supported is invariant under a change of coordinates, modulo smoothing operators. By an analysis of how these pseudodifferential operators change under a change of coordinates, we will be able to obtain a theory of pseudodifferential operators on manifolds.

Let $\kappa : U \to V$ be a diffeomorphism, where U and V are open sets in $\mathbb{R}^n$, and consider $a \in S^t_{\mathrm{loc}}(U)$ which gives a pseudodifferential operator $T : \mathcal{D}(U) \to \mathcal{E}(U)$ of order t. Define an operator $S : \mathcal{D}(V) \to \mathcal{E}(V)$ by setting

$$S f(\kappa(x)) = T(f \circ \kappa)(x) = \int a(x,\xi) e^{2\pi i \xi \cdot (x-y)} f(\kappa(y)) \, dy \, d\xi,$$

i.e. S is defined such that $\kappa_* \circ S = T \circ \kappa_*$. One can verify that S is pseudolocal, i.e. that it does not expand the wavefront sets of inputs, so one might expect S to be a pseudodifferential operator, of order t as well. By localizing, we may assume that $\mathrm{supp}_x(a)$ is compact, so that T extends to a continuous operator from $\mathcal{D}(U)^*$ to $\mathcal{D}(U)^*$, and S to a continuous operator from $\mathcal{D}(V)^*$ to $\mathcal{D}(V)^*$. If a symbol $b \in S^t_{\mathrm{loc}}(V)$ existed such that $S = b(y, D)$, then we would find

$$b(\kappa(x_0), \eta) = \left\{ S e^{2\pi i \eta \cdot z} \right\} (\kappa(x_0)) = \int a(x_0, \xi) e^{2\pi i [\xi \cdot (x_0 - x) + \eta \cdot \kappa(x)]} \, dx \, d\xi$$

Provided we can show that the right hand side defines a symbol $b \in S^t_{\mathrm{loc}}(V)$, then $b(y, D)$ will be a continuous operator whose kernel is compactly supported in $V \times V$, and the fact that $S e^{2\pi i \eta \cdot z} = b(y, D) e^{2\pi i \eta \cdot z}$ for all η implies, by continuity and the fact that linear combinations of exponentials are dense in $\mathcal{D}(V)^*$, that $S = b(y, D)$. Thus, given $a \in S^t_{\mathrm{loc}}(U)$, such that $\mathrm{supp}_x(a)$ is compact, our goal is an analysis of the function

$$b(\kappa(x_0), \eta) = \int a(x_0, \xi) e^{2\pi i [\xi \cdot (x_0 - x) + \eta \cdot \kappa(x)]} \, dx \, d\xi.$$

If $\phi \in \mathcal{D}(V)$ is a bump function equal to one in a neighborhood of the projections of the support of the kernel of S onto each coordinate, then

$$b(\kappa(x_0), \eta) = \left\{ \phi S \left\{ \phi \cdot e^{2\pi i \eta \cdot z} \right\} \right\} (\kappa(x_0)),$$

which implies that $b \in C^\infty(T^*V)$. If we set

$$\Phi(\xi, \eta) = \int \phi(x) e^{2\pi i [\eta \cdot \kappa(x) - \xi \cdot x]} \, dx,$$

then

$$b(\kappa(x_0),\eta) = \int a(x_0,\xi)\Phi(\xi,\eta)e^{2\pi i \xi \cdot x_0}\, d\xi.$$

Now Φ is a standard oscillatory integral, whose phase $\phi(x) = \eta \cdot \kappa(x) - \xi \cdot x$ has a stationary point for values of x such that $D\kappa(x)^T\eta = \xi$. Since $D\kappa(x)^T$ is invertible, this can only happen when $|\eta| \sim |\xi|$. More precisely, suppose $|D\kappa(x)|, |D\kappa(x)^{-1}| \leqslant C$. Then $|D\kappa(x)^T\eta - \xi| \geqslant |\xi|/2 \gtrsim_C |\xi| + |\eta|$, and $|D\kappa(x)^T\eta - \xi| \geqslant |\eta|/2 \gtrsim_C |\xi| + |\eta|$ for $|\xi| \leqslant |\eta|/2C$. Thus unless $2C|\eta| \geqslant |\xi| \geqslant |\eta|/2C$, we conclude by the principle of nonstationary phase that for each $N > 0$,

$$|\Phi(\xi,\eta)| \lesssim_N [1 + |\xi| + |\eta|]^{-N},$$

where the implicit constant will be uniformly bounded over a family of diffeomorphisms κ and a family of pseudodifferential operators T if the kernels of the resulting operators S are uniformly supported on a common compact subset of $V \times V$, and if we have uniform upper and lower bounds on the derivative of κ on the support of these kernels. If we write $1 = \chi_1 + \chi_2$, for $\chi_1, \chi_2 \in C^\infty(\mathbb{R}^d)$ with $\chi_1(\alpha)$ supported on $1/2C \leqslant |\alpha| \leqslant 2C$ and equal to one when $1/C \leqslant |\alpha| \leqslant C$, then this induces a decomposition $b(\kappa(x_0),\eta) = b_1(\kappa(x_0),\eta) + b_2(\kappa(x_0),\eta)$, where

$$b_i(\kappa(x_0),\eta) = \int \chi_i(\xi/|\eta|)a(x_0,\xi)\Phi(\xi,\eta)e^{2\pi i \xi \cdot x_0}\, d\xi.$$

The bound on $\Phi(\xi,\eta)$ above implies that $b_1 \in S^{-\infty}$. To analyze b_2, we apply the method of stationary phase. To simplify our formulas, let us assume without loss of generality that $x_0 = \kappa(x_0) = 0$ and that $B = D\kappa(x_0)$. If we fix $|\eta| = 1$, consider $\lambda > 1$, and do a change of variables, replacing ξ with $\lambda(\xi + B^T\eta)$, we find that

$$b_2(\kappa(x_0),\lambda\eta) = \lambda^d \int \psi(x,\xi)a(x,\lambda B^T\eta + \xi))e^{2\pi i \lambda \phi(x,\xi)}\, d\xi\, dx,$$

where $\psi(\xi,x) = \chi_2(\xi + B^T\eta)\phi(x_0 + x)$ and $\phi(x,\xi) = \eta \cdot (\kappa(x) - Bx) - \xi \cdot x$. This is a continuous family of non-degenerate stationary phase integrals, each with a unique stationary point when $(x,\xi) = (0,0)$. Since ψ is equal to one in a neighborhood of this point, it will not show up in the corresponding asymptotics. The Hessian at the origin is precisely $A + B$, where

$$A = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix}$$

661

and

$$C = \begin{pmatrix} \tilde{C} & 0 \\ 0 & 0 \end{pmatrix}$$

where $\tilde{C} = \tilde{C}(\eta)$ is the Hessian of the map $x \mapsto \eta \cdot \kappa(x)$ at the origin. If we set

$$r(x) = \eta \cdot (\kappa(x) - Bx) - (1/2)(x^T A x)$$

which satisfies $\partial^\alpha r(0) = 0$ for all $|\alpha| \leq 2$, then it follows that if $a_\lambda(x_0, \xi) = a(x_0, \lambda \xi)$, then

$$b_2(\kappa(x_0), \lambda\eta) \sim \sum_{2\nu \geq 3\mu} \frac{1}{(-2)^\nu} \frac{1}{\mu! \nu!} \frac{1}{(2\pi i \lambda)^{\nu-\mu}} \langle (A+B)^{-1}\nabla, \nabla \rangle^\nu \{r^\mu a_\lambda\}(x, D\kappa(x)^T).$$

The next Lemma implies that we can write this asymptotic development as

$$b_2(\kappa(x_0), \lambda\eta) \sim \sum_{\nu=0}^{\infty} \langle i \cdot \nabla_x / 2\pi, \nabla_\xi / \lambda \rangle^\nu \left\{ e^{2\pi i \lambda r(x)} a_\lambda(0, \xi) \right\}_{x=0, \xi=B^T \eta}$$

where if we sum over $\nu \leq N$, then the error term will be $O(\lambda^{(d-N)/2})$.

Lemma 40.12. *Let A be a symmetric, invertible matrix, let B be a symmetric matrix, and suppose $\det(A + tB)$ is independent of t. Then there exists k such that $(A^{-1}B)^k = 0$ for some k. For this k, and for any $N > 0$,*

$$\sum_{j<N} \frac{1}{j!} (2i\lambda)^{-j} \langle (A+B)^{-1}\nabla, \nabla \rangle^j u(0)$$

$$= \sum_{j<N} \frac{1}{j!} (2i\lambda)^{-j} \langle A^{-1}\nabla, \nabla \rangle^j \left\{ e^{2\pi i \lambda x^T B x / 2} u \right\}(0) + O(\lambda^{-N/k}).$$

Proof. See Hormander, Vol 1. Section 7.7. □

Thus we have proved the following representation formula for pseudodifferential operators under changes of coordinates.

Theorem 40.13. *Let $\kappa : U \to V$ be a diffeomorphism, where U and V are open sets in $\mathbb{R}^n$, and suppose $T = a(x, D)$ is a pseudodifferential operator of order t on U. If S is an operator defined on V defined such that $\kappa_* \circ S = T \circ \kappa_*$, then S is a pseudodifferential operator of order t, whose symbol a_κ has the asymptotic development*

$$a_\kappa(\kappa(x), \eta) \sim \sum \frac{1}{(2\pi i)^\alpha} \frac{1}{\alpha!} (\partial_\xi^\alpha a)(x, D\kappa(x)^T \eta) \left\{ \partial_y^\alpha e^{2\pi i \eta \cdot r_x(y)} \right\} \Big|_{y=x},$$

where $r_x(y) = \kappa(y) - \kappa(x) - D\kappa(x)(y-x)$. The term in the sum corresponding to a multi-index α is a symbol of order $t - \lceil |\alpha|/2 \rceil$, and thus this really is an asymptotic expansion.

Pseudodifferential operators defined by a local symbol of order t are therefore invariant under a change of coordinates, modulo smoothing operators. Moreover, if $\kappa_* \circ S = T \circ \kappa_*$, where T is a ΨDO with symbol $a \in S^t_{\text{loc}}(U)$, and S is a ΨDO with symbol $b \in S^t_{\text{loc}}(V)$, then

$$b(y, \eta) - a(\kappa^{-1}(y), D\kappa(x)^{-T} \cdot \eta)$$

is a pseudodifferential operator of order $t - 1$.

Given a manifold M, a continuous operator $T : \mathcal{D}(M) \to \mathcal{E}(M)$ is called a *pseudodifferential operator of order t* if whenever (x, U) is a coordinate chart on M, the operator $T_x : \mathcal{D}(x(U)) \to \mathcal{E}(x(U))$ given by applying T in the coordinate chart, is a pseudodifferential operator of order t on $x(U)$. We let $\Psi^t_{\text{loc}}(M)$ denote the family of operators of this form. The next Lemma shows that when $M = U$ is an open subset of $\mathbb{R}^n$, this really is the family of pseudodifferential operators defined by symbols in $\mathcal{S}^t_{\text{loc}}(T^*U)$.

Lemma 40.14. *Suppose $T : \mathcal{D}(U) \to \mathcal{E}(U)$ is a pseudodifferential operator on U of order t, where U is viewed as a manifold as above. Then we can find a symbol $a(x, \xi) \in \mathcal{S}^t_{loc}(U \times \mathbb{R}^d)$ such that $T - a(x, D)$ is a smoothing operator. The symbol a is uniquely determined up to a symbol in $\mathcal{S}^{-\infty}(\mathbb{R}^d \times \mathbb{R}^d)$.*

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. $\square$

Remark. Under the assumption that the operator has a kernel which is smooth away from the diagonal, an operator $T : \mathcal{D}(M) \to \mathcal{E}(M)$ is a pseudodifferential operator if and only if it is pseudodifferential operator when transferred in coordinates on a family of coordinates charts that form an atlas for M. The assumption on the kernel is needed, for instance, since if we take $M = \mathbb{R}$, we consider an atlas of the form $\{(n, n+1)\}$, and we consider $Tf(x) = f(x-2)$, then T vanishes on each of these coordinates charts, and so looks to be a pseudodifferential operator in these charts, whereas T clearly is not a pseudodifferential operator on $\mathbb{R}$ since it is not pseudolocal.

It is often simpler to discuss operators that can be asymptotically expanded in terms of homogeneous symbols. That is, we wish to discuss symbols $a \in \dot{\mathcal{S}}^t$ such that there exist a sequence of symbols $\{a_k\}$, where $a_k(x, \xi)$ is smooth away from $\xi = 0$, homogeneous of order $t - k$, and

$$a \sim \sum_{k=0}^{\infty} a_k.$$

Such symbols are called *classical*, and denoted $\mathcal{S}_{\mathrm{cl}}^t$, since this, roughly speaking, was the family of pseudodifferential operators initially studied by Kohn and Nirenberg. Given a manifold M, we write $\Psi_{\mathrm{cl}}^t(M) = \mathrm{Op}(\mathcal{S}_{\mathrm{cl}}^t)(M)$ for the class of all pseudodifferential operators T such that T_x is classical for any coordinate system (x, U) on M. It suffices to check this on a cover because of the asymptotic expansion for the change of variables formula. The homogeneous function which agrees with the leading term in the expansion is invariant under coordinate changes if we interpret it as a function on $T^*M - 0_M$, so for a classical pseudodifferential operator T in $\Psi_{\mathrm{cl}}^t(M)$, we can define the *principal symbol* $a \in C^\infty(T^*M - 0_M)$ to be that homogeneous function of order t which agrees with the leading term in the asymptotic expansions above. For nonclassical operators, there is no canonical choice of a principal symbol, though one can consider the principal symbol as an element of $\mathcal{S}_{\mathrm{loc}}^t(M)/\mathcal{S}_{\mathrm{loc}}^{t-1}(M)$. A pseudodifferential operator is then called *elliptic* if it's principal symbol is nonvanishing at a suitably distance away from the origin, or equivalently, if it is elliptic in coordinates ranging over the entirety of M.

Example. *Let* $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. *For any 'symbol of order t' on* $\mathbb{T}^d$, *i.e. any function* $a : \mathbb{T}_x^d \times \mathbb{Z}_\xi^d \to \mathbb{C}$ *with extends to an element of* $\mathcal{S}^t(\mathbb{T}_\xi^d \times \mathbb{R}_\xi^d)$, *we can consider the operator* $T : \mathcal{E}(\mathbb{T}^d) \to \mathcal{E}(\mathbb{T}^d)$ *defined by setting*

$$T\phi(x) = \sum_{\xi \in \mathbb{Z}^d} a(x, \xi)\hat{\phi}(\xi)e^{2\pi i \xi \cdot x},$$

where $\hat{\phi} : \mathbb{Z}^d \to \mathbb{C}$ *is the Fourier transform of ϕ on* $\mathbb{T}^d$. *We claim that T is a pseudodifferential operator. By translation invariance, it will suffice to find a neighborhood Ω of the origin upon and a coordinate system on Ω upon which the*

transfer of T is a pseudodifferential operator. Write

$$T\phi(x) = \sum_{\xi} a(x,\xi)\hat{\phi}(\xi)e^{2\pi i\xi \cdot x}$$

$$= \int_{\mathbb{T}^d} \left(\sum_{\xi} a(x,\xi)e^{2\pi i\xi \cdot (x-y)} \right) \phi(y) \, dy$$

$$= \int_{\mathbb{T}^d} K(x,y)\phi(y) \, dz,$$

where we treat

$$K(x,y) = \sum_{\xi} a(x,\xi)e^{2\pi i\xi \cdot (x-y)}.$$

as the distribution on $\mathcal{D}(\mathbb{T}^d \times \mathbb{T}^d)^*$ obtained as the distributional limit of the partial sums. Consider the periodic kernel $\tilde{K} \in \mathcal{D}(\mathbb{R}^d \times \mathbb{R}^d)^*$ induced by K. Then the Poisson summation formula implies that

$$\tilde{K}(x,y) = \sum_{\xi} a(x,\xi)e^{2\pi i\xi \cdot (x-y)} = \sum_{n}(\mathcal{F}_{\xi}^{-1}a)(x,(x-y)+n) = \sum_{n} K_a(x,y+n),$$

where K_a is the kernel of the psedodifferential operator $a(x,D)$. If $\tilde{\Omega} = (-1/4, 1/4)^d$, then for $x, y \in \tilde{\Omega}$, and $n \neq 0$,

$$\left| \partial_x^\alpha \partial_y^\beta K_a(x,y+n) \right| \lesssim_N |n|^{-N}.$$

This implies that

$$(x,y) \mapsto \sum_{n \neq 0} K_a(x,y+n)$$

lies in $C^\infty(\tilde{\Omega} \times \tilde{\Omega})$, and so induces a smoothing operator on $\tilde{\Omega}$. Thus $\tilde{T}$, restricted to $\tilde{\Omega}$, differs by a smoothing operator from the pseudodifferential operator $a(x,D)$. But this means that $\tilde{T} \in \dot{\Psi}^t(\tilde{\Omega})$. If Ω is the subset of $\mathbb{T}^d$ corresponding to $\tilde{\Omega}$, then this means T behaves like a pseudodifferential operator on $\tilde{\Omega}$. Thus T is actually a pseudodifferential operator of order t on $\mathbb{T}^d$. Modulo smoothing operators, working backwards through this argument also clearly shows that all pseudodifferential operators *of order t can be written in the form introduced at the beginning of this example.*

    As a particular example of this kind of construction, consider the operator $T : \mathcal{E}(\mathbb{T}) \to \mathcal{E}(\mathbb{T})$ given by

$$P_+\phi(t) = \sum_{n>0} \hat{\phi}(n)e^{2\pi nit},$$

then P_+ is a classical pseudodifferential operator on $\mathbb{T}$ of order zero, and it's principal symbol is $\mathbf{I}(\xi > 0)$. Similarily, if

$$P_-\phi(t) = \sum_{n<0} \widehat{\phi}(n)e^{2\pi nit},$$

then P_- is a classical pseudodifferential operator on $\mathbb{T}$ of order zero with principal symbol $\mathbf{I}(\xi < 0)$. For any non-zero complex values $a_+, a_- \in \mathbb{C} - \{0\}$, $a_+P_+ + a_-P_-$ is an elliptic classical pseudodifferential operator on $\mathbb{T}$ of order zero.

Example. *On a Riemannian manifold M, the operator $-\Delta$ is a pseudodifferential operator of order two, given in local coordinates as*

$$a(x,\xi) = \sum g^{ij}(x)\xi_i\xi_j.$$

If M is geodesically complete, then $-\Delta$ can be interpreted as an essentially self-adjoint operator on $L^2(M)$, and the resulting calculus of unbounded operators allows us to define the operator $\sqrt{-\Delta}$. Regardless, the fact that $-\Delta$ is an elliptic pseudodifferential operator allows us to define a pseudodifferential operator $\sqrt{-\Delta}$ of order one, such that $\sqrt{-\Delta}^2 = -\Delta$. This operator will agree with the definition given by the unbounded calculus, modulo smoothing operators, if M is geodesically complete.

40.7 ΨDOs and Microsupport

Fix an open set $\Omega \subset \mathbb{R}^d$. The theory of microlocal analysis asks us to think of the singularities of a distribution $u \in \mathcal{D}(\Omega)^*$ as having both position *and* direction, i.e. existing on the wavefront set $\mathrm{WF}(u) \subset T^*M$, and indicated by the directions that the Fourier transform of u does not decay rapidly, once sufficiently localized. The heuristics of pseudodifferential operators also ask us to view a distribution u as living on T^*M, such that a pseudodifferential operator S given by a symbol a acts on u by multiplying that part of u that 'lives at' the point (x, ξ) by the quantity $a(x, \xi)$. In particular, if a is made to decay rapidly in the directions that define the wavefront set u, then we should expect Su to be smooth, i.e. application of S annihilates the singularities of a. In this section, we elaborate on these ideas, as well as introducing further notions of the microlocal properties of pseudodifferential operators.

If $T : \mathcal{D}(\Omega) \to \mathcal{E}(\Omega)$ is a pseudodifferential operator with symbol a, then we define the microsupport μ-supp(T) of T to be equal to the microsupport μ-supp(a)

of it's symbol, defined in the parts of these notes on oscillatory integral distributions.

Lemma 40.15. *Let T be a pseudodifferential operator on Ω. Then it's canonical relation is equal to*

$$C_T = \{(x, x; \xi, \xi) : (x, \xi) \in \mu\text{-}supp(T)\}.$$

This follows because an open conic set $\Gamma \subset \Omega \times \mathbb{R}^d$ is disjoint from μ-supp(T) if and only if Γ is disjoint from WF(Tu) for any $u \in \mathcal{E}(\Omega)^*$.

Proof. Let $T = a(x, D)$ for some symbol $a(x, \xi)$. It follows from the general theory of oscillatory integral distributions that the canonical relation of T is equal to

$$C_T \subset \{(x, \xi; x, \xi) : (x, \xi) \in \mu\text{-supp}(T)\}.$$

Thus if Γ is disjoint from μ-supp(T), then it follows from the general theory that Γ is disjoint from WF(Tu) for any $u \in \mathcal{E}(\Omega)^*$. If we assume the converse, then for any $(x_0, \xi_0) \in \Gamma$, we can find a pseudodifferential operator $S : \mathcal{D}(\Omega) \to \mathcal{E}(\Omega)$ with a symbol $b(x, \xi)$ supported on Γ and equal to one in a conic neighborhood of (x_0, ξ_0). It follows that ST is a smoothing operator. It follows from the composition calculus that $a \cdot b \in \mathcal{S}^{-\infty}(\Omega \times \mathbb{R}^d)$. Thus $(x_0, \xi^0) \notin \mu$-supp$(a) = \mu$-supp(T). $\qquad\square$

Similar to the argument above, the composition calculus implies that for any two properly supported pseudodifferential operators T and S,

$$\mu\text{-supp}(TS) \subset \mu\text{-supp}(T) \cap \mu\text{-supp}(S),$$

and for any pseudodifferential operator T,

$$\mu\text{-supp}(T^t) = \{(x, -\xi) : (x, \xi) \in \mu\text{-supp}(T)\} \quad \mu\text{-supp}(T^*) = \mu\text{-supp}(T).$$

One can prove these either using the microsupport of the symbols defining T and S, or the properties of the canonical relation of T and S.

Theorem 40.16. *Let Γ be a closed conic set, and suppose T is a properly pseudodifferential operator with μ-supp$(T) \cap \Gamma = \varnothing$, then T maps $\mathcal{D}_\Gamma^*(\Omega)$ continuously into $\mathcal{E}(\Omega)$.*

Proof. TODO: For any open set $U \subset \Omega$, control the derivatives of $T\phi$ on U by decomposing T into inputs outside of U, a ΨDO with a symbol in $\mathcal{S}^{-\infty}$, and a ΨDO localized near μ-supp(T), and so on. $\qquad\square$

In fact, a sequence $\{u_n\}$ converges in $\mathcal{D}_\Gamma^*(\Omega)$ to some $u \in \mathcal{D}_\Gamma^*(\Omega)$ if and only if it converges distributionally to u, and $Tu_n \to Tu$, where T is an arbitrary properly supported ΨDO with $\Gamma \cap \mu\text{-supp}(T) = \varnothing$. The proof is left to the reader.

Theorem 40.17. *Fix $u \in \mathcal{D}(\Omega)^*$. Then $(x_0, \xi^0) \notin WF(u)$ if and only if there exists a conic neighborhood Γ of (x_0, ξ^0) such that for any properly supported pseudodifferential operator T on Ω with $(x_0, \xi^0) \in \mu\text{-supp}(T)$, $Tu \in \mathcal{E}(\Omega)$.*

Remark. It follows from this theorem, and the fact that the class of pseudodifferential operators and the microsupport of such operators are invariant under diffeomorphism, that the wavefront set of a distribution is invariant under diffeomorphisms.

We have already seen the regularity theory of pseudodifferential operators. Recall that if T is a properly supported pseudodifferential operator of order t, then T maps $H_{\text{loc}}^s(\Omega) \to H_{\text{loc}}^{s-t}(\Omega)$. By virtue of the fact that this is an isomorphism if T is elliptic, one can *define* $H_{\text{loc}}^s(\Omega)$ to be the space of all distributions $u \in \mathcal{D}(\Omega)^*$ such that $Tu \in L_{\text{loc}}^2(\Omega)$, where T can be an arbitrary properly supported ΨDO of order m.

Let us also now look at the *microlocal regularity* of T, i.e. the regularity of T where we only care about regularity in certain conical subsets of $T^*\Omega$. For a conic open set $\Gamma \subset \Omega \times \mathbb{R}^d$, we define $H_{\Gamma,\text{loc}}^s(\Omega)$ to be the family of all distributions $u \in \mathcal{D}(\Omega)^*$ such that for any properly supported pseudodifferential operator T of order zero with conically compact microsupport $\mu\text{-supp}(T) \subset \Gamma$, $Tu \in H_{\text{loc}}^s(\Omega)$. By the results above, and the Sobolev embedding theorem, $\lim_{s \to \infty} H_{\Gamma,\text{loc}}^s(\Omega) = \mathcal{D}_\Gamma^*(\Omega)$.

Theorem 40.18. *If T is a properly supported pseudodifferential operator of order t on Ω, then T maps $H_{\Gamma,loc}^s(\Omega)$ into $H_{\Gamma,loc}^{s-t}(\Omega)$.*

Proof. Fix a conically compact set $\Gamma_1 \subset \Gamma$. If S is a properly supported pseudodifferential operator of order zero with $\mu\text{-supp}(S) \subset \Gamma$, it suffices to show that ST maps $H_{\Gamma,\text{loc}}^s(\Omega)$ into $H_{\text{loc}}^{s-t}(\Omega)$. But $ST = STU_1 + STU_2$, where U_1 and U_2 are properly supported pseudodifferential operators of order zero, such that $\mu\text{-supp}(U_1)$ is conically compact and contained in Γ, and $\mu\text{-supp}(U_2)$ is disjoint from Γ_1. But this means that if $v \in H_{\Gamma,\text{loc}}^s(\Omega)$, then $U_1 v \in H_{\text{loc}}^s(\Omega)$, and since ST is a properly support pseudodifferential operator of order t, $STU_1 v \in H_{\text{loc}}^{s-t}(\Omega)$. On the other hand, $WF(U_2 v)$, and thus $WF(TU_2 v)$, is disjoint from Γ_1, which implies that STU_2 is smooth, and thus lies in $H_{\text{loc}}^{s-t}(\Omega)$ trivially. $\qquad \square$

Ellipticity can also be microlocalized. If T was an elliptic pseudodifferential operator, we found a pseudodifferential parametrix S for T, i.e. such that $S \circ T$ and $T \circ S$ are smoothing operators. We say a symbol $a \in \dot{S}^t(\Omega)$ is *elliptic* on a conic open set $\Gamma \subset \Omega \times \mathbb{R}^d$ if for $(x, \xi) \in \Gamma$,

$$|a(x, \xi)| \sim \langle \xi \rangle^t$$

with an implicit constant uniform in ξ, and locally uniform in x. In this case, we can find a 'parametrix' on *conically compact* subcones of Γ.

Theorem 40.19. *If T is a pseudodifferential operator with a symbol $a \in \dot{S}^t(\Omega)$ which is elliptic on a conic set Γ, then there exists a pseudodifferential operator $S \in \dot{S}^{-t}(\Omega)$ such that $T \circ S$ and $S \circ T$ are regularizing on Γ.*

Proof. If Γ_1 is a conic open subset of Γ, we can find a pseudodifferential operator S with a symbol b by a recursive formula similar to that of the construction of a parametrix of an elliptic symbol, though applying cutoffs outside of Γ_1 so that the symbol vanishes outside of Γ and remains smooth and well defined. $\square$

As a consequence, if a ΨDO T is elliptic on Γ, then for any compactly supported distribution u, it follows that

$$\mathrm{WF}(Tu) \cap \Gamma = \mathrm{WF}(u) \cap \Gamma,$$

i.e. the wavefront set is preserved on Γ.

If T is a properly supported classical pseudodifferential operator of order t with principal symbol $p(x, \xi)$, then the *characteristic set* of T is

$$\mathrm{Char}(T) = \{(x, \xi) : p(x, \xi) = 0\}.$$

The characteristic set is closed and conic, and T is elliptic on the complement of $\mathrm{Char}(T)$. Thus it follows that for $u \in \mathcal{D}(\Omega)^*$, $\mathrm{WF}(u) \subset \mathrm{WF}(Tu) \cup \mathrm{Char}(T)$. Therefore, if $Tu \in \mathcal{E}(\Omega)$, then $\mathrm{WF}(u) \subset \mathrm{Char}(T)$.

40.8 Vector-Valued Pseudodifferential Operators

For certain applications of pseudodifferential operators, e.g. to systems of partial differential equations, it is useful to have a theory of vector-valued pseudodifferentail operators acting on systems of distributions. The theory is almost entirely

analogous to the scalar-valued theory, except that products of pseudodifferential operators may not commute.

We consider only finite dimensional vector-valued quantities here, though the generalization to Banach spaces is not too difficult to imagine. If V is a finite dimensional vector space, and $\Omega \subset \mathbb{R}^d$, we can define a family $\mathcal{D}(\Omega; V)$ of V-valued test functions, and thus obtain a family $\mathcal{D}(\Omega, V)^*$ of V-valued distributions on Ω. These spaces are really just $\mathcal{D}(\Omega) \otimes^\wedge V$ and $\mathcal{D}(\Omega)^* \hat{\otimes} V$. Thus if V has a basis $\{e_1, \ldots, e_n\}$, then elements of $\mathcal{D}(\Omega, V)$ can be uniquely expanded as $\phi_1 e_1 + \cdots + \phi_n e_n$ for test functions $\{\phi_i\}$ in $\mathcal{D}(\Omega)$, and elements of $\mathcal{D}(\Omega, V)^*$ can be uniquely expanded as $u_1 e_1 + \cdots + u_n e_n$ for distributions $\{u_i\}$ in $\mathcal{D}(\Omega)^*$. Similarily, we can define $\mathcal{E}(\Omega, V)$, $\mathcal{E}(\Omega, V)^*$, $H^s(\Omega, V)$, and virtually all of the other function spaces of interest to us in analysis, and these also just turn out to be tensor products.

If V and W are both finite dimensional linear spaces, then we have a natural isomorphism

$$L(X \otimes V, Y \otimes W) \cong L(X, Y) \otimes L(V, W).$$

In particular, after fixing bases $\{e_1, \ldots, e_n\}$ and $\{e_1', \ldots, e_n'\}$ for V and W, an arbitrary operator T in $L(X \otimes V, Y \otimes W)$ can be written uniquely as

$$T\left(\sum_{i=1}^n x_i \otimes e_i\right) = \sum_{i=1}^n \sum_{j=1}^m T_{ij}(x_i) e_j$$

where T_{ij} lie in $L(X, Y)$. Thus every vector-valued continuous linear map is given by an $n \times m$ matrix of scalar-valued continuous linear maps.

It is trivial that we have a Schwartz kernel for such operators, i.e. for any $T : \mathcal{D}(\Omega_1, V) \to \mathcal{D}(\Omega_2, W)^*$, there exists a matrix-valued distribution $K \in \mathcal{D}(\Omega_2 \times \Omega_1, V \otimes W)^*$ such that

$$\langle T\phi, \psi \rangle = \int K(x, y)(\phi(x) \otimes \psi(y)) \, dx \, dy.$$

Let us now specialize to the study of vector valued pseudodifferential operators

We define a pseudodifferential operator T of order t on Ω, *valued in $L(V, W)$*, to be an operator from $\mathcal{D}(\Omega, V)$ to $\mathcal{E}(\Omega, W)^*$ induced by an element of $\Psi_{\mathrm{loc}}^t(\Omega) \hat{\otimes} L(V, W)$. With any such operator we can associate $a : \Omega \times \mathbb{R}^d \to L(V, W)$, a *vector-valued symbol* of order t, such that

$$T\phi(x) = \int a(x, \xi)\phi(y) e^{2\pi i \xi \cdot (x-y)} \, dy.$$

The symbolic calculus remains unpeturbed, except for a few modifications:

- The operators T^t and T^* are $L(W^*, V^*)$-valued.

- If T is an $L(V, W)$ valued proper pseudodifferential operator of order t, and S an $L(W, U)$ valued operator of order s, then the operator $T \circ S$ is an $L(V, U)$ valued pseudodifferential operator of order $t + s$ with an analogous asymptotic expansion to the scalar-valued case, but the non-commutativity implies that the commutator $[T, S] = T \circ S - S \circ T$ is *not necessarily* a pseudodifferential operator of $t + s - 1$.

- If T is a classical $L(V, W)$ valued pseudodifferentail operator with principal symbol $p(x, \xi)$, then the characteristic set of T is

$$\mathrm{Char}(T) = \{(x, \xi) : p(x, \xi) : V \to W \text{ is not injective}\}.$$

Then T is elliptic if $\mathrm{Char}(T) = \varnothing$. One can consider the analogous microlocal variants.

We can also consider pseudodifferential operators valued in linear maps between two vector bundles E and F on a space Ω, i.e. an operator $T : C^\infty(\Omega; E) \to C^\infty(\Omega; F)$ which, when taken in coordinates given by trivializations of E and F, look like matrix valued pseudodifferential operators. We can then associate such an operator T with a symbol, i.e. a family of sections $\mathbb{R}^n \to \Gamma(\Omega, L(E, F))$ with appropriate smoothness and decay. One can then define $\mathrm{Char}(T)$ as above, and the notion of an elliptic operator.

Example. *If M is a smooth manifold, then the exterior derivative $\Omega^n(M) \to \Omega^{n+1}(M)$ is a pseudodifferential operator. Let us focus on the case where M is an open submanifold of $\mathbb{R}^d$ so that it has a natural coordinate system. Since*

$$d(f\,dx^S) = \sum_{j \notin S} \frac{\partial f}{\partial x^j} dx^j \wedge dx^S = \sum_{j \notin S} (2\pi i)(D_x^j f) dx^j \wedge dx^S,$$

it follows that the symbol of the exterior derivative is

$$a(\xi) = \sum_{\substack{S \subset \{1,\dots,d\} \\ \#(S)=n}} \sum_{j \notin S} \sigma(j, S) \cdot 2\pi i \xi_j \cdot E_{S, S \cup \{j\}},$$

where $E_{S,T} : L(\Lambda^n, \Lambda^m)$ is simply the bundle map that maps dx^S to dx^T, and everything else to zero, and where $\sigma(j, S) \in \{-1, 1\}$ is chosen such that $dx^j \wedge$

$dx^S = \sigma(j, S)dx^{S \cup \{j\}}$. *For $n = 0$ in particular, we have*

$$a(\xi) = \sum_{j=1}^{d} 2\pi i \xi_j E_j,$$

thus we see that the exterior differential in this case is always injective, and so $d : \Omega^0(M) \to \Omega^1(M)$ is elliptic. For $n > 0$ on the other hand, is not *elliptic. For instance, in the case $n = 1$, we have*

$$a(\xi) = \sum_{i<j} 2\pi i (\xi_j E_{i,\{i,j\}} - \xi_i E_{j,i \cup \{i,j\}}).$$

But then we notice that for each fixed $\xi^0 \in \mathbb{R}^d$, $a(\xi^0)$ maps $\sum \xi_i^0 dx^i$, and thus $Char(T) = \Omega \times \mathbb{R}_\xi^d$.

Example. *We recall that the principal symbol of a classical pseudodifferential operator on a manifold M is invariantly defined on T^*M. Given a manifold M, we can consider the line bundle $Vol^{1/2}(M)$ of half scalar densities. We claim there is a more rich invariant that can be associated with pseudodifferential operators from $Vol^{1/2}(M)$ to itself, namely, the* subprincipal symbol, *given in coordinates such that if T has symbol $a \sim \sum_{i=0}^{\infty} a_{m-i}$, then*

$$a^{sp} = a_{m-1} + C \cdot \partial_j^x \partial_j^\xi a_m$$

TODO: GET COEFFICIENT C CORRECT. is invariantly defined as a $Hom(Vol^{1/2}(M))$ valued symbol of order $m - 1$ on T^*M. In particular, if a_0 vanishes up to second order, the a_{m-1} is invariantly defined on T^*M.

40.9   The Index Theorem

Let $T \in \Psi_{\text{loc}}^0(\Omega)$. Then T is a continuous operator from $L_c^2(\Omega)$ to $L_{\text{loc}}^2(\Omega)$. We begin this section by determining what conditions ensure that this operator is compact, i.e. what conditions ensure that T maps bounded subsets of $L_c^2(\Omega)$ to bounded subsets of $L_{\text{loc}}^2(\Omega)$. This is equivalent to the induced maps $L^2(K) \to L_{\text{loc}}^2(\Omega)$ being compact for any compact set $K \subset \Omega$, and our job is simplified considerably by the following lemma.

Lemma 40.20. *Let X be a Banach space, and let Y be a Fréchet space. Then a bounded linear operator $T : X \to Y$ is compact if and only if for any sequence $\{x_i\}$ converging weakly to zero in X, $\{Tx_i\}$ converges in the standard topology of Y.*

Proof. TODO: Maybe move to functional analysis notes?

Since Y is a Fréchet space, T is compact if and only if for any bounded sequence $\{x_i\}$ in X, $\{Tx_i\}$ has a convergent subsequence in Y.

If T is compact, and a sequence $\{x_i\}$ converges weakly to zero, then that sequence is bounded, hence $\{Tx_i\}$ is precompact. But $\{Tx_i\}$ converges weakly to zero, since T is continuous from the weak topology on X to the weak topology on Y. But this means $\{Tx_i\}$ converges in norm to zero, since any subsequence has a further subsequence converging in norm to zero.

Conversely, suppose T maps a sequence converging weakly to zero to a sequence converging in norm. If $\{x_i\}$ is a bounded sequence in X, then by Banach-Alaoglu, there exists a subsequence of $\{x_i\}$ which is Cauchy in the weak topology, which without loss of generality we will assume to be $\{x_i\}$ itself. To show this implies $\{Tx_i\}$ converges in Y, we note that if d_Y is a translation invariant metric defining Y, and $\{Tx_i\}$ did not converge, then there would be $\varepsilon > 0$ and a strictly increasing pair of sequences $\{i_j\}$ and $\{i'_j\}$ such that for all j, $d_Y(Tx_{i_j}, Tx_{i'_j}) \geq \varepsilon$. But $x_{i_j} - x_{i'_j}$ converges to zero in the weak topology, which gives a contradiction. $\square$

Thus an operator $T : L^2_c(\Omega) \to L^2_{\text{loc}}(\Omega)$ is compact if and only if, for any compact set $K \subset \Omega$, and any sequence $\{f_i\}$ in $L^2(K)$ converging weakly to zero, $\{Tf_i\}$ converges to zero in $L^2_{\text{loc}}(\Omega)$, which means that for any other compact set $K_1 \subset \Omega$,

$$\lim_{i \to \infty} \int_{K_1} |Tf_i(x)|^2 \, dx = 0.$$

Any operator TODO-

Any smoothing operator $T : L^2_c(\Omega) \to L^2_{\text{loc}}(\Omega)$ is compact, since if $K \in C^\infty(\Omega \times \Omega)$, then it is certainly true that $K \in$

The dual of $L^2_c(\Omega)$ can be identified with $L^2_{\text{loc}}(\Omega)$. Thus a family $\{f_i\}$ converges to zero weakly in $L^2_c(\Omega)$ if and only if

$$\int f_i(x) g(x) \, dx \to 0$$

A family $\{f_i\}$ in $L^2_c(\Omega)$ converges weakly to zero precisely when it converges in $L^2_{\text{loc}}(\Omega)$ to zero.

Thus an operator $T : L^2_c(\Omega) \to L^2_{\text{loc}}(\Omega)$ is compact if and only if

673

40.10 Mehler's Formula

Consider the Hermite operator $L = \partial_x^2 - x^2$. The L is diagonalized by the Hermite operators. Define $H_n = e^{x^2}(-\partial_x)^n\{e^{-x^2}\}$, the Hermite polynomial of degree n. These polynomials have coefficient expansions of the form

$$H_n(x) = n! \sum_{k \leqslant n/2} \frac{(-1)^k}{(n-2k)!k!}(2x)^{n-2k}.$$

If we define $\tilde{H}_n = e^{-x^2}H_n$, then $\{\tilde{H}_n\}$ are an orthogonal basis for $L^2(\mathbb{R})$, with

$$\|\tilde{H}_n\|_{L^2(\mathbb{R})} = 2^{n/2}(n!)^{1/2}\pi^{1/4},$$

i.e. so that the L^2 norm of $\tilde{H}_n$ grows on the order of $n^{n/2}$. We then have

$$L\tilde{H}_n = (2n)\tilde{H}_n.$$

Having diagonalized the operator, to study the heat equation $\partial_t + L = 0$, we can use symbolic calculus to define the heat propogators e^{-tL} with $e^{-tL}\tilde{H}_n = e^{-2tn}\tilde{H}_n$. The kernel can be written (formally at least), as

$$e^{-tL}(x,y) = \sum_{n=0}^{\infty} \frac{e^{-2tn}}{2^{n/2}(n!)^{1/2}\pi^{1/4}}H_n(x)H_n(y)\exp\left(-\frac{x^2+y^2}{2}\right).$$

Using the Taylor's formula for $e^{-(x-z)^2}$, one can prove that

$$\sum_{n=0}^{\infty} \frac{H_n(x)}{n!}z^n = e^{2xz-z^2}$$

then L is diagonalized by the functions $\{H_n e^{-x^2/2}\}$, i.e. $L\{H_n e^{-x^2/2}\} = 2nH_n e^{-x^2/2}$. Mehler's identity specifies that if we define an operator

$$T_t f(x) = \int (1-t^2)^{-1/2}\exp\left(-\frac{t^2(x^2+y^2)-2txy}{1-t^2}\right)f(y)\,dy,$$

then, , we have

$$T_t H_n$$

then $T_t H_n$

674

$$K(x, y, t) = (1 - t^2)^{-1/2} \exp\left(-\frac{t^2(x^2 + y^2) - 2txy}{1 - t^2}\right),$$

then K has an expansion in Hermite polynomials, i.e.

$$K(x, y, t) = \sum_{n=0}^{\infty} \frac{(t/2)^n}{n!} H_n(x) H_n(y).$$

Chapter 41

Fourier Integral Operators

Pseudodifferential Operators formalize a family of operators that can modulate the amplitude of wave packets. The theory of Fourier integral operators extends this theory by introducing operators that can not only modulate the amplitude of wave packets, but also move the position of wave packets in *phase space* (though the uncertainty principle necessitates that we do so in a *symplectic manner*). As a more down to earth description, Fourier integral operators are operators whose kernel is a distribution locally representable by an oscillatory integral, and so we will begin with a general study of such kernels.

41.1   Lagrangian Distributions

Now suppose a_1 and a_2 are two symbols, and $\phi_1 : U \times \mathbb{R}^{p_1}_\theta \to \mathbb{R}$ and $\phi_2 : U \times \mathbb{R}^{p_2}_{\theta'} \to \mathbb{R}$ are two non-degenerate phase functions, such that I_{a_1, ϕ_1} and I_{a_2, ϕ_2} differ by a smooth function. We claim that

$$\mu\text{-supp}(a_1) \cap \Lambda_{\phi_1} = \mu\text{-supp}(a_2) \cap \Lambda_{\phi_2}.$$

Indeed, suppose that $(x_0, \xi_0) \in \Lambda_{\phi_1}$. Then there is $(x_0, \theta_0) \in \Sigma_{\phi_1}$, such that $\nabla_x \phi(x_0, \theta_0) = \xi_0$. Consider ψ as above, satisfying the assumptions of the asymptotic theorem above. Then for any compactly supported, smooth function f equal to one in a nieghborhood of x_0, the $\lambda^{-\frac{p_1-d}{2}}$ term in the asymptotic expansion of

$$\int I_{a_1, \phi_1}(x) f(x) e^{-2\pi i \lambda \psi(x, \sigma)} \, dx$$

676

is equal to $a_1(x_\sigma, \lambda\theta_\sigma)$. Since (x_0, θ_0) lies in the essential support of a_1, we can find some θ_{σ_1} such that it is *not* true that $|a_1(x_{\sigma_1}, \lambda\theta_{\sigma_1})| \lesssim_N \lambda^{-N}$ for all $N > 0$. It then follows that it is not true that for all $N > 0$,

$$\left| \int I_{a_2,\phi_2}(x) f(x) e^{2\pi i\lambda\psi(x,\sigma_1)} \right| \lesssim_N \lambda^{-N}.$$

But this means that

$$(x_0, \xi_0) \in \mathrm{WF}(I_{a_2,\phi_2}) \subset \mu\text{-supp}(a_2) \cap \Lambda_{\phi_2}.$$

The converse follows by symmetry.

It turns out that the reverse is true, provided one is willing to change the symbol one is working with. This is precisely the *equivalence of phase theorem*. To prove this, we shall come up with a definition of the family of oscillatory integral distributions $I_{a,\phi}$ such that Λ_ϕ is an open submanifold of a given Lagrangian submanifold Λ of T^*U, that is independent of the choice of phase function used.

Let M be a manifold, and let $\Lambda \subset T^*M$ be a Lagrangian submanifold. We then say a distribution u is a *Lagrangian distribution* of order m associated to Λ, if for all properly supported pseudodifferential operators $P_1, \ldots, P_N \in \Psi^1_{\mathrm{cl}}(M)$, with principal symbols vanishing on Λ,

$$(P_1 \circ \cdots \circ P_N)u \in W^{2,\infty}_{-m-d/4,\mathrm{loc}}(M).$$

The class of all Lagrangian distributions of order m is denoted by $I^m(M, \Lambda)$. Since there is no occurence of N on the right hand side of the equation, we conclude by a Sobolev embedding argument that $\mathrm{WF}(u) \subset \Lambda$ for any $u \in I^m(M, \Lambda)$. The definition is easily checked to be coordinate independent.

Lemma 41.1. *Suppose ϕ is a non-degenerate phase function defined on a conic open subset of T^*M, with $\Lambda_\phi \subset \Lambda$. If $a \in S^\mu(M \times \mathbb{R}^p)$ is a symbol of order μ supported on this conic neighborhood, and if $m = \mu - d/4 + p/2$, then $I_{a,\phi} \in I^m(M, \Lambda)$.*

Proof. To perform this calculation, we may assume without loss of generality that $\mathrm{supp}_x(a)$ is compact, and that $\mathrm{supp}_\theta(a)$ is supported in a small conic neighborhood Γ of $\mathbb{R}^p - \{0\}$. Then we may choose a coordinate system containing $\mathrm{supp}_x(a)$, and a smooth, homogeneous function $H(\xi)$ in this coordinate system such that

$$\Lambda \cap \mu\text{-supp}(a) \subset \{(\nabla_\xi H(\xi), \xi) : \xi \in \mathbb{R}^d\}.$$

677

Let us calculate

$$J(\xi) = e^{2\pi i H(\xi)} \widehat{I_{a,\phi}}(\xi) = \int a(x,\theta) e^{2\pi i [H(\xi) + \phi(x,\theta) - x \cdot \xi]} \, dx \, d\theta.$$

If we write $\lambda = |\xi|$, and set $\omega = \xi/|\xi|$, then we can also write this quantity as

$$\lambda^p \int a(x, \lambda\theta) e^{2\pi i \lambda [H(\omega) + \phi(x,\theta) - x \cdot \omega]} \, dx \, d\theta.$$

If there are no stationary points in μ-supp(a), then this quantity is $O_N(|\xi|^{-N})$ for all $N > 0$. On the other hand, if there is a stationary point (x_0, θ_0, ξ_0) in this domain, then

$$\nabla_\theta \phi(x,\theta) = 0 \quad \text{and} \quad \nabla_x \phi(x,\theta) = \xi.$$

For any such triple (x_0, θ_0, ξ_0), we therefore have $(x_0, \xi_0) \in \Lambda_\phi \subset \Lambda$, so so that $x_0 = \nabla_\xi H(\xi_0)$. The point is non-degenerate, and so for each ξ in a small neighborhood of ξ_0, there exists a unique stationary point for the integral above, and this stationary point is precisely given by the function $x = \nabla_\xi H(\xi)$. Moreover, for each such ξ there is a unique θ_ξ such that $(\nabla_\xi H(\xi), \theta_\xi, \xi)$ is a stationary point of the integral. By homogeneity, the family of all such stationary points $\{\theta_\xi\}$ is compact, and thus has magnitudes lying in $[C^{-1}, C]$ for some large $C > 0$. If we choose $\chi \in C_c^\infty(\mathbb{R}^p)$ with $\chi(\theta) = 1$ for $|\theta| \in [C^{-1}, C]$, then the integral

$$\lambda^p \int (1 - \chi(\theta)) a(x, \lambda\theta) e^{2\pi i \lambda [H(\omega) + \phi(x,\theta) - x \cdot \omega]} \, dx \, d\theta$$

has no stationary points on it's support. A simple dyadic analysis for large θ, and for small θ, simply derivative bounds, show this quantity is actually Schwartz in the ξ variable. Thus to understand $J(\xi)$, it suffices to analyze the stationary phase integral

$$\lambda^p \int \chi(\theta) a(x, \lambda\theta) e^{2\pi i \lambda [H(\omega) + \phi(x,\theta) - x \cdot \omega]} \, dx \, d\theta.$$

We can now use the variable-coefficient analysis of stationary phase to bound this integral. Namely, if we set

$$\tilde{a}(x, \theta, \lambda) = \lambda^p \chi(\theta) a(x, \lambda\theta),$$

then

$$\left| \nabla_x^n \nabla_\theta^m \nabla_\lambda^k \tilde{a}(x, \theta, \lambda) \right| \lesssim_{n,m,k} \lambda^{p+\mu}.$$

Thus the variable-coefficient version of stationary phase implies that J is a local symbol of order $\mu + p/2 - d/2 = m - d/4$. If $P = P_1 + P_{\leq 0}$ is a properly supported classical pseudodifferential operator of order one with symbol $b \sim \sum b_{1-k}$, where b_1 vanishes on Λ, then

$$
\begin{aligned}
(PI_{a,\phi})(x) &= \int b(x,\eta)\widehat{I}_{a,\phi}(\eta)e^{2\pi i \eta \cdot x}\, d\eta \\
&= \int \{b(x,\eta)J(\eta)\}e^{2\pi i[\eta\cdot x - H(\eta)]}\, d\eta.
\end{aligned}
$$

In particular,

$$
e^{2\pi i H(\xi)}\widehat{PI_{a,\phi}}(\xi) = \int \{P(x,\eta)J(\eta)\}e^{2\pi i[(\eta-\xi)\cdot x - [H(\eta)-H(\xi)]]}\, d\eta\, dx.
$$

Now since b_1 vanishes on Λ, which is where the integral above is stationary, this quantity differs from

$$
\int \{b_{\leq 0}(x,\eta)J(\eta)\}e^{2\pi i[\eta\cdot x - H(\eta)]}\, d\eta
$$

by a Schwartz function. But now a dyadic decomposition and stationary phase implies that this quantity is a symbol of order $m - d/4$ in the ξ variable. Applying induction, it follows that for any $P_1, \ldots, P_N$ of the form above, the function

$$
(P_1 \circ \cdots \circ P_N)I_{a,\phi}
$$

has the property that it's Fourier transform is equal to a symbol c of order $m - d/4$, times $e^{-2\pi i H(\xi)}$. But then for all $j > 0$,

$$
\left(\int_{|\xi|\sim 2^j} |c(\xi)|^2\, d\xi \right)^{1/2} \lesssim 2^{j(m+d/4)},
$$

which is sufficient to conclude that $(P_1 \circ \cdots \circ P_N)I_{a,\phi} \in I^m(M,\Lambda)$. □

The *equivalence of phase theorem* now follows from the fact that *any* Lagrangian distribution is given by an oscillatory integral of the form above.

Lemma 41.2. *Suppose that $\Lambda \subset T^*\mathbb{R}^d$ is a Lagrangian distribution of the form*

$$
\Lambda = \{(\nabla H(\xi),\xi) : \xi \in \mathbb{R}^d\}
$$

for some homogeneous, smooth function H. Then if $u \in H_c^m(\mathbb{R}^d, \Lambda)$, then $e^{2\pi i H(\xi)}\widehat{u}(\xi)$ is a symbol of order $m - d/4$.

Proof. TODO: See Sogge, "Fourier Integrals in Classical Analysis", Proposition 6.1.1. □

Theorem 41.3. *Let $\Lambda \subset T^*M$ be a Lagrangian submanifold, and fix $(x_0, \xi_0) \in \Lambda$. Let ϕ be a nondegenerate phase function, such that Λ_ϕ agrees with Λ on a small neighborhood of (x_0, ξ_0). Then for any $u \in I^m(M, \Lambda)$, with $WF(u)$ contained in a sufficiently small neighborhood of (x_0, ξ_0), can be written, modulo an element of $C^\infty(M)$, can be written as $I_{a,\phi}$ for some symbol a of order $m + d/4 - p/2$.*

Proof. TODO: See Sogge, "Fourier Integrals in Classical Analysis", Proposition 6.1.4. □

Let us conclude this section by using the equivalence of phase to discuss the *reduction* of θ variables. Given any Lagrangian submanifold Λ, and any $u \in I^m(M, \Lambda)$, it is of interest to determine an oscillatory integral representation of u with as few phase variables as possible. To do this, consider any representation given locally in coordinates in terms of a non-degenerate phase function $\phi : U \times \mathbb{R}^p_\theta \to \mathbb{R}$. Consider the projection maps $\Pi_\Lambda : \Lambda_\phi \to U$ and $\Pi_\Sigma : \Sigma_\phi \to U$, then the non-degeneracy of ϕ and the fact that Σ_ϕ is d dimensional implies that

$$\dim(\mathrm{Ker}(\Pi_\Lambda)_*) = d - \mathrm{Rank}(\Pi_\Lambda)_*.$$

However, if a tangent vector v is in the kernel of $(\Pi_\Lambda)_*$, we can write it as $v = \sum v_i d\theta^i$, where $(H_\theta \phi)v = 0$. But the dimension of this space of tangent vectors is precisely p minus the rank of $H_\theta \phi$. Thus we conclude that

$$p - \mathrm{Rank}(H_\theta \phi) = d - \mathrm{Rank}(\Pi_\Lambda)_*.$$

This inequality can be used to reduce θ variables, if the rank of $(\Pi_\Lambda)_*$ is large. Let $r = \mathrm{Rank}(H_\theta \phi)$. TODO: See Sogge, Remark following Proposition 6.1.5. Then we can consider a phase function parameterizing Λ locally with only $p - r$ variables. In particular, if the rank of $(\Pi_\Lambda)_*$ is constant, and equal to r everywhere, then $X = \Pi_\Lambda(\Lambda)$ is an r dimensional submanifold of M. We then say that $I^m(\Lambda, M)$ is the space of *conormal distributions* to X of order m, since one can see that Λ is then the conormal bundle of X, given the assumption that Λ is Lagrangian.

Example. *We will later see in the theory of Fourier integral operators that on a compact manifold M equipped with a self-adjoint, positive-semidefinite elliptic pseudodifferential operator P of order one, we can approximate the half-wave*

propogators e^{itP} for small t, modulo C^∞ functions, by an operator with kernel K_t in $I^0(M \times M, \Lambda)$, where

$$\Lambda = \{(x, y; \xi, \eta) : (y, -\eta) = \Phi_t(x, \xi)\},$$

and where $\{\Phi_t\}$ is the phase flow for the Hamiltonian vector field associated with the principal symbol p of P. Under the assumption that the cospheres

$$\{\xi : p(x, \xi) = 1\}$$

have non-vanishing Gaussian curvature, $H_\xi \phi$ has rank $d - 1$. But this means that we should be able to locally find a one-dimensional non-degnerate phase function ψ, such that Λ_ψ agrees with Λ locally, and so there exists a symbol $b(t, x, y, s)$ of order $\mu = (d - 1)/2$ in the s variable, such that $K_t(x, y)$ is a finite sum of oscillatory integrals of the form

$$K_t(x, y) = \int b(t, x, y, s) e^{2\pi i \psi(t, x, y, s)} \, ds.$$

In the special case where M is Riemannian, and $P = \sqrt{-\Delta}$, one can use the fact that Φ_t is the geodesic flow to show $\psi(t, x, y, s) = s \cdot a(t, x, y) \cdot (|t| - d_g(x, y))$ for some non-vanishing function a.

41.2   Local Symplectic Geometry

Before we get into the general theory of Fourier integral operators, let's recall some results in the theory of symplectic geometry. We recall that a *symplectic vector space V* is a finite dimensional vector space equipped with a non-degenerate, skew-symmetric bilinear form ω. A fundamental example is that for any vector space W, $V = W^* \oplus W$ is naturally a symplectic vector space with the symplectic form $\omega((x_1^*, x_1), (x_2^*, x_2)) = x_2^*(x_1) - x_1^*(x_2)$. Spectral theory can be used to show that for any symplectic vector space V, V is even dimensional, and that we can find a pair of independent sets $\{e_i\}$ and $\{f^i\}$, forming a basis of V, such that $\omega(e_i, e_j) = \omega(f^i, f^j) = 0$, and $\omega(e_i, f^i) = 1$. Such a basis is called a *Darboux basis*, and we conclude from it's existence that all symplectic vector spaces of the same dimension are isomorphic.

   A *Lagrangian subspace* of a symplectic vector space V is a subspace W of V such that $W^\perp = W$, where

$$W^\perp = \{x \in V : \omega(x, x') = 0 \text{ for all } x' \in W\}.$$

A subspace W is Lagrangian if and only if $\dim(W) = \dim(V)/2$, and for any $w_1, w_2 \in W$, $\omega(w_1, w_2) = 0$. As an example, if we write $V = V_e \oplus V_f$, where V_e and V_f are the spans of the separate parts of a Darboux basis, then V_e and V_f are Lagrangian subspaces. If $A : W \to W^*$ is a linear map, then the graph $\Gamma_A = \{(x, Ax) : x \in W\}$ is a Lagrangian submanifold of $W \oplus W^*$ if and only if A is self-adjoint, where we identify W^{**} with W, i.e. if $\langle Ax_1, x_2 \rangle = \langle x_1, Ax_2 \rangle$ for all $x_1, x_2 \in W$.

Lemma 41.4. *Let V be a symplectic vector space, and let*

$$\{e_1, \ldots, e_{n_1}\} \cup \{f^1, \ldots, f^{m_2}\}$$

be linearly independent vectors such that $\omega(e_i, e_j) = \omega(f^i, f^j) = 0$, and $\omega(e_i, f^j) = \delta(i, j)$. Then we can extend these sets to a full Darboux basis for V.

Proof. Suppose first that $n_1 < n_2$. Since ω is non-degenerate, we can find a vector e such that $\omega(e, e_i) = 0$ for $1 \leqslant i \leqslant n_1$, and $\omega(e, f^i) = \delta(i, n_1 + 1)$. It follows that e is linearly independent from the previously selected elements of the basis and we can set $e_{n_1+1} = e$. Thus we can increase n_1 by one. A similar argument works for choosing f_{n_2+1} if $n_2 < n_1$. If $n_1 = n_2$, and we don't yet have a basis, then we choose $e = e_{n_1+1}$ linearly independent to the previous set such that $\omega(e, e_i) = \omega(e, f^i) = 0$. Iterating this selection procedure yields the required Darboux basis. $\square$

Lemma 41.5. *If V_0 and V_1 are Lagrangian subspaces of a symplectic vector space V, then we can find a third Lagrangian subspace V_2 which is transverse to both V_0 and V_1.*

Proof. Let V have dimension $2n$. It suffices to find a Darboux basis in which

$$V_0 = \mathrm{span}(e_1, \ldots, e_n) \quad \text{and} \quad V_1 = \mathrm{span}(e_1, \ldots, e_k, f^{k+1}, \ldots, f^n),$$

since we can then set V_2 to be the space spanned by $\{e_i + f^i : k + 1 \leqslant i \leqslant n\} \cup \{f^1, \ldots, f^k\}$. To do this, we pick $\{e_1, \ldots, e_n\}$ such that $\{e_1, \ldots, e_k\}$ spans $V_0 \cap V_1$, and $\{e_1, \ldots, e_n\}$ spans V_1. Next, we note that the map $T : V_0 \to \mathbb{R}^{n-k}$ given by

$$Tv = (\omega(e_{k+1}, v), \ldots, \omega(e_n, v))$$

has kernel equal to $V_0 \cap V_1$. This is because $V_0^\perp = V_0$, so if $v \in V_1$, then we automatically obtain that $\omega(e_1, v) = \cdots = \omega(e_k, v) = 0$, so that if $Tv = 0$, then $v \in V_0^\perp = V_0$. But this means that T is surjective, so we can find $\{f^{k+1}, \ldots, f^n\} \subset$

V_1 such that $\omega(e_i, f_j) = \delta(i, j)$. It follows that $\{e_1, \ldots, e_k\} \cup \{f^{k+1}, \ldots, f^n\}$ are linearly independent, and thus form a basis for V_1. And we can now use the previous Lemma to extend these vectors to a Darboux basis. $\square$

If V_X and V_Y are symplectic vector spaces, then $V_X \oplus V_Y$ can be made into a symplectic vector space, if we equip it either with with the symplectic form $\omega_X - \omega_Y$, or with $\omega_Y - \omega_X$. In either case, we call a Lagrangian subspace C of $V_X \oplus V_Y$ a *linear canonical relation*. We recall that a *symplectic linear map* $A : V_1 \to V_2$ between symplectic vector spaces is a map preserving the symplectic form.

Lemma 41.6. *Let $C \subset V_X \oplus V_Y$ be a linear canonical relation. Then we can find orthogonal decompositions $V_X = V_{X,1} \oplus V_{X,2}$ and $V_Y = V_{Y,1} \oplus V_{Y,2}$ such that*

$$C = C_X \oplus \Gamma \oplus C_Y,$$

where C_X is a Lagrangian submanifold of $V_{X,1}$, C_Y is a Lagrangian submanifold of $V_{Y,1}$, and Γ is the graph of a symplectic isomorphism $A : V_{X,2} \to V_{Y,2}$.

Proof. Let

$$C_X = \{x \in V_X : (x, 0) \in C\} \quad \text{and} \quad C_Y = \{y \in V_Y : (0, y) \in C\}.$$

Because C is Lagrangian, C is contained in $C_X^{\perp} \oplus C_Y^{\perp}$. In other words, if $x_1 \in C_X$, $y_1 \in C_Y$, and $(x_2, y_2) \in C$, then

$$\omega(x_1, x_2) = \omega(y_1, y_2) = 0.$$

In particular, $C_X \subset C_X^{\perp}$ and $C_Y \subset C_Y^{\perp}$. Find orthogonal $V_{X,2}$ and $V_{Y,2}$ such that

$$C_X^{\perp} = C_X \oplus V_{X,2} \quad \text{and} \quad C_Y^{\perp} = C_Y \oplus V_{Y,2}.$$

Then $C \subset C_X \oplus C_Y \oplus (V_{X,2} \oplus V_{Y,2})$. We can thus find $\Gamma \subset V_{X,2} \oplus V_{Y,2}$ such that $C = C_X \oplus C_Y \oplus \Gamma$. We claim Γ projects bijectively onto $V_{X,2}$ and $V_{Y,2}$. For instance, suppose $x \in V_{X,2}$ and $y_1, y_2 \in V_{Y,2}$ are such that (x, y_1) and (x, y_2) lie in Γ. Then $(0, y_1 - y_2)$ lies on Γ and on C_Y, so by orthogonality, $y_1 = y_2$. Thus Γ is the graph of a symplectic isomorphism $A : V_{X,2} \to V_{Y,2}$. To define $V_{X,1}$ and $V_{Y,1}$, let

$$\dim(V_X) = 2n \quad \dim(V_Y) = 2m \quad \dim(C_X) = a \quad \dim(C_Y) = b.$$

We can then consider Darboux bases

$$\{e_{X,1}, \ldots, e_{X,n}\} \cup \{f_{X,1}, \ldots, f_{X,m}\}$$

683

and
$$\{e_{Y,1}, \ldots, e_{Y,m}\} \cup \{f_{Y,1}, \ldots, f_{Y,m}\}$$

for V_X and V_Y, such that
$$C_X = \mathrm{span}(\{e_{X,1}, \ldots, e_{X,a}\})$$

and
$$C_Y = \mathrm{span}(\{e_{Y,1}, \ldots, e_{Y,b}\}).$$

But then we immediately see that
$$V_{X,2} = \mathrm{span}(\{e_{X,a+1}, \ldots, e_{X,n}\} \cup \{f_{X,a+1}, \ldots, f_{X,n}\})$$

and
$$V_{Y,2} = \mathrm{span}(\{e_{Y,b+1}, \ldots, e_{Y,m}\} \cup \{f_{Y,a+1}, \ldots, f_{Y,m}\}).$$

We then simply define
$$V_{X,1} = \mathrm{span}(\{f_{X,1}, \ldots, f_{X,a}\})$$

and
$$V_{Y,1} = \mathrm{span}(\{f_{Y,1}, \ldots, f_{Y,b}\}),$$

and the remaining parts of the proof follow immediately. $\qquad \square$

A *symplectic manifold M* is a manifold equipped with a symplectic two form ω, i.e. a two form which gives each of the tangent spaces of M a symplectic structure. The basic example here is $M = T^*X$, where X is any smooth manifold; the natural two form here is

$$\omega = dx \wedge d\xi = d\theta,$$

where $\theta = \sum \xi_i dx^i$ is the *tautological one form* on T^*M. It has the property that for any section $s : M \to T^*M$, we have $s^*\theta = s$.

Like with the Riemannian form on a Riemannian manifold, the symplectic form on a symplectic manifold gives us a natural bundle isomorphism $J : T^*M \to TM$. In particular, given a function $f : M \to \mathbb{R}$, we can define the *symplectic gradient* $\nabla_\Xi f$ as the vector field which is identified under the bundle isomorphism with the covector field df. As an example, if $M = T^*X$, then in local coordinates (x, ξ) on T^*X, the identification is given by

$$J(d\xi^i) = \frac{\partial}{\partial x^i} \quad \text{and} \quad J\left(dx^i\right) = -\frac{\partial}{\partial \xi^i}.$$

684

Thus for a function $f : T^*X \to \mathbb{R}$, we have

$$\nabla_\Xi f = \sum \frac{\partial f}{\partial \xi^i} \frac{\partial}{\partial x^i} - \frac{\partial f}{\partial x^i} \frac{\partial}{\partial \xi^i}.$$

This gradient is closely related to the theory of Hamiltonian vector fields, i.e. since, given a Hamiltonian H on some phase space representing a physical system, $\nabla_\Xi H$ gives the Hamiltonian flow of that physical system. We can define the *Poisson bracket* of two functions $f, g : M \to \mathbb{R}$ by setting $\{f, g\} = \nabla_\Xi f(g) = dg(\nabla_\xi f)$. For $M = T^*X$, we can write this in coordinates as

$$\{f, g\} = \sum \frac{\partial f}{\partial \xi_i} \frac{\partial g}{\partial x_i} - \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial \xi_i},$$

which agrees with the classical Poisson bracket.

A (immersed) *Lagrangian submanifold* of a symplectic manifold M is a submanifold N such that it's tangent space is a Lagrangian subspace of the tangent space of M at each point. To check that an immersion $i : X \to M$ gives a Lagrangian submanifold, it suffices to show that $i^*\omega = 0$, and that $\dim(X) = \dim(M)/2$. We have already seen these kinds of subspaces in our discussion of oscillatory integral distributions, since the wavefront sets of these distributions form Lagrangian submanifolds of the cotangent space of the space the distributions live on. Another example includes particular kinds of sections $X \to T^*X$.

Lemma 41.7. *The image of a section $s : X \to T^*X$ is an immersed Lagrangian submanifold if and only if locally we can write*

$$s = df$$

for some function $f : X \to \mathbb{R}$.

Proof. If $\theta = \sum \xi^i dx^i$, then $d\theta$ is the symplectic form ω, and $s^*\theta = s$. Thus s gives a Lagrangian manifold if and only if $s^*\omega = s^*(d\theta) = d(s^*\theta) = ds = 0$. The result above now follows by Poincaré's Lemma. $\square$

A similar result holds if Λ is a conic Lagrangian submanifold of T^*X.

Lemma 41.8. *Suppose that $\Lambda \subset T^*X - \{0_X\}$ is a conic Lagrangian manifold containing some covector $(x_0, v) \in T^*X$. Then local coordinates (x, U) can be chosen, centered at x_0, inducing coordinates (x, ξ) on T^*U, such that the map $(x, \xi) \mapsto \xi$ is a diffeomorphism in an open neighborhood of (x_0, ξ_0) in Λ, and there exists a smooth, homogeneous function H such that, in a small neighborhood of (x_0, ξ_0) in T^*X, Λ is equal to the set of all pair $(\nabla H(\xi), \xi)$.*

685

Proof. We first choose coordinates (x, U) such that $(x, \xi) \mapsto \xi$ is a diffeomorphism. Begin by choosing coordinates (y, V) centered at x_0 such that $v = dy_1$. The tangent plane V_0 to Λ at v must be Lagrangian. If V_0 is transverse to the Lagrangian plane given in (y, η) coordinates by

$$V_1 = \{(0, a) : a \in \mathbb{R}^d\},$$

then we can set x to be y. Otherwise, we find V_2 Lagrangian and transverse to both V_0 and V_1. Since V_2 is transverse to V_1, it can be identified with a linear section, and the last result thus implies that we can find a quadratic form $Q : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$V_2 = \{(a, dQ(a)) : a \in \mathbb{R}^d\}.$$

If we set $x_1 = y_1 + Q(y)$, and $x_i = y_i$ for $2 \leqslant i \leqslant n$, then these are the required coordinates.

We claim now that if (x, U) gives a diffeomorphism, then we can find H. Shrinking U if necessary, there exists a radial $\phi : \mathbb{R}^d \to U$ such that

$$\Lambda \cap U = \{(\phi(\xi), \xi) : \xi \in \mathbb{R}^d\}.$$

Since Λ is Lagrangian, if $\psi(\xi) = (\phi(\xi), \xi)$, then

$$\psi^* \theta = \sum \xi_i d\phi_i = 0.$$

If we set $H(\xi) = \sum \xi_i \phi_i(\xi)$, then $\nabla H = \phi$, giving the required result. $\square$

The structure of linear canonical relations can give us results about 'nonlinear' canonical relations, i.e. a conic Lagrangian submanifold of $T^*X \times T^*Y$ for two manifolds X and Y.

Lemma 41.9. *Let X and Y be smooth manifolds, and let C be a conic Lagrangian submanifold of $T^*X \times T^*Y$. Fix $(x_0, y_0; \xi_0, \eta_0) \in C$, and assume that the vector*

$$\xi_0 \frac{\partial}{\partial \xi} + \eta_0 \frac{\partial}{\partial \eta}$$

is not tangent to C. Then X and Y have coordinate systems $x = (x', x'')$ and $y = (y', y'')$ centered at x_0 and y_0, such that $\xi_0 = (1, \ldots, 0)$, $\eta_0 = (1, \ldots, 0)$, and the tangent plane to C is given by

$$dx' = dy' \quad and \quad d\xi' = d\eta' \quad and \quad d\xi'' = 0 \quad and \quad d\eta'' = 0.$$

Then (x', x'', η', y'') can be used as local coordinates for C, and we can find a phase function $\phi(x', x'', y'', \eta')$ such that C is parameterized by ϕ, in the sense that C locally agrees with Λ_ϕ on the coordinate system.

41.3 Local Theory

Let us recall the theory of oscillatory integral distributions. Let $U \subset \mathbb{R}^n$ be open. Consider a real-valued phase function $\phi \in C^\infty(U_x \times \mathbb{R}^p_\theta)$, homogeneous in the θ variable, and such that $\nabla_{x,\theta}\phi$ is non-vanishing on the support of a symbol $a \in S^m(U \times \mathbb{R}^p)$. Then we can define a distribution u, formally speaking, by the equation

$$u(x) = \int_{\mathbb{R}^p} a(x,\theta)e^{2\pi i \phi(x,\theta)} \, d\theta.$$

If we consider the conic set

$$\Sigma_\theta = \{(x,\theta) \in U \times \mathbb{R}^p : \nabla_\theta \phi(x,\theta) = 0\},$$

then $\mathrm{WF}(u) \subset \Sigma_\theta$. Let us now assume the phase is *nondegenerate*, in the sense that whenever $\nabla_\theta \phi = 0$, $D_{x,\theta}(\nabla_\theta \phi)$ is an invertible matrix. Then Σ_θ will be an n dimensional manifold in $U \times \mathbb{R}^p$, and the map $\Sigma_\theta \to T^*U$ given by $(x,\theta) \mapsto \nabla_x \phi(x,\theta)$ will be an immersion, which we will denote by Λ_θ. This immersed manifold will be a *Lagrangian submanifold* of T^*U, in the sense that the Tangent spaces W at each point of Λ_θ will satisfy $W^\perp = W$ with respect to the symplectic form $d\xi \wedge dx$.

A *Fourier integral operator* is a continuous operator $T : C_c^\infty(Y) \to \mathcal{D}^*(X)$ whose Schwartz kernel K is given, modulo a smoothing kernel, by a sum of oscillatory integral distributions of the form

$$K(x,y) = \int_{\mathbb{R}^p} a(x,y,\theta)e^{2\pi i \phi(x,y,\theta)} \, d\theta,$$

for some symbol $a \in S^\mu(X \times Y \times \mathbb{R}^N)$. It's *canonical relation* is

$$C_\phi = \{(x,y;\nabla_x\phi(x,y,\theta), -\nabla_y\phi(x,y,\theta)) : \nabla_\theta\phi(x,y,\theta) = 0\},$$

which roughly speaking, tells us how T moves wave packets around in phase space. The *order* of the operator is $\mu - n/4 + p/2$; this strange choice might be better understood if $\phi(x,y,\xi) = (x-y) \cdot \xi$, in which case the Fourier integral operator is actually a pseudodifferential operator, and then the two orders agree with one another. TODO: More intuitive explanation, maybe using stationary phase?

687

Example. *The solution to the half-wave equation $\partial_t f = \sqrt{-\Delta} f$ on $\mathbb{R}^d$ is given by the equation*

$$e^{it\sqrt{-\Delta}} f = \int e^{2\pi i(\xi \cdot x + t|\xi|)} \widehat{f}(\xi) = \int e^{2\pi i(\xi \cdot (x-y) + t|\xi|)} f(y) \, d\xi \, dy,$$

which is a Fourier integral operator which phase $\phi_t(x, y, \xi) = \xi \cdot (x - y) + t|\xi|$ and symbol $a_t(x, y, \xi) = 1$. Given two input functions f_0 and f_1, if we write

$$f_+ = \frac{f_0 - i\sqrt{-\Delta}^{-1} f_1}{2} \quad and \quad f_- = \frac{f_0 + i\sqrt{-\Delta}^{-1} f_1}{2},$$

then the function

$$u(x, t) = e^{it\sqrt{-\Delta}} f_+ + e^{-it\sqrt{-\Delta}} f_-$$

solves the wave equation $\partial_t^2 u = \Delta u$, with initial conditions

$$u(x, 0) = f_+ + f_- = f_0$$

and

$$\partial_t u(x, 0) = i\sqrt{-\Delta} f_+ - i\sqrt{-\Delta} f_- = f_1.$$

Expanding out the definition of u, we obtain a sum of operators applied to f_0 an f_1, of the form

$$T f(x, t) \mapsto \int e^{2\pi i(\xi \cdot x \pm t|\xi|)} |\xi|^{-j} \widehat{f}(\xi).$$

Each of these operators is a Fourier integral operator with canonical relation

$$C_\phi = \{y \mp t\xi/|\xi|, y; \xi, \xi\}.$$

Similarily, one may define wave propogators on a compact Riemannian manifold M. Locally in coordinates, these propogators can be expressed as Fourier integral operators, with a very similar canonical relation, i.e.

$$C_\phi = \{\exp_y(t\xi/|\xi|), y; \xi, \xi\},$$

so that singularities travel along geodesics.

Example. *Let A_t be the spherical averaging operator, i.e. $A_t f(x)$ is the average of f on a sphere of radius t centered at x. We can write $A_t = f * \sigma_t$, where σ_t is the surface measure of the sphere of radius t. Recall that stationary phase tells us that*

$$\widehat{\sigma}(\xi) = \sum_\pm e^{\pm 2\pi i|\xi|} a_\pm(\xi)$$

688

for symbols $a_\pm$ of order $-(n-1)/2$. Thus we can write

$$A_t f(x) = \sum_\pm \int e^{2\pi i(\xi \cdot x \pm t|\xi|)} a_\pm(t\xi) \hat{f}(\xi),$$

which relates the study of averaging operators to the theory of Fourier integral operators. Alternatively, if $\Phi(x, y, t) = |x - y|^2/t^2 - 1$, then

$$A_t f(x) = \int \delta(\Phi(x, y, t)) f(y) \, dy$$

$$= \int e^{2\pi i \lambda \Phi(x, y, t)} f(y) \, d\lambda \, dy.$$

Thus we can express A_t in many different ways as Fourier integral operators.

The last example indicates an important problem with the theory of Fourier integral operators; there is a lack of uniqueness in the choice of phase defining the operator. Nonetheless, we have seen that the *wavefront set* is essentially uniquely verified, because it is connected to the relation between $\mathrm{WF}(Tu)$ and $\mathrm{WF}(u)$, which do not need a phase representation to be defined. The *equivalence of phase theorem* says that the canonical relation is essentially the only invariant of the representation of the operator as a Fourier integral; if two phases have the same canonical relations, then Fourier integral operators defined in terms of one phase can be converted into an operator defined in terms of the other operator, and one has an asymptotic formula relating the two symbols of the associated representations.

TODO: Merge these notes with later notes.

For some purposes, it has become convenient to determine the class of Lagrangian distributions in a representation independent way. A *Lagrangian distribution* of order m on a manifold M, associated with a Lagrangian submanifold Λ of T^*M, is a distribution u on M, such that for any $N > 0$, and any properly supported pseudodifferential operators $P_1, \ldots, P_N$ of order one, with principal symbols vanishing on Λ, we have

$$(P_1 \circ \cdots \circ P_N)u \in H^\infty_{-m-n/4,\mathrm{loc}}(M).$$

The set of all Lagrangian distributions of order m is denoted $I^m(M, \Lambda)$. The condition above means, roughly, that one can differentiate u in directions away from Λ while maintaining the regularity of u. In particular, applying some kind of Sobolev embedding gives that if $u \in I^m(M, \Lambda)$, then $\mathrm{WF}(u) \subset \Lambda$. It will be easy

to see what kinds of distributions are Lagrangian once we prove the equivalence of phase theorem, which allows one to reduce the study of Lagrangian distributions to oscillatory integrals like the ones above.

The next lemma is useful for the proof of the equivalence of phase theorem, and is, in fact, a special case.

Lemma 41.10. *Suppose that u is a compactly supported Lagrangian distribution in $I^m(\mathbb{R}^n, \Lambda)$, where*

$$\Lambda = \{(\nabla H(\xi), \xi) : \xi \in \mathbb{R}^n\}$$

for some smooth, homogeneous function H. Then for $|\xi| \geq 1$, there exists a symbol $v \in S^{m-n/4}(\mathbb{R}^n)$ such that

$$\widehat{u}(\xi) = e^{-2\pi i H(\xi)} v(\xi),$$

i.e. so that formally speaking, u is given by the oscillatory integral distribution.

$$u(x) = \int v(\xi) e^{2\pi i(\xi \cdot x - H(\xi))} \, d\xi$$

Proof. TODO: 6.1.1 of Sogge. $\qquad \square$

Theorem 41.11. *Let ϕ be a non-degenerate phase function defined on an open conic neighborhood of a point $(x_0, \xi_0) \in T^* \mathbb{R}^n$, and consider the conic Lagrangian manifold Λ_ϕ. If $a \in S^\mu(\mathbb{R}^n \times \mathbb{R}^p)$ is supported on a sufficiently small conic neighborhood of (x_0, ξ_0), then the oscillatory integral distribution*

$$u(x) = \int a(x, \theta) e^{2\pi i \phi(x, \theta)} \, d\theta$$

lies in $I^m(\mathbb{R}^n, \Lambda)$, where $m = \mu - n/4 + p/2$. Moreover, there exists $v \in S^{m-n/4-1}$ such that

$$\widehat{u}(\xi) = (2\pi)^{n/2-p/2} e^{-2\pi i H(\xi)} \left(a(x, \theta) |Det(H\phi)|^{-1/2} e^{(i\pi/4)sgn(H\phi)} + v(\xi) \right),$$

where (x, θ) is the unique solution to the differential equation $\nabla_\theta \phi(x, \theta) = 0$ and $\nabla_x(x, \theta) = \xi$. Conversely, every Lagrangian distribution $u \in I^m(\mathbb{R}^n, \Lambda)$ with $WF(u)$ contained in a small enough neighborhood of (x_0, ξ_0) can be written in this form.

Example. *Let $\phi(x, y, \xi)$ be smooth away from $\xi = 0$, homogeneous of degree one, and satisfy*

$$\partial_\xi^\beta \{\phi(x, y, \xi) - (x - y) \cdot \xi\} \lesssim_\beta |x - y|^2 |\xi|^{1-\beta}.$$

690

For any symbol $a \in S^m(\mathbb{R}^n \times \mathbb{R}^n)$ supported on $|x - y| \lesssim 1$ and $|\xi| \gtrsim 1$ in such a way that $|\nabla_\xi \phi| \gtrsim |x - y|$ and $|\nabla_x \phi| \gtrsim |\xi|$ on the support of a, define

$$K(x, y) = \int a(x, y, \xi) e^{[2\pi i \phi(x, y, \xi)]} \, d\theta$$

We have seen one proof that K is a pseudodifferential operator. Let's give another proof using the equivalence of phase theorem. The distribution K is defined by an oscillatory integral distribution with Σ_ϕ contained in $\{(x, x; \xi) : x \in \mathbb{R}^n, \xi \in \mathbb{R}^n\}$. The fact that $\phi(x, y, \xi) \approx (x - y) \cdot \xi$ means that the resulting Lagrangian manifold is then

$$\Lambda_\phi = \{(x, x; \xi, -\xi)\},$$

which is the same Lagrangian manifold associated pseudodifferential operators. To determine the Fourier transform of K using equivalence of phase, fix (x_0, ξ_0). Then at $p = (x_0, x_0; \xi_0, -\xi_0)$, the tangent space to Λ_ϕ is the span of

$$\{\partial_{x_i} + \partial_{y_i}\} \cup \{\partial_{\xi_i} - \partial_{\eta_i}\}.$$

Let V be the Lagrangian subspace of $T_p(T^*M)$ spanned by

$$\{(a, 0, c, a + c)\}$$

$$\{\partial_{\eta_i}\} \cup \{\partial_{x_i} + \partial_{\xi_i} - \partial_{\eta_i}\}$$

$$e_1, \ldots, e_n = \partial_{x_i} + \partial_{y_i}$$
$$e'_1, \ldots, e'_n = \partial_{x_i}$$
$$f'_i = \partial_{\xi_i} - \partial_{\eta_i}$$
$$f_i = \partial_{\eta_i}$$

$z^1 = \xi_0 dx - \xi_0 dy$
Choose coordinates $\xi_0 dx - \xi_0 dy = dz^1$
Now $u \in I^m(\mathbb{R}^n, \Lambda_\phi)$. The result above says that

$$\hat{K}(\xi, \eta) = (2\pi)^{n/2} e^{-2\pi i H(\xi)}$$

41.4 Fourier Integral Operators

Pseudodifferential operators formalize the family of all operators that modulate the amplitude of wave packets. The theory of Fourier integral operators extends this theory by not only modulating the amplitude of wave packets, but also moving them around in phase space in a *symplectic manner*, i.e. in a way which, roughly speaking, obeys the uncertainty principle. Basic examples of Fourier integral operators include the translation operators

$$\text{Trans}_{x_0} f(x) = f(x - x_0),$$

modulation operators

$$\text{Mod}_{\xi_0} f(x) = e^{2\pi i \xi_0 \cdot x} f(x),$$

the change of variables operator T_A associated with an invertible linear transformation $A : \mathbb{R}^n \to \mathbb{R}^n$, i.e. such that

$$T_A f(x) = |\det(A)|^{-1/2} f(A^{-1} x),$$

and the Fourier transform

$$\mathcal{F} f(\xi) = \hat{f}(\xi).$$

These four operators are all unitary, so might be thought of as preserving the amplitude of wave packets, but they move wave packets around in phase space in different ways:

- The translation operators move wave packets in phase space according to the diffeomorphism
  $$\Phi(x, \xi) = (x + x_0, \xi).$$

- The modulation operators move packets in phase space according to the diffeomorphism
  $$\Phi(x, \xi) = (x, \xi + \xi_0).$$

- The change of variables operator moves packets in space according to the diffeomorphism
  $$\Phi(x, \xi) = (Ax, (A^T)^{-1} \xi).$$

- The Fourier transform move packets according to the diffeomorphism
  $$\Phi(x, \xi) = (\xi, -x).$$

All four of the diffeomorphisms $\Phi : T^* \mathbb{R}^d \to T^* \mathbb{R}^d$ are *symplectomorphisms*, i.e. they preserve the symplectic form $\omega = \sum dx^i \wedge d\xi_i$, i.e. the bilinear form

$$\omega((x_1, \xi_1), (x_2, \xi_2)) = \xi_2(x_1) - \xi_1(x_2),$$

which is the derivative of the *tautological one form* $\theta = \sum \xi_i dx^i$, a fact very important to the tractability of the study of the associated operators, because of the uncertainty principle.

Let us see why this is essential. With any diffeomorphism $\Phi : T^* \mathbb{R}^n \to T^* \mathbb{R}^n$, we can try and construct a unitary operator T_Φ from $L^2(\mathbb{R}^n)$ to itself, which roughly speaking, has the property that it maps a wave packet localized at a point (x_0, ξ_0) to a wave packet localized near $\Phi(x_0, \xi_0)$. If we consider an pseudodifferential operator $S_a = a(x, D)$ with symbol a, then S amplifies wave packets localized at (x_0, ξ_0) by a quantity $a(x_0, \xi_0)$. Thus, morally speaking, we should expect the operator $T_\Phi^{-1} \circ S_a \circ T_\Phi$ to fix the location of wave packets, and amplify wave packets localized at (x_0, ξ_0) by the quantity $a \circ \Phi$, i.e. so that

$$T_\Phi^{-1} \circ S_a \circ T_\Phi \approx (a \circ \Phi)(x, D).$$

We should at least expect this to hold up to first order. If we have another operator S_b, then we should expect that, using the formula above, up to first order we should have

$$\begin{aligned} [(a \circ \Phi)(x, D), (b \circ \Phi)(x, D)] &\approx [T_\Phi^{-1} \circ S_a \circ T_\Phi, T_\Phi^{-1} \circ S_b \circ T_\Phi] \\ &\approx T_\Phi^{-1} \circ [S_a, S_b] \circ T_\Phi \\ &\approx ([a, b] \circ \Phi)(x, D). \end{aligned}$$

Taking principal symbols of either side of the equation yields to the exact equation

$$\{a \circ \Phi, b \circ \Phi\} = \{a, b\} \circ \Phi,$$

where $\{f, g\} = \omega(\nabla f, \nabla g)$ is the Poisson bracket. But this means that the equation above can only be true if $\Phi^* \omega = \omega$, i.e. Φ is a symplectomorphism. In this case, the graph of Φ, namely the set

$$\Lambda_\Phi = \{(x, y; \xi, \eta) : (x, \xi) = \Phi(y, \eta)\} \subset T^*(\mathbb{R}^n \times \mathbb{R}^n)$$

will be a Lagrangian submanifold of $T^* \mathbb{R}^n_X \times T^* \mathbb{R}^n_Y$ with respect to the symplectic form

$$\omega_X - \omega_Y = \sum dx^i \wedge d\xi_i - dy^i \wedge d\eta_i.$$

693

The set Λ_Φ is called the *canonical relation* of Φ. To see that the canonical relation is Lagrangian, note that $T_{(x,y;\xi,\eta)}\Lambda_\Phi$ can be identified with pairs of tangent vectors

$$v \in T_{(x,\xi)}(T^* \mathbb{R}_X^n) \quad \text{and} \quad w \in T_{(y,\eta)}(T^* \mathbb{R}_Y^n)$$

such that $v = \Phi_*(w)$, and $\omega_X(\Phi_*(w_1), \Phi_*(w_2)) = \omega_Y(w_1, w_2)$.

The fact that we are reducing ourselves to the study of symplectomorphisms Φ gives us a hint as to how to define the resulting operator T_Φ, at least microlocally. Intuitively speaking, our discussion above shows that the wave front set of the kernel of Φ must be contained in Λ_Φ, because spectral singularities at a point (y, η) will be moved to singularities at a point (x, ξ), so we should expect that $\mathrm{WF}(T_\Phi f) = \Lambda_\Phi \circ \mathrm{WF}(f)$. We will see in the next section on *oscillatory integral distribution* (one can also see the idea immediately from stationary phase heuristics) that an operator of the form

$$T_\Phi f(x) = \int a(x, y, \theta) e^{2\pi i \phi(x,y,\theta)} f(y) \, d\theta \, dy$$

would have this property, provided that we choose the *phase function* $\phi : \mathbb{R}_x^n \times \mathbb{R}_y^n \times \mathbb{R}_\theta^N \to \mathbb{R}$ such that

$$\Lambda_\Phi = \{(x, y; -\nabla_x \phi(x, \theta), \nabla_y \phi(x, \theta)) : \nabla_\theta \phi(x, y, \theta) = 0\}$$

Λ_Φ

$$\int a(x, \theta) e^{2\pi i \phi(x,\theta)} \, d\theta$$

have a wave-front set contained in $\Lambda_\phi = \{(x, \nabla_x \phi(x, \theta)) : \nabla_\xi \phi(x) = 0\}$, and, provided ϕ is non-degenerate, Λ_ϕ is a Lagrangian manifold. Thus, given a symplectomorphism $\Phi : T^* \mathbb{R}^n \to T^* \mathbb{R}^n$, if we can find a phase $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that $\Lambda_\Phi = \Lambda_\phi$, then, modulo C^∞ kernels, we might expect to find a symbol a such that

$$T_\Phi f(x) = \int a(x, y, \theta) e^{2\pi i \phi(x,\theta)} f(y) \, d\theta \, dy.$$

More generally, we might not be able to find a phase ϕ that works globally for Φ, but we can localize, and once localized, the symplectic structure of Λ_Φ will actually guarantee the existence of ϕ. But this means that our operators might be given by finite sums of integral operators of the form above. We are now naturally reaching the study of general Fourier integral operators.

41.5 Hyperbolic Equations

Fourier integral operators were initially introduced to obtain parametrices for hyperbolic equations. To see how these arise, let us begin with a constant coefficient linear differential operator on $\mathbb{R}_t \times \mathbb{R}_x^n$, given by $P(\partial_t, D_x)$, where $P(\tau, \xi)$ is a polynomial, which we will assume can be written in the form $\tau^m + \tau^{m-1}Q_1(\xi) + \cdots + Q_m(\xi)$, where $Q_i(\xi)$ is a polynomial of degree at most i. Then the hyperbolic equation is

$$L = \partial_t^m + \partial_t^{m-1}Q_1(D_x) + \cdots + \partial_t Q_{m-1}(D_x) + Q_m(D_x).$$

We note that here ∂_t is the standard derivative operator, whereas D_x^α is the derivative operator, normalized by dividing by an appropriate power of $2\pi i$ so that it is the Fourier multiplier of ξ^α. We recall the Cauchy-Kovalevskaya theorem, which gives unique analytic solutions to the Cauchy problem $Lu = f$ given initial conditions $u_0, \partial_t u_0, \ldots, \partial_t^{m-1}u_0$, given that f, and the initial conditions are analytic functions on $\mathbb{R}^n$. But we are interested in more general existence results.

The surprising feature of this problem is that these operators need not even have solutions if we switch from studying analytical initial conditions to say, compactly supported smooth initial conditions. For instance, suppose there exists a distribution u on $\mathbb{R}_t \times \mathbb{R}_x$, tempered in the x-variable, such that $Lu = 0$, where $L = \partial_t - D_x$, and we let $u_0 \in \mathcal{S}(\mathbb{R}_x)^*$ be the initial value of the distribution. Then, taking Fourier transforms in the x variable, we conclude that $\partial_t \hat{u}(t, \xi) = \xi \hat{u}(t, \xi)$, which implies that $\hat{u}(t, \xi) = \hat{u}_0(\xi)e^{\xi t}$. But this distribution is *never* tempered in the ξ variable; if it was tempered for one positive value of t, and one negative value of t, then we could conclude that

$$|\hat{u}_0(\xi)| \lesssim e^{-\varepsilon|\xi|}$$

for some $\varepsilon > 0$. But the Paley-Wiener theorem and it's variants therefore imply that u_0 is analytic, and actually extends to a holomorphic function on a small strip containing the real line. Thus the existence of solutions to the Cauchy problem $\partial_t u - Du = 0$ is very delicate; in particular, there are no solutions with initial conditions in $\mathcal{D}(\mathbb{R}_x)$. This hints at the fact that to make the solution to the Cauchy problem tractable, we must ensure that the polynomial $P(\tau, \xi) = 0$ have *imaginary roots*. A desire to find a more powerful existence statement for solutions to such equations, tempered in the x-variable, will force us to choose polynomials $P(\tau, \xi)$ whose principal part has purely imaginary roots in the τ variable.

Let us discuss this situation more precisely. Let $\tilde{E}_0, \ldots, \tilde{E}_{m-1} : \mathbb{R}^n_\xi \times \mathbb{R}_t \to \mathbb{C}$ be the analytic solutions to the Cauchy problem $P(\partial_t, \xi) = 0$ with initial conditions

$$\partial_t^i \tilde{E}_j(0, \xi) = \delta_{ij}$$

for $0 \leqslant i \leqslant m - 1$. If we are to expect $P(\partial_t, D_x) = 0$ to be a well posed differential equation, then we should expect each solution $\tilde{E}_i$ to be tempered in the ξ variable. We can calculate the functions $\{\tilde{E}_i\}$ explicitly. If we fix ξ_0, and assume first that the roots of $P(\tau, \xi_0)$ in the τ variable are distinct, then the roots are distinct locally around ξ_0. If we let $\tau_1(\xi), \ldots, \tau_m(\xi)$ be the roots of the equation, then these are analytic functions in the ξ variable locally around ξ_0. The functions $h_i(t, \xi) = e^{i\tau_i(\xi)t}$ then satisfy the Cauchy problem $P(\partial_t, \xi) = 0$ with initial conditions

$$\partial_t^i h_j(0, \xi) = \tau_j(\xi)^i$$

for $0 \leqslant i \leqslant m - 1$. The uniqueness of analytic solutions guaranteed by the Cauchy-Kovalevsky theorem imply that

$$h_i(t, \xi) = \tilde{E}_0(t, \xi) + \tilde{E}_1(t, \xi)\tau_i(\xi) + \cdots + \tilde{E}_{m-1}(t, \xi)\tau_i^{m-1}(\xi).$$

If $\tilde{E} = (\tilde{E}_0, \ldots, \tilde{E}_{m-1})$ and $h = (h_0, \ldots, h_{m-1})$, then we can summarize this in the matrix equation

$$h = \begin{pmatrix} 1 & \tau_1 & \ldots & \tau_1^{m-1} \\ 1 & \tau_2 & \ldots & \tau_2^{m-1} \\ \vdots & \ddots & \ldots & \vdots \\ 1 & \tau_m & \ldots & \tau_m^{m-1} \end{pmatrix} \tilde{E}.$$

Provided that the roots $\{\tau_i\}$ are distinct, we can solve this equation to find the functions $\{\tilde{E}_i\}$ in terms of the functions $\{h_i\}$ using Cramer's rule. Namely, if $V(\tau_1, \ldots, \tau_m)$ is the Vandermonde determinant, i.e. the determinant of the matrix $\{\tau_i^j\}$, and if $V_i(\tau_1, \ldots, \tau_m; t)$ is the determinant of the matrix obtained by replacing the ith column with the vector $\{e^{t\tau_j(\xi)}\}$, then

$$\tilde{E}_i(t, \xi) = \frac{V_i(\tau_1(\xi), \ldots, \tau_m(\xi); t)}{V(\tau_1(\xi), \ldots, \tau_m(\xi))}.$$

For instance, if $m = 2$, then

$$\tilde{E}_1(t, \xi) = \frac{e^{t\tau_1(\xi)}\tau_2(\xi) - e^{t\tau_2(\xi)}\tau_1(\xi)}{\tau_2(\xi) - \tau_1(\xi)} \quad \text{and} \quad \tilde{E}_2(t, \xi) = \frac{e^{t\tau_2(\xi)} - e^{t\tau_1(\xi)}}{\tau_2(\xi) - \tau_1(\xi)}$$

696

Note that the function

$$(\tau_1, \ldots, \tau_m, t) \mapsto \frac{V_i(\tau_1, \ldots, \tau_m; t)}{V(\tau_1, \ldots, \tau_m)}$$

are analytic in t, symmetric in the variables $\{\tau_i\}$, and *entire* in the variables $(\tau_1, \ldots, \tau_m) \in \mathbb{C}^n$. For instance, in the case $m = 2$, when $\tau_1(\xi_0) = \tau_2(\xi_0) = \tau$ we have

$$\tilde{E}_1(t, \xi_0) = e^{t\tau}(1 - t\tau) \quad \text{and} \quad \tilde{E}_2(t, \xi_0) = te^{t\tau}.$$

Thus the equation we have constructed continues to specify the functions $\{\tilde{E}_i\}$ even when the roots of P in the ξ variable are not distinct.

By virtue of the roots switching around, one cannot necessarily define the functions $h_1, \ldots, h_n$ globally for all $\xi \in \mathbb{R}^n$ if roots coincide. But quantities symmetric in the variables $\{h_i\}$ are well defined, for instance, the function

$$S(t, \xi) = |\text{Re}(h_1(t, \xi))| + \cdots + |\text{Re}(h_n(t, \xi))|.$$

is globally defined. If $\tilde{E}_0, \ldots, \tilde{E}_{m-1}$ are tempered in the ξ variable, then it follows that the function s is tempered in the ξ variable, and thus satisfies some equation of the form

$$|S(t, \xi)| \lesssim_t \langle \xi \rangle^{N_t}$$

for some $N_t > 0$. But this means that

$$|\text{Re}(\tau_1(\xi))| + \cdots + |\text{Re}(\tau_m(\xi))| \lesssim 1 + \log\langle \xi \rangle.$$

The theory of semialgebraic sets (which applies because $(\tau_1, \ldots, \tau_m)$ are the projections onto the τ variable of solutions to the polynomial equation $P(\tau, \xi) = 0$) implies that this can only be possible if

$$|\text{Re}(\tau_1(\xi))| + \cdots + |\text{Re}(\tau_m(\xi))| \lesssim 1,$$

i.e. because we can only have polynomial growth on semialgebraic sets. This is actually a necessary and sufficient condition for the Cauchy problem to be solvable (a result of Garding). An operator with this property will be called *hyperbolic*.

There is an equivalent specification of being hyperbolic which is very useful to the study of such equations. If $L = P(\partial_t, D_x)$ is a hyperbolic constant-coefficient partial differential equation of order m, then we can consider the degree m polynomial $P_m(\tau, \xi)$. We claim the roots of P_m in the τ variable differ from the roots of P by at most $O(1)$ for $|\xi| \gg 1$. Note that this is the only part of the discussion

where we have used the fact that the polynomials Q_i have degree at most i. This means that the condition

$$|\text{Re}(\tau_1(\xi))| + \cdots + |\text{Re}(\tau_m(\xi))| \lesssim 1$$

holds if and only if the roots of P_m are all purely imaginary. Thus $P(\partial_t, D_x)$ is hyperbolic if and only if the roots of P_m in the τ variable are all purely imaginary.

In addition, we say a differential operator $P(\partial_t, D_x)$ is *strongly*, or *strictly hyperbolic* if in addition to being hyperbolic, the imaginary roots of P_m are all distinct from one another for all $\xi \in \mathbb{R}^d$, which implies the roots of P are distinct for large ξ. If we label the roots of P_m as $i\lambda_1(\xi), \ldots, i\lambda_m(\xi)$, then the functions $\{\lambda_i\}$ are real-valued analytic functions in $\mathbb{R}^n - \{0\}$, which are homogeneous of degree one. Thus $|\lambda_i(\xi) - \lambda_j(\xi)| \gtrsim |\xi|$ for $i \neq j$.

TODO: Construct Parametrix.

Remark. A similar phenomenon to the study of parametrices for hyperbolic equations arises in the study of the half-wave equation associated with positive semidefinite pseudodifferential operators. More precisely, if P is a positive-semidefinite pseudodifferential operator of order one on a compact manifold M, then one can construct a parametrix to the half-wave equation

$$\partial_t = 2\pi i P$$

which also has an expression as a Fourier integral. This is done in more detail in the later parts of these notes on the geometry of eigenfunctions.

41.6   Propogation of Singularities

One important relation between u and $\text{WF}(u)$ 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 $(x, \xi) \in \text{WF}(u) - \text{WF}(v)$,

$$q(x, \xi) = \sum_{|\alpha| \leqslant K} a_\alpha(x)\xi^\alpha = 0,$$

and $WF(u) - WF(v)$ 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 $WF(v_t) \subset WF(u_{tt})$.

Then the Paley-Wiener theorem implies that $\hat{u}$ is an analytic function on $\mathbb{R}^d$. If $\hat{u}$ decays rapidly, then u is also a smooth function. However, even if u is not smooth, $\hat{u}$ 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 $x \in \mathbb{R}^d$ if u is locally a C^∞ function in a neighbourhood of x, i.e. there exists a bump function $\phi \in C^\infty(\mathbb{R}^d)$ with $\phi(x) \neq 0$ such that $\phi u \in C^\infty(\mathbb{R}^d)$. The *singular support* of a compactly supported distribution u to be the set of all points $x \in \mathbb{R}^d$ upon which u is not nonsingular.

A degree m constant coefficient linear differential operator $P(D)$ on $\mathbb{R}^n$ is said to be of *real principal type* if the principal symbol P_m is a real-coefficient polynomial, and ∇P_m is non-vanishing on $\mathbb{R}^n - \{0\}$.

Lemma 41.12. *If $P(D)$ is a differential operator of real principal type, then there exists distributions E_+ and E_- which are parametrices for P,*

$$WF(E_+) \subset (\{0\} \times \mathbb{R}^n) \cup \{(t \cdot \nabla P_m(\xi), \xi) : t > 0, \xi \in Char(P), P_m(\xi) = 0\},$$

and

$$WF(E_+) \subset (\{0\} \times \mathbb{R}^n) \cup \{(t \cdot \nabla P_m(\xi), \xi) : t < 0, \xi \in Char(P), P_m(\xi) = 0\}.$$

41.7 Littlewood Paley Theory

Let T be a Fourier integral operator from an open set $U \subset \mathbb{R}^n$ to $V \subset \mathbb{R}^m$ associated with a Lagrangian distribution

$$\Lambda \subset (T^*X - 0_X) \times (T^*Y - 0_Y).$$

Qualitatively speaking, wavefront set analysis tells us the following:

- T maps $\mathcal{D}(X)$ into $\mathcal{E}(Y)$.

- T extends to a map from $\mathcal{E}(X)^*$ into $\mathcal{D}(Y)^*$.

Let's see some more quantitative consequences, namely, that we can apply *Littlewood-Paley theory* to these operators.

It will be easiest to work locally in coordinates. We can therefore assume that U and V are open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, and that T has an integral expression of the form

$$Tf(x) = \int \int_{\mathbb{R}^p} a(x, y, \theta) e^{2\pi i \phi(x,y,\theta)} f(y) \, dy \, d\theta,$$

where ϕ has the property that for any $(x, y, \theta) \in \mu\text{-supp}(a)$ with $\nabla_\theta \phi(x, y, \theta) = 0$, we have *both* $\nabla_x \phi(x, y, \theta) \neq 0$ and $\nabla_y \phi(x, y, \theta) \neq 0$, and a is a symbol of order μ in the θ variable, which (working modulo the family of smoothing operators), we may assume to be supported on a δ neighborhood of Λ. We fix a smooth, compactly supported function $\eta \in C_c^\infty(\mathbb{R}^d)$ supported on $\{1/2 \leqslant |\xi| \leqslant 2\}$, and

consider the Littlewood-Paley cuttoff operator $L_k = \eta(D/2^k)$. We now compute the operator

$$L_{k_1} \circ T \circ L_{k_2}.$$

Firstly, we calculate that $L_{k_1} \circ T$ can be expressed as on oscillatory integral of the form

$$\int a_{k_1}(x, y, \theta) e^{2\pi i \phi(x,y,\theta)} \, d\theta,$$

where if $\phi_1(x, y, z, \theta) = \phi(z, y, \theta) - \phi(x, y, \theta) + (x - z) \cdot \nabla_x \phi(x, y, \theta)$, then

$$a_{k_1}(x, y, \theta) = e^{-2\pi i \phi(x,y,\theta)} \eta(D/2^{k_1}) \{ e^{2\pi i \phi(x,y,\theta)} a(x, y, \theta) \}$$

$$\sim \sum_\alpha \frac{1}{2^{|\alpha| k_1} \cdot \alpha!} (\partial^\alpha \eta) \left(\frac{\nabla_x \phi(x, y, \theta)}{2^{k_1}} \right) D_z^\alpha \left\{ e^{2\pi i \phi_1(x,y,z,\theta)} a(z, y, \theta) \right\} \big|_{z=x}.$$

We therefore calculate that $L_{k_1} \circ T \circ L_{k_2}$ can be expressed as

$$\int a_{k_1,k_2}(x, y, \theta) e^{2\pi i \phi(x,y,\theta)} \, d\theta,$$

where, if $\phi_2(x, y, w, \theta) = \phi(x, w, \theta) - \phi(x, y, \theta) + (y - w) \cdot \nabla_y \phi(x, y, \theta)$, then

$$a_{k_1,k_2}(x, y, \theta) \sim \sum_\beta \frac{1}{2^{|\beta| k_2} \cdot \beta!} (\partial_\xi^\beta \eta) \left(-\frac{\nabla_y \phi(x, y, \theta)}{2^{k_2}} \right) D_w^\beta \{ e^{2\pi i \phi_2(x,y,w,\theta)} a_{k_1}(x, w, \theta) \} \big|_{w=y}$$

$$\sim \sum_{\alpha,\beta} \frac{1}{\alpha! \beta!} \frac{1}{2^{|\alpha| k_1 + |\beta| k_2}} (\partial_\xi^\beta \eta) \left(-\frac{\nabla_y \phi(x, y, \theta)}{2^{k_2}} \right)$$

$$D_z^\alpha D_w^\beta \left\{ e^{2\pi i [\phi_1(x,z,w,\theta) + \phi_2(x,y,w,\theta)]} (\partial^\alpha \eta) \left(\frac{\nabla_x \phi(x, w, \theta)}{2^{k_1}} \right) a(z, w, \theta) \right\} \Big|_{\substack{z=x \\ w=y}}.$$

Now we apply our assumption. The set

$$\{ (x, y, \theta) : |\theta| = 1, \nabla_\theta \phi(x, y, \theta) = 0 \}$$

is compact. Applying this compactness, together with homogeneity, we can find $C > 0$ such that if $\nabla_\theta \phi(x, y, \theta) = 0$, then

$$(C/2)^{-1} |\theta| \leqslant |\nabla_x \phi(x, y, \theta)| + |\nabla_y \phi(x, y, \theta)| \leqslant (C/2) |\theta|.$$

Modulo smoothing operators, we have $\text{supp}(a_{k_1,k_2}) \subset \text{supp}(a)$. Moreover, for $(x, y, \theta) \in \text{supp}(a)$, we can find (x', y', θ') such that $|\theta| = |\theta'|$,

$$|x - x'| + |y - y'| + \frac{|\theta - \theta'|}{|\theta|} \leqslant \delta,$$

701

and $\nabla_\theta \phi(x', y', \theta) = 0$. If δ is chosen appropriately, depending only on ϕ, we conclude that

$$C^{-1}|\theta| \leqslant |\nabla_x \phi(x, y, \theta)| + |\nabla_y \phi(x, y, \theta)| \leqslant C|\theta|.$$

In order for a_{k_1, k_2} to not be smoothing on a neighborhood of (x, y, θ), we therefore must have

$$C^{-1} 2^{k_2 - 1} \leqslant |\theta| \leqslant C 2^{k_2 + 1}.$$

and

$$C^{-1} 2^{k_1 - 1} \leqslant |\theta| \leqslant C 2^{k_1 + 1},$$

i.e. so that the gradients $\nabla_x \phi$ and $\nabla_y \phi$ are contained in the support of $\eta(\cdot/2^{k_1})$ and $\eta(\cdot/2^{k_2})$ respectively. But this is impossible if $|k_2 - k_1| \geqslant \log_2(4C^2)$. Thus we should expect in this circumstance that the operator $L_{k_1} \circ T \circ L_{k_2}$ is arbitrarily well behaved. Moreover, under this assumption we should expect that the kernel of $L_{k_1} \circ T \circ L_{k_2}$ is infinitely differentiable, and each derivative is $O_N(2^{-N(k_1+k_2)})$ for all $N > 0$. This means that in regularity problems involving these operators, if we perform a decomposition

$$T = \sum_{k_1, k_2} = L_{k_1} \circ T \circ L_{k_2},$$

then we should only have to worry about the analysis of the components $L_{k_1} \circ T \circ L_{k_2}$ with $|k_1 - k_2| \lesssim 1$. By applying the triangle inequality (so that you can assume that non-zero Littlewood-Paley parts of the operator are sparse), or widening the support of the function η, we can actually assume that we have an equality of the form

$$T \approx \sum_k L_k \circ T \circ L_k,$$

where the difference between the two sides is a smoothing operator.

A similar phenomenon happens in the study of spectral multipliers on a Riemannian manifold M. It is then natural to simultaneously apply the Littlewood-Paley theory in coordinates, i.e. the operators $\{L_k\}$, but in addition, Littlewood-Paley theory on the manifold, i.e. considering the operators $\{L'_k\}$ given by the equation

$$L'_k = \eta(P/2^k),$$

where P is a classical self-adjoint positive-semidefinite elliptic pseudodifferential operator of order one, on M, with symbol compactly localized in space to the

coordinate system U, such that in this coordinate system, P has symbol p, and principal symbol p_1, and ψ is another compactly supported smooth function on M localizing the multiplier to a coordinate system. It is then natural to consider expressions of the form

$$L_{k_1} \circ T \circ L'_{k_2}.$$

For two symbols a and b, let

$$a \circ b \sim \sum_\alpha \frac{1}{\alpha!} D_\xi^\alpha a \cdot \partial_x^\alpha b$$

be the composition operator obtained via the calculus of pseudodifferential operators. In Strichartz (1972), it is shown that $\eta(P/2^k)$ is a pseudodifferential operator, and has a symbol $\tilde{\eta}_k$ which can be asymptotically expanded as

$$\tilde{\eta}_k(x,\xi) \sim \sum_{m \leqslant n} \frac{(-1)^m}{m!(n-m)!} \frac{1}{2^{kn}} \{(\partial_x^n \eta)(p_1/2^k) p_1^m\} \circ \overbrace{(p \circ \cdots \circ p)}^{n-m \text{ times}}.$$

Thus, in particular, we have that

$$|D_x^\alpha D_\xi^\beta \{\tilde{\eta}_k(x, 2^k \xi)\}| \lesssim_{\alpha,\beta,N} \langle \xi \rangle^{-N},$$

uniformly in k. Applying the composition calculus for FIOs, as above, shows that for $|k_1 - k_2| \gtrsim 1$, we also have that $L_{k_1} \circ T \circ L'_{k_2}$ is infinitely differentiable, and each derivative is $O_N(2^{-N(k_1+k_2)})$.

Part V

Linear Partial Differential Equations

Here we study *constant coefficient* linear partial differential equations

$$Lf(x) = \sum a_\alpha D^\alpha f(x)$$

and the *non-constant coefficient* linear partial differential equations

$$Lf(x) = \sum a_\alpha(x) D^\alpha f(x)$$

using methods of harmonic analysis. The central problems of the theory are *existence* and *regularity*: Given f, does a solution u to the equation $Lu = f$ exist, and if so, how does the regularity of f relate to the regularity of u. We often phrase the regularity problems in terms of function spaces. Let $\{X^s\}$ be a family of function spaces, indexed by a parameter $s \in \mathbb{R}$, measuring the *regularity* of functions in these spaces. Given a linear partial differential operator L, a regularity result might say that for some s and t, if $Lu = f$, where $f \in X^s$, then u in X^t. Such results often require one to use the special properties of the operator L; general theories are rarely available, except for certain very special families of partial differential equations, especially the *elliptic partial differential equations*.

Chapter 42

Green's Functions

Let Ω be a bounded, open set on $\mathbb{R}^n$. Recall Green's second identity,

$$\int_\Omega (u\Delta v - v\Delta u) = \int_{\partial\Omega} u\frac{\partial v}{\partial \eta} - v\frac{\partial u}{\partial \eta}.$$

We introduce the quantity

$$u_x(y) = \frac{1}{|x - y|^{n-2}},$$

which is Newton's gravitational potential, or Coulomb's electrostatic potential given a point mass at the point x. A simple computation shows that $\Delta u_x(y) = 0$ for $y \neq x$, perhaps using the fact that if $u(x)$ is a radial function, with $u(x) = f(r)$ for $r = |x|$, then

$$\Delta u(x) = f''(r) + \frac{n - 1}{r}f'(r).$$

Plugging this function into Green's identity, if $v \in C^2(\Omega)$ is a Harmonic function on a domain Ω, continuous on the closure of Ω, then for any x, applying Green's identity on the domain $\Omega_{x,\varepsilon}$, obtained from Ω by removing the ball B of radius ε centered at x, where ε is small enough so that the entire ball is contained in Ω, we find that

$$0 = \int_{\partial\Omega_{x,\varepsilon}} \left(u_x\frac{\partial v}{\partial \eta} - v\frac{\partial u_x}{\partial \eta} \right)$$

$$= \int_{\partial\Omega} \left(u_x\frac{\partial v}{\partial \eta} - v\frac{\partial u_x}{\partial \eta} \right) - \int_{\partial B} u_x\frac{\partial v}{\partial \eta} - v\frac{\partial u_x}{\partial \eta}.$$

706

Now for $y \in \partial B$, with $y = x + z$, $\eta(y) = \varepsilon^{-1}z$. And

$$\nabla u_x(y) = -(n-2)\frac{z}{|z|^n},$$

so

$$\int_{\partial B} u_x \frac{\partial v}{\partial \eta} - v\frac{\partial u_x}{\partial \eta} = \int_{|z|=1} \varepsilon \cdot \nabla v(x + \varepsilon z) \cdot z + (n-2)v(x + \varepsilon z).$$

Applying a Taylor series shows that as $\varepsilon \to 0$, this quantity converges to $(n-2)C_d v(x)$, where C_d is the surface area of S^{d-1}. Thus we conclude that

$$v(x) = \frac{-1}{C_d(n-2)}\int_{\partial\Omega}\left(u_x\frac{\partial v}{\partial \eta} - v\frac{\partial u_x}{\partial \eta}\right)$$

Thus v is determined by it's boundary values, and boundary derivatives in the direction normal to Ω. To get rid of these normal boundary derivatives, we suppose we can compute a function h_x, which is harmonic in Ω, and equal to u_x on $\partial\Omega$. Then we can apply Green's second identity again, which yields

$$0 = \int_{\partial\Omega} v\frac{\partial h_x}{\partial \eta} - \frac{\partial v}{\partial \eta}G$$

$$= \int_{\partial\Omega} v\frac{\partial h_x}{\partial \eta} - \frac{\partial v}{\partial \eta}u_x.$$

Thus we conclude that

$$v(x) = \frac{-1}{C_d(n-2)}\int_{\partial\Omega} v\left(\frac{\partial h_x}{\partial \eta} - \frac{\partial u_x}{\partial \eta}\right).$$

The function

$$g_x(y) = \frac{-1}{C_d(n-2)}(h_x(y) - u_x(y))$$

is called the *Green's function* on the domain Ω. It solves the equation $\Delta g_x = \delta_x$ on Ω, while vanishing on the boundary. We have calculated this by virtue of the fact that for any harmonic function v on Ω,

$$v(x) = \int_{\partial\Omega} \frac{\partial g_x}{\partial \eta}v \, dy.$$

Thus v is uniquely determined by it's boundary values. What's more, we now have a natural operator which, given some sufficiently boundary conditions on $\partial\Omega$, allows us to find a unique solution to Laplace's equation on Ω.

707

Chapter 43

Elliptic Equations

Consider a partial differential operator $Lf(x) = \sum a_\alpha(x) D^\alpha f(x)$, where $a_\alpha \in C^\infty(\mathbb{R}^n)$. Associated with this operator is it's symbol $a(x,\xi) = \sum a_\alpha(x)\xi^\alpha$, and if L has order m, it's principle symbol is $a_m(x,\xi) = \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha$. We say L is *elliptic* near x_0 if $a_m(x_0,\xi) = 0$ only when $\xi = 0$. This is equivalent to assuming that there exists $M > 0$, and two constants c_1 and c_2 and a small neighborhood U of x_0 such that $c_1|\xi|^m \leqslant |a(x,\xi)| \leqslant c_2|\xi|^m$ for $|\xi| \geqslant M$.

Elliptic operators have a near complete theory, because they are *invertible* modulo smoothing operators. More precisely

Chapter 44

Propogation Speed

One important feature of a partial differential equation is the manner in which data *propogates* through a solution. Our goal in this chapter is to analyze when partial differential equations have a *finite speed of propogation*. One feature of many important partial differential equations (of which the wave equation is an exemplary example) is the fact that values of the equation *propogate through time* at a finite speed c. This is a property of many different equations, whose study is the aim of this chapter. So let us consider a general linear partial differential equation of the form $\partial_t u = Lu$, where we do not have to deal with higher-derivatives in the t-variable because we allow u instead to be *vector-valued*, i.e. u is a function from $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^m$ for some m, and $(Lf)_i = \sum c_{i,j,\alpha}(x,t)D_x^\alpha f_j$.

To determine what properties of the equation are *necessary* for the equation to have a finite speed of propogation, we assume that the problem of existence and uniqueness for the Cauchy problem of this equation has been solved, in the sense that for any function $f \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^m)$ and any $t_0 \in \mathbb{R}$, there exists a unique solution $u(x,t)$ (let's say, lying in the space of distributions), defined for $t \geqslant t_0$, solving the differential equation $\partial_t u = Lu$, and with $u(x,t_0) = f(x)$ for all $x \in \mathbb{R}^d$. Thus we can consider operators $\{S_{t_0,t_1} : t_0 \leqslant t_1\}$ such that, with the notation as above, $S_{t_0,t_1}f(x) = u(x,t_1)$. Since we are working with linear differential equations, the operators S_{t_0,t_1} are *linear operators*, and by assumption they are Schwartz, i.e. they map $\mathcal{D}(\mathbb{R}^d, \mathbb{R}^m)$ to $\mathcal{D}(\mathbb{R}^d, \mathbb{R}^m)^*$. In particular, the linearity implies that the *principle of superposition* applies, i.e. that $S_{t_0,t_1}(f+g) = S_{t_0,t_1}f + S_{t_0,t_1}g$.

To formally analyze how data propogates in solutions to the equations above, we introduce the idea of *influence* and *dependence*. For $t < s$, we say that a point (x,t) in space-time *does not influence* a point (y,s) if there exists a spatial neighborhood U of x, and a space-time neighborhood V of (y,s), such that for

any f_1 and f_2 which are equal outside of U, $S_{t,s}f_1$ is equal to $S_{t,s}f_2$ on V. Using linearity, this is equivalent to the property that, for any f supported on U, $S_{t,s}f$ is supported outside of V. The *domain of influence* of (x,t) is then the set of all (y,s) which *are* influenced by (x,t). The *domain of dependence* $\mathcal{D}_D(y,s)$ of a point (y,s) is precisely the set of all (x,t) such that (y,s) is influenced by (x,t). We define

$$\mathcal{D}_{I,s}(x,t) = \{y : (y,s) \in \mathcal{D}_I(x,t)\} \quad \text{and} \quad \mathcal{D}_{D,t}(y,s) = \{x : (x,t) \in \mathcal{D}_D(y,s)\}$$

to be the domains at particular time slices.

We can now formally state our goal: what are the conditions which ensure a *finite speed of propogation c*, i.e. such that the domain of influence of any point (x_0, t_0) is contained in the cone

$$\Gamma_{x_0,t_0} = \{(x_1, t_1) : |x_1 - x_0| \leqslant c|t_1 - t_0|\}.$$

We will begin by analyzing the properties of equations with a finite speed of propogation, in order to try and obtain a condition that can be reversed to ensure a finite speed of propogation. So we now assume that our equation has a finite speed of propogation c. Here are some immediate properties.

Theorem 44.1. *Suppose that $\partial_t = L$ is a partial differential equation with a finite speed of propogation c. Then the following properties hold:*

- *(Finite Propogation of Support) If $\text{supp}(f) \subset K$, then the support of $S_{t_0,t_1}f$ is contained in the $t_1 - t_0$ thickening of the set K.*

- *(Domains of Determinacy) Suppose that $t < s$, and U is an open neighborhood of $\mathcal{D}_{D,t}(y,s)$. Then if f vanishes on U, then $S_{t,s}f$ vanishes on a neighborhood of (y,s).*

- *(Huygen's Wave Construction): For $t_0 < s < t_1$, $\mathcal{D}_{I,t_1}(x_0,t_0)$ is contained in the union of all of the sets $\mathcal{D}_{I,t_2}(y,s)$, where y ranges over all elements of $\mathcal{D}_{I,s}(x_0,t_0)$.*

It is now clear that for any point (x_0, t_0), $\mathcal{D}_I(x_0, t_0)$ intersects the hyperplane $\Sigma = \{t = t_0\}$ at the unique point (x_0, t_0), i.e. $\mathcal{D}_{I,t_0}(x_0, t_0) = \{(x_0, t_0)\}$. More strongly, $\mathcal{D}_I(x_0, t_0)$ must intersect this point on Σ transversally, i.e. at angles $\theta \in [0, \pi]$ bounded from below by some $\delta > 0$, i.e. such that

$$|\cos(\theta)| \leqslant \frac{c}{(1+c^2)^{1/2}} < 1.$$

Related to this property, we say a hypersurface $\Sigma = \{t = p(x)\}$ is *space-like* if no pair of points on Σ influences each other, and *strictly space-like* if there exists $\delta > 0$ such that for any $(x,t) \in \Sigma$, $\mathcal{D}_I(x,t)$ intersects the tangent space to $T_{(x,t)}\Sigma$ at an angle greater than δ. It follows from the assumptions that if $|\nabla p(x)| < 1/c$, then Σ is strictly space-like.

Let us consider a partial differential equation of the form $\partial_t u = L_t u$, where $u(x,t)$ takes on values in some finite-dimensional vector space V, and L_t is a linear partial differential equation. We will be interested in the

Let us assume that the Cauchy problem for such equations has been solved, i.e. that given the specification of u at time $t = 0$, one can uniquely solve for the equation at future times, i.e. so that there exists a family of solution operators $\{S_{t_0,t_1} : t_0 < t_1\}$ taking initial conditions at a particular time into the future. We say that the space-time point (x_0, t_0) *influences* a space-time point (x_1, t_1) if $S_{t_0,t_1} u_0$

We say a point p influences a point q if there exists two

Chapter 45

Dispersive Equations

TODO: Move vector valued approach somewhere else: We begin with constant-coefficient linear dispersive equations i.e. considering a first order system $\partial_t u = Lu$, where $u : V \times \mathbb{R} \to \mathbb{R}$ for some finite dimensional Hilbert space V, and where $Lu = \sum c_\alpha D_x^\alpha u$ for some $c_\alpha \in \text{End}(V)$. In order to have any hope of a well posed theory finite speed of propogation, we assume that L is formally skew-adjoint, i.e. such that for all test functions $u, v \in \mathcal{D}(V)$, $\langle Lu, v \rangle = -\langle u, -Lv \rangle$. If we write $L = ih(D_x)$ for some polynomial $h : \mathbb{R}^d \to \text{End}(V)$, then this is equivalent to the coefficients of h being self-adjoint. The polynomial h is called the *dispersion relation* of the equation. If $V = \mathbb{R}^m$, then the assumptions imply that h is a system of real-coefficient polynomials.

A partial differential equation is *dispersive* if wave-like solutions to the partial differential equation with differing frequencies travel in different directions and speeds, so that linear combinations of such solutions tend to 'spread out in space' over time. This phenomenon presents itself in various *dispersive estimates* for such partial differential equations.

We begin with a constant coefficient partial differential equation of the form $P(D_t, D_x) = 0$ which is degree m in time, where the roots of the polynomial $P(\omega, \xi) = 0$ in the ω variable are all real-valued. This equation is called the *dispersion relation* of the equation. In the case where the equation is of the form $\partial_t = ih(D_x)$, the solutions are of the form $\omega = h(\xi)$, and we call the function h the dispersion relation. For higher degrees the function h must be treated as a multi-valued function.

The *planar waves* of the equation are precisely those of the form

$$u(x, t) = e^{2\pi i(\xi \cdot x + \omega t)}$$

where $\xi \in \mathbb{R}^d$, $\omega \in \mathbb{R}$, and $P(\omega, \xi) = 0$. Thus the dispersion relation $P(\omega, \xi) = 0$ relates the *wave number ξ* of the planar wave with the *angular frequency ω* of the planar waves involved. In particular, looking at the phase $\xi \cdot x + \omega t$ and trying to solve in the x-variable, the high and low points of the waves in space will look like they are travelling at a *phase velocity*

$$-\frac{\xi}{|\xi|} \frac{h(\xi)}{|\xi|}.$$

In particular, the phase speed is $|h(\xi)|/|\xi|$.

More generally, we can consider superposition solutions using the Fourier inversion formula, i.e. writing

$$u(x, t) = \int e^{2\pi i(\xi \cdot x + h(\xi)t)} \widehat{u_0}(\xi) \, d\xi$$

Taking $u_0(\xi) = e^{2\pi i \xi \cdot x}$ gives the planar wave solutions above. A *wave packet* solution is obtained by taking u_0 to be supported on a small region of phase-space, i.e. localized to a particular position in space, and a particular frequency. One example localized near (x_0, ξ_0) would be of the form

$$u_0(x) = e^{2\pi i \xi_0 \cdot x} \phi\left(\frac{x - x_0}{\delta}\right),$$

713

where ϕ is a Schwartz function whose Fourier transform ψ is smooth and compactly supported. Fourier inversion tells us that

$$u(x,t) = \left(\int \psi(\xi) e^{2\pi i [(\xi_0 + \xi)\cdot x + h(\xi_0 + \xi)t - \xi \cdot x_0]} \, d\xi \right).$$

The principle of nonstationary phase tells us that $u(x,t)$ is spatially concentrated where $x = x_0 - \nabla h(\xi_0)t + O(t)$. Thus the phase envelope of u starts travelling starting at x_0, and travels at a *group velocity* $-\nabla h(\xi_0)$. The slight differences in phase velocity for indiidual planar waves under which u is composed in superposition can lead to the phase envelope spreading apart in time at a rate of $O(t)$, i.e. *dispersing in space*.

Example. *For $\omega_0 \in \mathbb{R}$, we can consider the* phase rotation equation

$$\partial_t u = 2\pi i \omega_0,$$

which can be explicitly solved by writing

$$u(x,t) = e^{2\pi i \omega_0 t} u_0(x).$$

Thus solutions have the envelope defined by u_0, and oscillate in time at the angular frequency ω_0. In particular, we have planar-wave solutions of the form

$$e^{2\pi i (\xi \cdot x + \omega_0 t)}.$$

The phase velocity here is $v_p(\xi) = -\omega_0/\xi$, i.e. planar waves with a high spatial frequency will appear to travel through space much more slowly. The group velocity, however, is 0, so a superposition of waves with similar spatial frequency will remain in the same position.

Example. *Consider the* transport equation

$$\partial_t u = -v_0 \cdot \nabla_x u$$

for some $v_0 \in \mathbb{R}^d$. The solution to this equation is given by

$$u(x,t) = u_0(x - v_0 t).$$

The planar-wave solutions are of the form

$$e^{2\pi i [\xi \cdot x - t(\xi \cdot v_0)]}.$$

714

The dispersion relation is given by $h(\xi) = -v_0 \cdot \xi$. Thus we obtain that the phase velocity is $v_p(\xi) = \xi(v_0 \cdot \xi)/|\xi|^2$; if ξ is aligned in the direction v_0, then the phase velocity will be v_0, and if ξ is orthogonal to v_0, the phase velocity will be zero, since the solution will be stationary in time. The group velocity is v_0, i.e. a wave packet will always roughly move in the direction v_0.

Example. *In the* free Schrödinger equation $i\partial_t u = -(\hbar/2m)\Delta u$, *the dispersion relation is $h(\xi) = -(\hbar/2m)|\xi|^2$, and the phase velocity is $v_p(\xi) = (\hbar/2m)\xi$. The group velocity is seen to be $v_g(\xi) = (\hbar/m)\xi$. Thus we see that the phase of a wave packet travels at half it's group velocity. More interesting is the group velocity equation, which in physics is used as a derivation of* De Broglie's law $mv = \hbar\xi$, *i.e. that the frequency of a wave packet is tied to it's velocity in time.*

Example. *The* one-dimensional Airy equation *is $\partial_t u = \partial_x^3 u$. It has dispersion relation $h(\xi) = |\xi|^3$. The phase velocity is $v_p(\xi) = \xi|\xi|$, and the group velocity is $v_g(\xi) = 3\xi|\xi| = 3v_p(\xi)$.*

Example. *Consider the* wave equation $\Box u = 0$, *where $\Box$ is the d'Alembertian operator $(2\pi)^{-2}\Delta - \partial_t^2$. The resulting polynomial is $P(\tau, \xi) = -(\xi^2 + \tau^2)$. The dispersion relation here is thus given by the equation $\xi^2 - h(\xi)^2 = 0$, i.e. we have $h(\xi) = \pm\xi$. The phase velocities are $v_p(\xi) = \pm\xi/|\xi|$, and the group velocities are also $v_g(\xi) = \pm\xi/|\xi|$.*

Example. *The* Klein-Gordan equation $\Box u = u$ *describes the movement of electrons where relativity is relevant. It is given by the polynomial $P(\tau, \xi) = -(|\xi|^2 + \tau^2 + 1)$, and has dispersion relation $h(\xi) = \pm\sqrt{1 + |\xi|^2}$. For large ξ, we thus see waves here behave like the wave equation, though for small ξ, waves behave like the transport equation.*

Our goal is to study *dispersive equations*, i.e. equations where wave packet solutions tend to spread apart over time. Since the velocity of a wave packet travelling at frequency ξ_0 is $\nabla h(\xi_0)$, wave packets near ξ_0 will travel at different frequencies if the *Hessian $H(\xi_0)$* of h is invertible. If this is true, the equation is called *fully dispersive*, like the Schrödinger equation and Airy equation. In dimension one, for the wave equation, the Hessian is zero, so in this case the wave equation is very non-dispersive. But in d dimensions, we see the Hessian has rank $d - 1$, so we at least have 'some dispersion', but not in the radial direction.

For an equation of the form $\partial_t u = Lu$, taking Fourier transforms in the x-variable tells us that

$$\widehat{u}(\xi, t) = e^{ih(\xi)t}\widehat{u}_0(\xi).$$

715

Thus the propogators $\{e^{Lt}\}$ are Fourier multipliers, with symbol $m_t(\xi) = e^{ih(\xi)t}$. In particular, taking inverse Fourier transforms yields the convolution kernels K_t such that $e^{Lt}f = K_t * f$. The kernels $\{K_t\}$ are precisely the fundamental solutions to the equation.

Example. *The propogators for the free Schrödinger equation $\partial_t u = i(\hbar/2m)\Delta u$ are Fourier multipliers with symbol $m_t(\xi) = -(\hbar/2m)|\xi|^2$. Using some complex analysis and the Fourier transform of the Gaussian, we can obtain the inverse Fourier transform, the* Schrödinger kernel

$$K_t(x) = t^{-d/2} \cdot (2\pi i\hbar/m)^{-d/2} e^{i(m/2\hbar t)|x|^2}.$$

This formula is quite remarkable, since it is not obvious that this kernel converges to the Dirac delta as $t \to 0$, given a balance between the growing singularity near the origin and the more and more rapid oscillation of the phase. The kernel is not *integrable – it does not even have any decay at infinity. But it is locally integrable, which results in the phenomenon that if u_0 is compactly supported and integrable, then $K_t * u_0$ will be* smooth, *but not locally integrable. This is an instance of the phenomenon of* local smoothing *– the group velocity $v_g(\xi) = (\hbar/m)\xi$ tells us that wave packets localized near large frequencies travel at a large frequency. Thus if we have a sum of wave packets localized in space, the high frequency packets will quickly leave this region, leaving only the smaller frequency packets, which result in a smooth function being left behind.*

Example. *The fundamental solution for the one dimension Airy equation $\partial_t u = \partial_x^3 u$ has a fundamental solution of the form*

$$K_t(x) = t^{-1/3} Ai(x/t^{1/3}),$$

where $Ai(x)$ is the Airy *function. The group velocity $v_g(\xi) = |\xi|\xi$ leads to high frequency terms moving much faster, so we should expect a much greater local smoothing for this equation. On the other hand, this should lead to wave packets spreading apart much faster, resulting in solutions having less decay.*

Example. *The wave equation $\Box u = 0$ has two fundamental solutions. This is because if we take Fourier transforms we are lead to the equation*

$$\hat{u}(\xi, t) = \cos(t|\xi|) \cdot \hat{u}_0(\xi) + \frac{\sin(t|\xi|)}{|\xi|} \widehat{\partial_t u_0}(\xi).$$

Thus we have $u(\xi, t) = [K^0 * u_0] + [K^1 * (\partial_t u_0)]$. TODO: 71 of Tao discussing these.

45.1 Strichartz Estimates

Our goal is to obtain *space-time estimates* of the form $\|u\|_{L_t^q L_x^r}$ on solutions u to a dispersive partial differential equation. To be concrete, we'll work with the Schrödinger equation $i\partial_t u = -(1/2)\Delta u$. The wave propogators $T_t = e^{itL}$ for this equation are Fourier multipliers with a symbol $m_t(\xi) = e^{it|\xi|^2/2}$ with $|m_t(\xi)| = 1$ for all ξ. It follows from Plancherel that these operators preserve the L^2 norm

$$\|e^{itL}f\|_{L_x^2} = \|f\|_{L_x^2}.$$

More generally, all the L^2 Sobolev norms are conserved, i.e.

$$\|e^{itL}f\|_{H_x^s} = \|f\|_{H_x^s}.$$

Thus the best space-time estimate we can obtain here is $L_x^2 \to L_t^\infty L_x^2$, and more generally, $H_x^s \to L_t^\infty H_x^s$ estimates.

We have explicitly computed the convolution kernel associated with the wave propogators in the last section, namely, we have

$$K_t(x) = t^{-d/2} \cdot (2\pi i)^{-d/2} e^{it|x|^2/2}.$$

Simply by applying the triangle inequality, we thus obtain that

$$\|T_t f\|_{L_x^\infty} = \|K_t * f\|_{L_x^\infty} \leqslant \|K_t\|_{L_x^\infty} \|f\|_{L_x^1} \lesssim |t|^{-d/2} \|f\|_{L_x^1}.$$

This is a *dispersion estimate*, since it says that for large t, despite the L^2 norm concentration, the solution is spreading out in space.

The L^2 energy of f might be conserved by the propogators, but it is spread out further and further in space over time, more thinly as $t \to \infty$. Interpolation yields that

$$\|T_t f\|_{L_x^{p*}} \lesssim |t|^{-d(1/p-1/2)} \|f\|_{L_x^p}$$

for $1 \leqslant p \leqslant 2$. These are the complete fixed-time $L^p \to L^q$ estimates that are available for the Schrödinger equations.

Theorem 45.1. *If we have $L^p \to L^q$ bounds of the form*

$$\|T_t f\|_{L_x^q} \leqslant A(t)\|f\|_{L_x^p}$$

then we have $q = p^*$, we have $p \leqslant 1$, and we have $A(t) \sim |t|^{-(d/2)(1/p-1/q)}$.

Proof. We first note that by Littlewood's Principle and the translation invariance of the Schrödinger equation"that the only $L^p \to L^q$ estimates for T_t must have $q \geqslant p$. If we can show $q = p^*$, this implies that $p \leqslant 1$ is necessary given the other equation $q = p^*$. Next, we use the *parabolic rescaling* of the Schrödinger equation, namely that

$$T_t\{\text{Dil}_R f\} = \text{Dil}_R\{T_{t/R^2} f\}.$$

It follows that if we had an $L^p \to L^q$ estimate for T_t with operator norm $A(t)$, then

$$R^{d/q}\|T_{t/R^2} f\|_{L_x^q} = \|T_t\{\text{Dil}_R f\}\|_{L_x^q} \leqslant A(t)\|\text{Dil}_R f\|_{L_x^p} = A(t)R^{d/p}\|f\|_{L_x^p}$$

Since f was arbitrary, we obtain that *if* we have $L^p \to L^q$ bounds, then $A(t/R^2) \leqslant A(t)R^{d(1/p-1/q)}$. Thus we conclude that $A(t) \sim |t|^{-(d/2)(1/p-1/q)}$. If $f \in \mathcal{S}_x$ has Fourier support in the unit ball, then the principle of non-stationary phase implies that for $|x| \geqslant 5|t|$,

$$|T_t f(x)| \lesssim_{N,M} \langle x \rangle^{-N} \langle t \rangle^{-M},$$

and for $|x| \leqslant 5|t|$, the principle of stationary phase implies that

$$|T_t f(x)| \sim \langle t \rangle^{-d/2}.$$

Thus $\|T_t f\|_{L_x^q} \sim \langle t \rangle^{d/q-d/2}$ for $|t| \geqslant 1$. Parabolic rescaling actually implies that for $|t| \leqslant 1$, $\|T_t f\|_{L_x^q} \sim |t|^{d/q-d/2}$. It thus follows that

$$|t|^{d/q-d/2} \lesssim |t|^{-(d/2)(1/p-1/q)}.$$

Thus we have $1/p + 1/q = 1$. $\qquad\qquad\square$

One can insert derivative Fourier multipliers into the estimates we have obtained to conclude that

$$\|T_t f\|_{W_x^{s,p*}} \lesssim t^{-d(1/p-1/2)}\|f\|_{W_x^{s,p}}$$

in the same range. Sobolev embedding allows one to obtain some higher integrability properties of the left hand side, i.e. that if we have p and q, with $q > p^*$, and $s = d(1/p^* - 1/q)$, then

$$\|T_t f\|_{L_x^q} \lesssim \|T_t f\|_{W_x^{s,p*}} \lesssim t^{-d(1/p-1/2)}\|f\|_{W_x^{s,p}}.$$

In particular, if f has Fourier support on $|\xi| \sim R$, then we conclude that

$$\|T_t f\|_{L_x^q} \lesssim t^{-d(1/p-1/2)}\|f\|_{W_x^{s,p}} \lesssim R^{d(1/p^*-1/q)} t^{-d(1/p-1/2)}\|f\|_{L_x^p}.$$

Thus we have frequency localized dispersive estimates. However, one can never *gain regularity* by applying the Schrödinger equation, contrasting the elliptic case where functions are smoothed by propogators.

Lemma 45.2. *There exists no estimates of the form*

$$\|T_t f\|_{W_x^{s_2,q}} \lesssim \|f\|_{W_x^{s_1,p}}$$

with $s_2 > s_1$.

Proof. We rely on the Galilean invariance of the Schrödinger equation, namely that

$$T_t\{\mathrm{Mod}_v f\} = e^{it|v|^2/2}\mathrm{Mod}_v\mathrm{Trans}_{vt}\{T_t f\}.$$

For $|v| \geqslant 1$ we have $\|\mathrm{Mod}_v f\|_{W_x^{s_1,p}} \sim |v|^{s_1}$. On the other hand, we have

$$\|T_t\{\mathrm{Mod}_v f\}\|_{W_x^{s_2,q}} \sim |v|^{s_2}.$$

Thus if we had a $W_x^{s_1,p} \to W_x^{s_2,q}$ bound, we have

$$|v|^{s_2} \sim \|T_t\{\mathrm{Mod}_v f\}\|_{W_x^{s_2,q}} \lesssim \|\mathrm{Mod}_v f\|_{W_x^{s_1,q}} \sim |v|^{s_1}.$$

Thus $|v|^{s_2} \lesssim |v|^{s_1}$, so as we take $|v| \to \infty$ we must have $s_2 \leqslant s_1$. $\qquad\square$

Combining these estimates with a duality argument, we obtain Strichartz time-space estimates for the Schrödinger equation, that allow us to obtain good bounds on the behaviour of the equation at all times, rather than just at a fixed, large time. To state these results, we thus define the operator

$$Tf(x,t) = T_t f(x).$$

This is precisely the *solution operator* for the Schrödinger equation.

Theorem 45.3. *If* $2/q + d/r = d/2$, *for* $2 \leqslant r, q \leqslant \infty$, *then*

$$\|Tf\|_{L_t^q L_x^r} \lesssim \|f\|_{L_x^2}.$$

This immediately implies that

$$\|Tf\|_{L_t^q W_x^{s,r}} \lesssim \|f\|_{H_x^s}.$$

Proof. The formal adjoint T^* of T maps functions on $\mathbb{R}^d \times \mathbb{R}$ to functions on $\mathbb{R}^d$ by the equation

$$T^*F(x) = \int e^{-2\pi itL}F(x,t)\,dt = \int T_{-t}F(x,t)\,dt.$$

719

Thus
$$TT^*F(x,t) = \int e^{2\pi i(t-s)L} F(x,s) \, ds = \int T_{t-s} F_s(x) \, ds.$$

We have the dispersive estimate
$$\|e^{2\pi i(t-s)} f\|_{L_x^r} \lesssim |t-s|^{-(d/2-d/r)} \|f\|_{L_x^{r*}}$$

for $1/r + 1/r^* = 1$. Thus
$$\|TT^*F\|_{L_t^q L_x^r} \leqslant \left\| \int \|T_{t-s} F_s\|_{L_x^r} \, ds \right\|_{L_t^q} \lesssim \left\| \int \frac{\|F_s\|_{L_x^{r*}}}{|t-s|^{d/2-d/r}} \, ds \right\|_{L_t^q}.$$

Now applying Hardy-Littlewood-Sobolev gives the required bound (*provided that* $q \neq 2$), i.e. that

$$\|TT^*F\|_{L_t^q L_x^r} \lesssim \left\| \int \frac{\|F_s\|_{L_x^{r*}}}{|t-s|^{d/2-d/r}} \, ds \right\|_{L_t^q} \lesssim \|F_s\|_{L_s^{q*} L_x^{r*}}.$$

The TT^* argument thus gives $\|Tf\|_{L_t^q L_x^r} \lesssim \|f\|_{L_x^2}$. For $q = 2$, we must instead apply more sophisticated techniques due to Keel and Tao (1998). □

We can view this equation in two ways. Firstly, the equation says that 'on average', the Schrödinger propogators result in an increase of integrability of f, from the L^2 norm to the L^r norm for $r > 2$, *but only by averaging in the L^q norm.* As an example application, let u_0 be a function with L^2 norm 1, and with Fourier transform supported on frequencies $|\xi| \sim R$. The uncertainty principle implies u_0 is locally constant at a scale $1/R$, and thus we expect that u_0 has size at most $R^{d/2}$ on any of these balls. The propogated solutions $T_t u_0$ also have L^2 norm 1, and have Fourier transform supported on frequencies $|\xi| \sim 1$. However, the Strichartz estimate thus implies that $T_t u_0$ can only have points with size $R^{d/2}$ on a small set of times with measure at most $O(1/R^2)$. This makes sense, since if the function has height $R^{d/2}$, it's mass would be highly concentrated near this point, and so we would have a spatial uncertainty of $1/R$. This implies a frequency uncertainty of R, and the dispersion relation implies that concentration in such a small ball only occurs for a time at most $1/R^2$ before the frequencies begin to separate from one another.

Theorem 45.4. *s*

720

Part VI

Geometry of Eigenfunctions

In these parts of the notes, we study *eigenfunctions* for the Laplacian on compact manifolds and bounded domains. This is a very classical problem, and it is where many of the classical techniques of harmonic analysis began. For instance, let us consider the study of the wave and heat equations

$$\left(\frac{\partial^2}{\partial t^2} - \Delta\right) u = 0 \quad \text{and} \quad \left(\frac{\partial}{\partial t} - \Delta\right) u = 0$$

on a vibrating membrane Ω fixed at the boundary, i.e. with $u(x, t) = 0$ for $x \in \partial\Omega$. If φ_λ solves the *Helmholtz equation* $\Delta\varphi_\lambda = -\lambda^2\varphi_\lambda$, vanishing on $\partial\Omega$, then a family of solutions to the wave equation and heat equations are given by

$$u_\lambda(t, x) = \cos(\lambda t + \phi)\varphi_\lambda(x) \quad \text{and} \quad u_\lambda(t, x) = e^{-t\lambda}\varphi_\lambda(x).$$

In this setting, the functions φ_λ are referred to as the *modes* of the wave equation, giving the wave profiles under which the resulting functions u_λ oscillate. Studying how these eigenfunctions interact with one another allows us to understand more general solutions to the wave equation.

Another application of eigenfunction problems occurs in the study of the Schrödinger equation

$$\left(-\frac{\hbar^2}{2}\Delta + V\right) \psi = E\psi,$$

which describes the evolution of a wave function ψ under a potential V. The case $V = 0$ and $E = E_0$ a constant value (the free Schrödinger equation) is equivalent to the Helmholtz equation, i.e. with λ proportional to $E_0/\hbar$. The values E_0 for which a solution exists are the *energy states* of the system. A useful physical intuition is that high energy states behave 'approximately classically', i.e. behaving like the classical dynamics of the Hamiltonian system given by the equation $H(x, \xi) = V(x) + |\xi|^2/2$. Thus the study of eigenfunctions is closely related to *semiclassical analysis*, i.e. understanding the relation between classical and quantum mechanics as the scale under which a problem is being studied changes.

There are various properties one can study of the eigenfunctions $\{\varphi_\lambda\}$. For instance, one could study their L^p norms, or other norms. One could also study the geometric properties of TODO

Chapter 46

The Laplace Beltrami Operator

Let

Chapter 47

Self-Adjoint, Elliptic Operators on a Compact Manifold

Let M^d be a compact manifold. Given a scalar density $d\omega$ on a manifold M, we can consider $L^2(M)$ as a Hilbert space induced by the inner product

$$\langle u, v \rangle = \int u(x)\overline{v(x)} \, d\omega.$$

Thus given a scalar density we can consider the adjoint of a pseudodifferential operator on M by considering the operator as an unbounded operator on $L^2(M)$ with domain $H_c^k(M)$, and in particular, we can consider *self-adjoint unbounded* operators on M; we are being somewhat lax with an overloaded set of terminology here, by a *self-adjoint operator* we mean an operator from $\mathcal{D}(M)$ to $\mathcal{D}(M)^*$, which is formally self adjoint on $\mathcal{D}(M)$, i.e. such that for all f, g *in* $\mathcal{D}(M)$,

$$\int_M Tf \cdot \overline{g} \, d\omega = \int_M f \cdot \overline{Tg} \, d\omega.$$

By a *self-adjoint unbounded operator*, we mean an operator from $\mathcal{D}(M)$ to $\mathcal{D}(M)^*$ which is essentially self-adjoint as an unbounded operator on $L^2(M)$, in the sense of the theory of self-adjointness. This is slightly stronger than being a self-adjoint operator, though we will soon see these properties are equivalent for an elliptic pseudodifferential operator. We now analyze the theory of self-adjoint, elliptic pseudodifferential operators on a compact manifold, in particular, proving the existence of the eigenfunction expansions of such operators. We begin by showing that such operators are automatically self-adjoint in the sense of an unbounded operator.

Theorem 47.1. *Suppose T is a self-adjoint, elliptic pseudodifferential operator of order t on a compact manifold M. Then T is a closed, self-adjoint unbounded operator.*

Proof. Let us begin by showing T is closed. We begin by noting that T is a continuous operator from $\mathcal{D}(M)^*$ to itself. We desire to show

$$\{(u,v) \in H^t(M) \times L^2(M) : Tu = v\}$$

is closed in $L^2(M) \times L^2(M)$, and so T is a closed operator on $L^2(M)$ when viewed with domain $H^t(M)$. If $\{u_i\}$ is a sequence in $H^t(M)$ converging in the $L^2(M)$ norm to some $u \in L^2(M)$, and $\{Tu_i\}$ converges in the $L^2(M)$ norm to some $v \in L^2(M)$, then $\{u_i\}$ converges distributionally to u, and so $Tu = v$ by the distributional continuity of T. Applying elliptic regularity, since $v \in L^2(M)$, we conclude that $u \in H^t(M)$. Thus we have shown the set above is closed.

Next, we show that T is a self-adjoint unbounded operator. Because T is continuous from $H^t(M)$ to $L^2(M)$, it is simple to see via a density argument that on $L^2(M)$, we have

$$\langle Tu, v \rangle = \langle u, Tv \rangle.$$

Thus to show T is self-adjoint as an unbounded operator, it suffices to show that if $v \in L^2(M)$ induces an inequality of the form

$$|\langle Tu, v \rangle| \lesssim \|u\|_{L^2(M)}$$

for all $u \in H^t(M)$, then $v \in H^t(M)$. But this again follows by elliptic regularity. Indeed, the inequality then holds for $u \in C^\infty(M)$, and the distributional theory thus implies that for such u,

$$\langle Tu, v \rangle = \langle u, Tv \rangle.$$

Thus we have an inequality of the form

$$|\langle u, Tv \rangle| \lesssim \|u\|_{L^2(M)}$$

for all $u \in C^\infty(M)$, and a density argument thus shows that $Tv \in L^2(M)$, so that $v \in H^t(M)$. $\square$

Now let T be a classical, self-adjoint pseudodifferential operator of order t. It follows from the calculus of pseudodifferential operators that it's principal symbol is real-valued. If T is self-adjoint and also *elliptic*, and M is connected and has

dimension bigger than one, it follows that the principal symbol is either positive or negative. By swapping T out with it's negation, we might as well assume that the principal symbol is positive. We will begin by consider the case $t = 1$, though the general case will reduce to this question. Given that the principal symbol is positive, we might imagine that T is 'positive-definite' up to first order, an intuition which leads to the following lemma.

Lemma 47.2. *Let T be a classical self-adjoint elliptic pseudodifferential operator of order 1 on a manifold M. Then there exists a classical self-adjoint elliptic pseudodifferential operator S of order $1/2$ such that $T - S^*S$ is a smoothing operator.*

Proof. We proceed as in the proof that $\sqrt{-\Delta}$ is a classical pseudodifferential operator. Let S_0 be a classical self-adjoint pseudodifferential operator of order $1/2$ with principal symbol $b_0(x, \xi) = a(x, \xi)^{1/2}$, then $S_0^*S_0$ is positive-semidefinite, and $T - S_0^*S_0$ is a classical pseudodifferential operator of order zero. Let r_1 denote the principal symbol of this operator. More generally, let us suppose that $T - (S_0 + \cdots + S_n)^*(S_0 + \cdots + S_n)$ is a classical pseudodifferential operator of order $-n$, with principal symbol r_n. If we consider a classical psuedodifferential operator S_{n+1} of order $1/2 - (n + 1)$ with principal symbol $b_{n+1}(x, \xi) = r_n(x, \xi)/2b_0(x, \xi)$, then $T - (S_0 + \cdots + S_{n+1})^*(S_0 + \cdots + S_{n+1})$ is a classical pseudodifferential operator of order $-(n + 1)$. Thus if we choose a self-adjoint operator S of order $1/2$ such that $S \sim \sum_{i=0}^{\infty} S_i$, then $T - S^*S$ is smoothing. $\qquad\square$

Lemma 47.3. *Suppose T is a classical self-adjoint elliptic pseudodifferential operator of order 1 with a non-negative principal symbol on a manifold M. For any compact subset K of M, there exists $\lambda_0 > 0$ such that for $\lambda \geqslant \lambda$, and $u \in H^1(M)$ supported on K,*
$$\langle (T + \lambda_0)u, u \rangle \sim \|u\|_{H^{1/2}(M)}^2.$$
In particular, if M is compact, one can do this anlaysis globally.

Proof. Without loss of generality, we may assume that the support of the kernel of T is compact. Choose a classical, self-adjoint, elliptic pseudodifferential operator S of order $1/2$ such that $T - S^*S$ is smoothing. Then for $u \in H_c^1(M)$, there exists $\gamma_0 > 0$ such that
$$\left| \langle Tu, u \rangle - \|Su\|_{L^2(M)}^2 \right| = \left| \int_M (T - S^*S)u(x)\overline{u}(x) \, dx \right| \leqslant \gamma_0 \|u\|_{L^2(M)}^2.$$

Thus if $\gamma > \gamma_0$,

$$\langle (T + \gamma)u, u \rangle = \langle Tu, u \rangle + \gamma \|u\|_{L^2(M)} \geqslant \|Su\|^2_{L^2(M)} + (\gamma - \gamma_0) \|u\|_{L^2(M)}.$$

Thus we conclude that $T + \gamma$ is positive definite. Since S has principal symbol $a(x, \xi)^{1/2}$, it is elliptic, and so for $u \in H^{1/2}(M)$, if $\gamma > \gamma_0$.

$$\begin{aligned}
\|u\|^2_{H^{1/2}(M)} &\lesssim \|Su\|^2_{L^2(M)} + \|u\|^2_{L^2(M)} \\
&\lesssim |\langle Tu, u \rangle| + \|u\|^2_{L^2(M)} \\
&= |\langle (T + \gamma)u, u \rangle - \langle \gamma u, u \rangle| + \|u\|^2_{L^2(M)} \\
&\leqslant \langle (T + \gamma)u, u \rangle + \langle \gamma u, u \rangle + \|u\|^2_{L^2(M)} \\
&\leqslant \langle (T + 2\gamma + 1)u, u \rangle.
\end{aligned}$$

Thus if $\lambda_0 > 2\gamma_0 + 1$, then

$$\langle (T + \lambda_0)u, u \rangle \lesssim \|Su\|_{L^2(M)} + \|u\|_{L^2(M)} \lesssim \|u\|_{H^{1/2}(M)}. \qquad \square$$

Theorem 47.4. *Suppose T is a classical, self-adjoint pseudodifferential operator of order one on a compact manifold M, with a non-negative principal symbol. Then for suitably large $\lambda_0 > 0$,*

- *For all $k > 0$, $(T + \lambda_0)^k$ is an isomorphism from $H^k(M)$ to $L^2(M)$.*

- *There exists an increasing family of real numbers $\{\lambda_i\}$ with $\lambda_i \to \infty$, and an orthogonal basis $\{e_i\}$ for $L^2(M)$ consisting solely of elements of $C^\infty(M)$, such that $Te_i = \lambda_i e_i$ for each $i > 0$. The functions $\{e_i\}$ satisfy*

$$\|e_i\|_{H^k(M)} \sim_k (\lambda_0 + \lambda_i)^k.$$

- *If $u \in H^k(M)$, then $\langle u, e_i \rangle \lesssim_k \|u\|_{H^k(M)}(\lambda_0 + \lambda_i)^{-k}.$*

Proof. The last result implies that $T + \lambda_0$ is an injective operator from $H^1(M)$ to $L^2(M)$ for sufficiently large $\lambda_0 > 0$. We have already seen that, since M is compact, $T + \lambda_0$ is a closed, self-adjoint unbounded operator on $L^2(M)$. But the theory of unbounded operators thus implies that because $T + \lambda_0$ is injective, it has dense image in $L^2(M)$. Thus we can define it's inverse $R = (T + \lambda_0)^{-1}$, which is a densely defined continuous operator in $L^2(M)$, and thus extends to a continuous operator on all of $L^2(M)$. Because $T + \lambda_0$ is positive-definite, i.e. $\langle (T + \lambda_0)u, u \rangle > 0$ for all non-zero $u \in H^1(M)$, density implies that $\langle Rv, v \rangle > 0$

727

for all $v \neq 0$. Thus R is positive definite. If $u \in H^1(M)$, and $(T + \lambda_0)u = v$, i.e. $u = Rv$, then ellipticity implies that

$$\|Rv\|_{H^1(M)} \lesssim \|v\|_{L^2(M)} + \|Rv\|_{L^2(M)} \lesssim \|v\|_{L^2(M)}.$$

Conversely, continuity of pseudodifferential operators implies that

$$\|v\|_{L^2(M)} = \|(T + \lambda_0)u\|_{L^2(M)} \lesssim \|u\|_{H^1(M)}.$$

Thus density implies that R is an *isomorphism* from $L^2(M)$ to $H^1(M)$. In particular we note that that $T + \lambda_0$ is an isomorphism from $H^1(M)$ to $L^2(M)$, i.e. it was surjective in the first place rather than merely having dense image. Moving on, The Rellich-Kondrachov theorem thus implies that R is a *compact*, positive definite operator on $L^2(M)$. Thus the spectral analysis of such operators implies that there exists an orthogonal basis $\{e_i\}$ and a family of positive, decreasing eigenvalues $\{\gamma_i\}$ such that if E_i is orthogonal projection onto the span of e_i, then $(T + \lambda_0)^{-1} = \sum_i \gamma_i^{-1} E_i$. If $\lambda_i = \gamma_i - \lambda_0$, then this means that $T = \sum_i \lambda_i E_i$.

Applying elliptic regularity, since $(T + \lambda_0)e_i = \gamma_i e_i$ lies in $L^2(M)$, we conclude that $e_i \in H^1(M)$. But iterating this argument gives that $e_i \in H^N(M)$ for any $N > 0$. Applying the Sobolev embedding theorem thus shows that $e_i \in C^\infty(M)$ for each i.

A simple generalization of the argument above (the spectral theory, ellipticity, and so on) shows that $(T + \lambda_0)^k$ is actually an isomorphism from $H^k(M)$ to $L^2(M)$ for each $k > 0$. We thus find that for any $u \in H^k(M)$,

$$\|u\|^2_{H^k(M)} \sim_k \|(T + \lambda_0)^k u\|^2_{L^2(M)} = \sum (\lambda_i + \lambda_0)^{2k} |\langle f, e_i \rangle|^2.$$

Thus we find that

$$|\langle u, e_i \rangle| \lesssim_k \|u\|_{H^k(M)} |\lambda_i + \lambda_0|^{-k} \lesssim_k \|u\|_{H^k(M)} |1 + \lambda_0|^{-k}. \qquad \square$$

The following, known as *Hörmander's square root trick*, allows us to extend the theory above to elliptic, self-adjoint operators of arbitrary order.

Theorem 47.5. *Fix integers p and q with $q \neq 0$, and set $t = p/q$. Suppose T is a classical, positive-definite, elliptic pseudodifferential operator of order t on a compact manifold M. Then the operator $T^{q/p}$, defined using the spectral functional calculus, is a classical, positive-definite, elliptic pseudodifferential operator of order 1, with principal symbol $a(x, \xi)^{q/p}$.*

Proof. The case $t = -1$ is easily seen to be true, since the assumptions imply T is invertible, and it's inverse is a pseudodifferential operator modulo smoothing. Thus, without loss of generality, we may assume that $q = 1$, and that $p > 0$. We can certainly *construct* an operator $S \in \Psi^1_{cl}(M)$ which is self-adjoint, with principal symbol $a(x, \xi)^{1/p}$, such that $S^p - T$ is a smoothing operator. Applying the theory above, we see that S has an eigenvalue decomposition, with eigenvectors lying in $C^\infty(M)$, and only finitely many eigenvalues being negative. Thus, modulo a smoothing operator, i.e. a finite sum of the projection operators $\{E_i\}$, we may actually replace S with a positive-definite operator. We claim that $S - T^{1/p}$ is also a smoothing operator, which would complete the proof since the principal symbol of S is equal to $a(x, \xi)^{q/p}$. Since S and T are both positive-definite, they are injective, and so by the spectral calculus of unbounded operators, T^{-1} and S^{-1} is a well-defined bounded, positive-definite operator on $L^2(M)$. If γ is a contour in the complex plane which winds around each element of $\sigma(T^{-1})$ exactly once, then we find using the holomorphic functional calculus that

$$T^{1/p} = \frac{1}{2\pi i} \int_\gamma z^{-1/p} (z - T^{-1})^{-1} \, dz$$

and

$$S = \frac{1}{2\pi i} \int_\gamma z^{-1/p} (z - S^{-p})^{-1} \, dz$$

Now S^{-p} is a pseudodifferential operator of order $-p$. Since $S^p - T$ is a smoothing operator, the composition calculus tells us that so is $T^{-1}(S^p - T)S^{-p} = T^{-1} - S^{-p}$. Write $S^{-p} = T^{-1} - R$. Then

$$\begin{aligned}
S - T^{1/p} &= \frac{1}{2\pi i} \int_\gamma z^{-1/p} \left((z - T^{-1} + R)^{-1} - (z - T^{-1})^{-1} \right) \\
&= \frac{1}{2\pi i} \int_\gamma z^{-1/p} \left((z - T^{-1} + R)^{-1} R (z - T^{-1})^{-1} \right).
\end{aligned}$$

This is a integral over a compact curve over a continuous family of smoothing operators, which is therefore a smoothing operator. $\qquad\square$

Remark. As above, finding an operator S of order $t/2$ such that $T - S^*S$ is smoothing allows one to show that if T is a classical, elliptic, self-adjoint pseudodifferential operator of order t, then $T + \lambda_0$ is positive definite for suitably large $\lambda_0 > 0$, so that we may apply the theorem above to a wider situation.

Using the trick, we quickly extend the theory above to general operators of this form.

Theorem 47.6. *Suppose T is a classical, self-adjoint, elliptic pseudodifferential operator of order $t = p/q$ on a compact manifold M. Then for sufficiently large $\lambda_0 > 0$,*

- *$(T + \lambda_0)^k$ is an isomorphism from $H^{tk}(M)$ to $L^2(M)$ for all $k > 0$.*

- *There is a basis $\{e_i\}$ for $L^2(M)$ consisting of elements of $C^\infty(M)$ and an increasing sequence of positive real numbers $\{\lambda_i\}$ with $\lambda_i \to \infty$, such that $Te_i = (\lambda_0 + \lambda_i)^t e_i$. The functions $\{e_i\}$ satisfy*

$$\|e_i\|_{H^k(M)} \sim_k (\lambda_0 + |\lambda_i|)^k$$

- *If $u \in H^k(M)$, then $|\langle u, e_i \rangle| \lesssim_k \|u\|_{H^k(M)} (\lambda_0 + |\lambda_i|)^{-k}$.*

The special case of this result, where Δ is the Laplace-Beltrami operator on a compact Riemannian manifold, was first conjectured by Ohm and Raleigh, and proved by Hilbert in 1904.

Example. *Let M be a compact Riemannian manifold. Then the Laplace-Beltrami operator Δ is a classical pseudodifferential operator of order two on M, and $-\Delta$ is formally positive-semidefinite. Thus we can see λ_0 above to be any positive number, which gives that there exists a discrete subset $\Lambda(M)$ of $[0, \infty)$, and a family of finite dimensional subspaces E_λ of $C^\infty(M)$ for each $\lambda \in \Lambda(M)$, such that $\Delta e_\lambda = -\lambda^2 e_\lambda$ for each $e_\lambda \in E_\lambda$. If $u \in H^k(\mathbb{T}^n)$ and $e_\lambda \in E_\lambda$, then*

$$|\langle u, e_\lambda \rangle| \lesssim_{\varepsilon, k} \|e_\lambda\|_{L^2(M)} \|u\|_{H^k(M)} (\varepsilon + \lambda_i)^{-k}$$

On $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$, we can compute this expansion explicitly. The space E_λ consists of the span of any functions of the form $e_\xi(x) = e^{2\pi i \xi \cdot x}$, where $\xi \in \mathbb{Z}^n$, and $\lambda = 2\pi|\xi|$. In this case the Fourier transform explicitly tells us the slightly stronger identity

$$|\langle u, e_\xi \rangle| = |\hat{u}(\xi)| \lesssim_k \min\left(\|u\|_{L^1(\mathbb{T}^n)}, \|D^k u\|_{L^1(\mathbb{T}^n)} |\xi|^{-k}\right).$$

Chapter 48

Distribution of Eigenvalues and the Half Wave Operator

Consider the setup we began our study of in the last section. We have a classical, self-adjoint, elliptic pseudodifferential operator T of order one on a compact manifold M with principal symbol $p(x, \xi) \geqslant 0$, and we have associated with T a sequence of orthonormal smooth functions $\{e_i\}$ in $C^\infty(M)$ which form a basis for $L^2(M)$, and an increasing sequence $\{\lambda_i\}$ tending to ∞, such that $T e_i = \lambda_i e_i$. Our goal now is to further the study of the behaviour of the eigenfunctions $\{e_i\}$, and the distributions of the eigenvalues $\{\lambda_i\}$. A key object in the study of the distribution of the eigenvalues is the function $N(\lambda)$, which gives the number of eigenvalues of T less than or equal to λ. Our results already give a very basic estimate for N, which we will considerably sharpen over the course of this chapter.

Theorem 48.1. *The function N is tempered in λ. More precisely, for $\lambda > 0$,*

$$N(\lambda) \lesssim_\varepsilon \lambda^{n+\varepsilon}.$$

Proof. Define the operators S_λ, which are the projections in $L^2(M)$ onto the eigenspaces corresponding to eigenvalues with value $\leqslant \lambda$. Then S_λ clearly has kernel

$$K_\lambda(x, y) = \sum_{\lambda_i \leqslant \lambda} e_i(x) \overline{e_i(y)}.$$

This function is smooth, and

$$\int K_\lambda(x, x) \, dx = \sum_{\lambda_i \leqslant \lambda} |e_i(x)|^2 = N(\lambda).$$

731

Thus $N(\lambda)$ is the trace of the operator S_λ. Because $\|e_i\|_{H^k(M)} \lesssim_k (1 + |\lambda|)^k$ if $\lambda_i \leqslant \lambda$, we have the elementary estimate

$$\|S_\lambda f\|_{H^k(M)} \lesssim_k \lambda^k \|f\|_{L^2(M)}.$$

The Sobolev embedding theorem thus implies that if $\dim(M) = d$, and if $\varepsilon > 0$, then

$$\|S_\lambda f\|_{L^\infty(M)} \lesssim_\varepsilon \|S_\lambda f\|_{H^{d/2+\varepsilon}(M)} \lesssim \lambda^{d/2+\varepsilon} \|f\|_{L^2(M)}.$$

Thus we conclude that, for any $f \in C^\infty(M)$, and any $x \in M$,

$$\left| \int K_\lambda(x, y) f(y) \, dy \right| \lesssim_\varepsilon \lambda^{d/2+\varepsilon} \|f\|_{L^2(M)}.$$

By density, this implies that if $g_{\lambda,x}(y) = K_\lambda(x, y)$, then

$$\|g_{\lambda,x}\|_{L^2(M)} \lesssim_\varepsilon \lambda^{d/2+\varepsilon}.$$

But $g_{\lambda,x} \in \mathrm{span}\{e_i : \lambda_i \leqslant \lambda\}$, so this means that

$$\|g_{\lambda,x}\|_{L^\infty(M)} = \|S_\lambda g_{\lambda,x}\|_{L^\infty(M)} \lesssim_\varepsilon \lambda^{d/2+\varepsilon} \|g_{\lambda,x}\|_{L^2(M)} \lesssim_\varepsilon \lambda^{d+2\varepsilon}.$$

But this means we conclude that $\|K_\lambda\|_{L^\infty(M \times M)} \lesssim_\varepsilon \lambda^{d+\varepsilon}$, and thus that

$$N(\lambda) = \int K_\lambda(x, x) \, dx \lesssim_\varepsilon \lambda^{d+\varepsilon}. \qquad \square$$

Our next goal is now to establish the *Sharp Weyl formula*. If

$$c = \int_{\{(x,\xi) \in T^*M : a(x,\xi) \leqslant 1\}} d\xi \, dx,$$

then the result states that

$$N(\lambda) = c\dot{\lambda}^n + O(\lambda^{n-1}).$$

We will continue to use the representation $N(\lambda) = \mathrm{Tr}(S_\lambda)$. However, a key step here is to recognize that S_λ is a *function* of T, namely $S_\lambda = \chi_\lambda(T)$, where $\chi_\lambda(\xi) = \mathbf{I}(\xi \leqslant \lambda)$. The next trick is to use the Fourier multiplication formula to obtain a

useful representation of the structure of S_λ. For $f \in C^\infty(M)$, we can write

$$
\begin{aligned}
S_\lambda f &= \int_{-\infty}^\infty \chi_\lambda(\tau) \left(\sum_i \delta_{\lambda_i}(\tau) E_i f \right) d\tau \\
&= \int_{-\infty}^\infty \hat{\chi}_\lambda(t) \left(\sum_i e^{2\pi i t \lambda_i} E_i f \right) dt \\
&= \int_{-\infty}^\infty \hat{\chi}_\lambda(t) e^{2\pi i t T} dt \\
&= \frac{1}{2\pi i} \int_{-\infty}^\infty \frac{e^{2\pi i t (T-\lambda)} f}{t + i0} dt.
\end{aligned}
$$

Thus we are lead to study the *half wave operator* $\partial_t - 2\pi i T$, since one sees that $(\partial_t - 2\pi i T)\{e^{2\pi i t T} f\} = 0$ for all $f \in L^2(M)$; it is easily verified for the eigenfunctions of T, and then the result follows by the distributional continuity of all operations involved.

48.1   Construction of the Lax Parametrix

We will understand the operators $e^{2\pi i t T}$ by constructing a *parametrix* for the equation $\partial_t - 2\pi i T$ over small times, i.e. for $|t| \lesssim 1$. The equation $\partial_t = 2\pi i T$ is a pseudodifferential variant of a basic hyperbolic partial differential equation, so based on the type of parametrices one can construct in the hyperbolic setting, one might hope to find a parametrix defined by an oscillatory integral of the form

$$
S f(x, t) = S(t) f(x) = \int s(t, x, y, \xi) e^{2\pi i \Phi(t, x, y, \xi)} f(y) \, dy \, d\xi,
$$

such that:

- $\Phi(t, x, y, \xi) = \phi(x, y, \xi) + t p(y, \xi)$, where ϕ is smooth away from $\xi = 0$, homogeneous of degree one, and $\phi(x, y, \xi) \approx (x - y) \cdot \xi$ in the sense that on the support of s,

$$
\partial_\xi^\beta \{\phi(x, y, \xi) - (x - y) \cdot \xi\} \lesssim_\beta |x - y|^2 |\xi|^{1-\beta}.
$$

  In particular, this implies that $\phi(x, y, \xi) = 0$ when $(x - y) \cdot \xi = 0$.

733

- s is a symbol of order zero, supported on $|x - y| \lesssim 1$ and on $|\xi| \geqslant 1$, in such a way that

$$|\nabla_\xi \phi(x, y, \xi)| \gtrsim |x - y| \quad \text{and} \quad |\nabla_x \phi(x, y, \xi)| \gtrsim |\xi|$$

  for $(x, y) \in \operatorname{supp}_x(s) \times \operatorname{supp}_y(s)$.

If $(\partial_t - 2\pi i T) \circ S = 0$ is a smoothing operator on $(-\varepsilon, \varepsilon) \times M$ and $S(0)$ differs from the identity operator by a smoothing operator, then $S - e^{2\pi i t T}$ is smoothing for $|t| \leqslant \varepsilon$. The construction of the parametrix S will therefore give us much more information about the behaviour of the propogators $e^{2\pi i t T}$ over small times.

To find a choice of ϕ and s which gives us this parametrix, let us start by determining what properties these functions should satisfy. Let us fix a coordinate system (x, U), where $x(U)$ is a precompact subset of $\mathbb{R}^n$. Let us assume that in these coordinates, T has symbol $a(x, \xi)$. Then the kernel of $(\partial_t - 2\pi i T) \circ S$ in this coordinate system is

$$\int (\partial_t + 2\pi i T(x, D)) \left\{ s(t, \cdot, y, \xi) e^{2\pi i \Phi(t, \cdot, y, \xi)} \right\} \, d\xi.$$

If we set

$$s'(t, x, y, \xi) = e^{-2\pi i \Phi(t, x, y, \xi)} (\partial_t - 2\pi i T(x, D)) \left\{ s(t, \cdot, y, \xi) e^{2\pi i \Phi(t, \cdot, y, \xi)} \right\}$$

then the kernel is

$$\int s'(t, x, y, \xi) e^{2\pi i \Phi(t, x, y, \xi)} \, d\xi.$$

Provided that s' is a symbol of order $-\infty$ for $0 < |t| \leqslant \varepsilon$, integration by parts shows that $(\partial_t + 2\pi i T) \circ S$ is smoothing, and so we will try to choose ϕ and s so as to obtain such a result.

In our discussion of pseudodifferential operators, we have already discussed an asymptotic formula for s', namely, if

$$r_{x,y}(z) = \nabla_x \phi(x, z, \xi) \cdot (x - z) - \{\phi(x, y, \xi) - \phi(z, y, \xi)\}.$$

then for any $N > 0$, if $a \sim \sum_{k=-\infty}^{1} a_k$, where a_k is homogeneous of degree k, and

if $\xi_\phi = \nabla_x \Phi(t, x, y, \xi) = \nabla_x \phi(x, y, \xi)$,

$$s'(t, x, y, \xi)$$
$$= \underbrace{(p(y, \xi) - a(x, \xi_\phi)) \cdot s(t, x, y, \xi)}_{\text{symbols of order 1}}$$
$$+ \underbrace{\partial_t s(t, x, y, \xi)}_{\text{symbol of order 0}}$$
$$- \sum_{1 \leqslant |\beta| < N} \frac{2\pi i}{\beta! \cdot (2\pi i)^\beta} \cdot \underbrace{\partial_\xi^\beta a(x, \xi_\phi) \partial_z^\beta \{e^{2\pi i r_{x,y}(z)} s(t, z, y, \xi)\}|_{z=y}}_{\text{symbols of order } 1 - \lceil |\beta|/2 \rceil}$$
$$+ R_N(t, x, y, \xi).$$

where, because $|\nabla_x \Phi(t, x, y, \xi)| \gtrsim |\xi|$ on the support of s,

$$\langle \xi \rangle^{t - \lceil N/2 \rceil} R_N \in L^\infty((-\varepsilon, \varepsilon) \times U \times U \times \mathbb{R}^d).$$

It is simple to establish estimates of the form

$$|\partial_x^\alpha \partial_y^\beta \partial_\xi^\lambda s'(t, x, y, \xi)| \lesssim \langle \xi \rangle^{N_{\alpha\beta\lambda}}.$$

Thus if we can justify that $|s'(t, x, y, \xi)| \lesssim_N \langle \xi \rangle^{-N}$ for all $N > 0$, then it will follow that s' is a symbol of order $-\infty$. We now determine the propoerties of the symbol s and the symbol ϕ which will give us these estimates.

To begin with, let us specify the function ϕ. In order to guarantee that s' is a symbol of order zero, the expansion above shows that $(p(y, \xi) - p(x, \xi_\phi)) \cdot s(t, x, y, \xi)$ must be a symbol of order zero. This will be true if we can pick ϕ such that, on the support of s, and for $|\xi| \gtrsim 1$,

$$p(x, \nabla_x \phi(x, y, \xi)) = p(y, \xi).$$

This is an example of an *Eikonal equation*, e.g. an equation of the form

$$q(z, \nabla_z \psi(z)) = 0$$

for some function $q(z, \zeta)$. In our case, $z = (x, y, \xi)$, so $\zeta = (\zeta_x, \zeta_y, \zeta_\xi)$, and so

$$q(z, \zeta) = p(x, \zeta_x) - p(y, \xi).$$

Let us make some further remarks we desire about our choice of function ϕ:

735

- We want ϕ to be homogeneous and smooth away from the origin. If we solve the equation for all $|\xi| = 1$, and then extend ϕ such that for $\lambda > 0$ and $|\xi| = 1$,

$$\phi(x, y, \lambda\xi) = \lambda\phi(x, y, \xi)\psi(\lambda),$$

where ψ is smooth, equal to one for $|\lambda| \geqslant 3/4$, and vanishing for $|\lambda| \leqslant 1/2$, then ϕ will satisfy the equation for all $|\xi| \gtrsim 1$. This means that

$$p(x, \nabla_x\phi(x, y, \xi)) - p(y, \xi)$$

is smooth and supported on $|\xi| \lesssim 1$, which implies it is a symbol of order $-\infty$, which suffices for our construction. Thus it suffices to solve the equation for $|\xi| = 1$.

- Since ϕ is smooth away from the origin and homogeneous, the equation

$$|\partial_\xi^\beta \{\phi(x, y, \xi) - (x - y) \cdot \xi\}| \lesssim_\beta |x - y|^2 \langle\xi\rangle^{1-\beta}$$

holds if, for $|\xi| = 1$, we have $\phi(x, y, \xi) = 0$ whenever $(x - y) \cdot \xi = 0$, and $\nabla_x\phi(x, y, \xi) = \xi$ whenever $x = y$. Thus we have some *initial conditions* for our Eikonal equation.

The second condition constitutes a type of initial condition for ϕ, since it specifies it's behaviour on a hypersurface, a kind of Cauchy condition, and thus we should expect these are close to the conditions that give unique solutions to the equation. And the following Lemma indeed shows that there is a unique function ϕ, defined for $|x - y| \lesssim 1$ and $|\xi| = 1$ with these properties.

Lemma 48.2. *Let Z be a smooth manifold, and let $q(z, \zeta)$ be a real-valued, smooth function defined locally around a point $(z_0, \zeta_0) \in T^*Z$. Let S be a smooth hypersurface in Z passing through z_0 with conormal vector ζ_S at z_0, such that*

$$\frac{\partial q}{\partial \zeta_S}(z_0, \zeta_0) = \lim_{t \to 0} \frac{q(z_0, \zeta_0 + t\zeta_S) - q(z_0, \zeta_0)}{t}$$

is nonzero. Suppose that ψ is any smooth function defined on S locally about z_0, such that $d\psi(z_0)$ agrees with the action of ζ_0 on $T_{x_0}S$. Then there exists a unique smooth function ϕ defined in a neighborhood of z_0, which agrees with ψ on S, satisfies the Eikonal equation $q(z, \nabla_z\phi(z)) = 0$, and has $\nabla_z\phi(z_0) = \zeta_0$.

Proof. TODO: See Sogge, Theorem 4.1.1. □

736

In our case,
$$Z = \{(x, y, \xi) : |\xi| = 1\}.$$
We have $z_0 = (x_0, x_0, \xi_0)$, $\zeta_0 = (\xi, \xi, 0)$, and

$$S = \{(x, y, \xi) : |\xi| = 1 \quad \text{and} \quad (x - y) \cdot \xi = 0\}.$$

The conormal vector ξ_S of S at z_0 is a multiple of $(\xi_0, -\xi_0, 0)$, and so by homogeneity,
$$\frac{\partial q}{\partial \xi_S} = \lim_{t \to 0} \frac{p(x_0, (1 + t)\xi_0) - p(x_0, \xi_0)}{t} = p(x_0, \xi_0),$$

which is nonvanishing because T is elliptic. If we define ψ equal to zero on S, then $d\psi = 0$, which agrees with the action of ζ_0 on S. Thus the theorem applies local uniqueness and existence to solutions to the Eikonal equation, and by compactness of Z we can patch such solutions together to find a solution defined for all $|x-y| \lesssim 1$.

We therefore conclude that there exists a unique choice of ϕ such that, if s has small enough support, $s'(t, x, y, \xi)$ is a symbol of order zero. Next, let us see what constraints are forced on us in order to ensure that $S(0)$ differs from the identity by a smoothing operator. The kernel of U is precisely

$$\int s(0, x, y, \xi) e^{2\pi i \phi(x, y, \xi)} \, d\xi.$$

We now show that this operator is actually a *pseudodifferential operator* of order zero, and determine it's symbol up to first order.

To do this, we write $\phi_\alpha(x, y, \xi) = (1 - \alpha)\phi(x, y, \xi) + \alpha(x - y) \cdot \xi$. Let U_α be the operator with kernel

$$\int s(0, x, y, \xi) e^{2\pi i \phi_\alpha(x, y, \xi)} \, d\xi.$$

Assume the support of s is close enough to the diagonal such that

$$|\nabla_\xi \phi_\alpha(x, y, \xi)| \gtrsim |x - y|$$

on the support of s. Then $\partial_\alpha^n U_t$ has kernel

$$\int (2\pi i)^n (\phi_1 - \phi_0)^n s(0, x, y, \xi) e^{2\pi i \phi_\alpha(x, y, \xi)} \, d\xi.$$

737

This is an oscillatory integral defined by a symbol of order n. However, when $t = 1$, the fact that $\phi(x, y, \xi) \approx (x - y) \cdot \xi$, together with the formula for converting pseudodifferential operators with compound symbols into standard Kohn-Nirenberg type symbols shows that $\partial_\alpha^n U_1$ is actually a pseudodifferential operator of order $-n$. Integration by parts, similarily, shows that $\partial_\alpha^n U_t$ is defined by an oscillator integral against a symbol of order $-n$. But this means that if we define a pseudodifferential operator by the asymptotic formula

$$V \sim \sum \frac{(-1)^n}{n!} \partial_\alpha^n U_1,$$

then $U - V$ is smoothing. Indeed, for any n, by Taylor's formula we have

$$U = \sum_{k=0}^{n-1} \frac{(-1)^n}{n!} \partial_\alpha^n U_1 + \frac{(-1)^n}{n!} \int_0^1 \alpha^{n-1} \partial_\alpha^n U_\alpha \, d\alpha$$

The integral here is an oscillatory integral defined against a symbol of order $-n$, and thus taking $n \to \infty$ verifies the claim.

It is an important remark that reversing this argument shows that *any* pseudodifferential operator can be written in the form above for the particular choice of ϕ we have given. This is a special case of the *equivalence of phase functions* theorem. This in particular guarantees that we can choose a symbol $I(x, y, \xi)$ of order zero such that $U - 1$ is smoothing if and only if $s(0, x, y, \xi) - I(x, y, \xi)$ is a symbol of order $-\infty$. The symbol I can be chosen to be vanishing for $|x - y| \gtrsim 1$, since the difference will be a smoothing pseudodifferential operator.

Next, the quantity

$$\partial_t s(t, x, y, \xi)$$

$$+ \sum_{k=1}^d \partial_\xi^k a(x, \xi_\phi) \partial_x^k s(t, x, y, \xi)$$

$$+ \left(a_0(x, \xi_\phi) + \frac{1}{2\pi} \sum_{|\beta|=2} \partial_\xi^\beta p(x, \xi_\phi) \partial_x^\beta \phi(x, y, \xi) \right) s(t, x, y, \xi).$$

must be a symbol of order -1. But because the coefficients of this equation are smooth, and all derivatives are bounded, it follows from the general theory of transport equations that there exists a unique smooth, function s_0 defined for $|t| \leqslant \varepsilon$, which is a symbol of order zero, such that $s_0(0, x, y, \xi) = I(x, y, \xi)$, s_0 vanishes

for $|x - y| \gtrsim 1$, and satisfies the transport equation

$$\partial_t s_0(t, x, y, \xi)$$

$$+ \sum_{k=1}^{d} \partial_\xi^k a(x, \xi_\phi) \partial_x^k s_0(t, x, y, \xi)$$

$$+ \left(a_0(x, \xi_\phi) + \frac{1}{2\pi} \sum_{|\beta|=2} \partial_\xi^\beta p(x, \xi_\phi) \partial_x^\beta \phi(x, y, \xi) \right) s_0(t, x, y, \xi) = 0.$$

We have thus justified that the quantity

$$R_0(t, x, y, \xi) = e^{-2\pi i \Phi(t,x,y,\xi)} (\partial_t - 2\pi i T) (s_0(t, \cdot, y, \xi) e^{2\pi i \Phi(t,x,y,\xi)})$$

is a symbol of order -1. Now we come to a quirk of this parametrix, which does not occur in the study of hyperbolic partial differential equations. Since the operator $P(x, D)$ is only *pseudolocal* rather than completely local, the remainder term R_0 is *not* necessarily supported on a neighborhood of the origin. To fix this, we now successively define the terms $\{s_k\}$ for $k < 0$, which are symbols of order $-k$, such that $s_k(0, x, y, \xi) = 0$, and

$$TODO : SPECIFYREQUIREDEQUATION.$$

Again, solutions exist for small time periods. And this implies that $e^{-2\pi i \Phi(t,x,y,\xi)} (\partial_t + 2\pi i T) ((s_0 + \cdots + s_{-k}) e^{2\pi i \Phi(t,x,y,\xi)})$ is a symbol of order $-k$ (TODO: Is It), and we can continue the calculation to complete the argument.

48.2 The Sharp Weyl Formula

Let us now use the Lax parametrix to prove the sharp Weyl formula, which immediately is given by the following, more precise estimate.

Theorem 48.3. *Let M be a compact manifold, and let T be a classical, self-adjoint, elliptic pseudodifferential operator of order one on M, with principal symbol $p(x, \xi)$. If S_λ is the projection operator onto the span of eigenfunctions with eigenvalue $\leqslant \lambda$, and K_λ is the kernel of the operator S_λ, then*

$$K_\lambda(x, x) = c(x)\lambda^n + O(\lambda^{n-1}),$$

where

$$c(x) = \int_{p(x,\xi) \leqslant 1} d\xi.$$

Proof. Recall that we have

$$S_\lambda f(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{2\pi i t(T-\lambda)} f}{t + i0} \, dt.$$

This integral only has a singularity when $t = 0$, so it makes sense to decompose the integral into two parts using an even function $\rho \in C_c^\infty(\mathbb{R})$ supported on $|t| \leqslant \varepsilon/2$ and equal to one for $|t| \leqslant \varepsilon/4$, i.e. writing

$$S_\lambda f = S_{\lambda,\text{low}} f + S_{\lambda,\text{high}} f,$$

where

$$S_{\lambda,\text{low}} f = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\rho(t)}{t + i0} e^{2\pi i t(T-\lambda)} f \, dt$$

and

$$S_{\lambda,\text{high}} f = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{1 - \rho(t)}{t} e^{2\pi i t(T-\lambda)} f \, dt.$$

To understand $S_{\lambda,\text{low}} f$, we use the parametrix for the wave equation. If we switch to coordinates around a particular point, then we find that the kernel of $S_{\lambda,\text{low}}$ is equal to $\tilde{K}_\lambda + R_\lambda$, where $\tilde{K}_\lambda$ is supported on a neighborhood of the diagonal, and locally in coordinates we can write

$$\tilde{K}_\lambda(x, y) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\rho(t)}{t + i0} s(t, x, y, \xi) e^{2\pi i (\phi(x,y,\xi) + t(p(y,\xi) - \lambda))} \, d\xi \, dt.$$

and where

$$R_\lambda(x, y) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\rho(t)}{t + i0} A(t, x, y) e^{2\pi i t\lambda} \, dt,$$

where A is smooth for $|t| \leqslant \varepsilon$ and for $x, y \in M$. But since $\rho \cdot A$ is smooth, with compact support in the t variable taking inverse Fourier transforms in the t variable shows that

$$|R_\lambda(x, y)| = \left| \int_{-\infty}^{\lambda} \mathcal{F}_t^{-1}\{\rho A\}(\lambda - \tau, x, y) \, d\tau \right| \lesssim 1,$$

By Taylor's formula, we can write $s(t, x, x, \xi) = s(0, x, x, \xi) + t \cdot r(t, x, \xi)$, where

740

r is a symbol of order zero. But if N is sufficiently large, we have

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\rho(t)}{t+i0}(t \cdot r(t,x,\xi))e^{2\pi it(p(x,\xi)-\lambda)} \, d\xi \, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \rho(t)r(t,x,\xi)e^{2\pi it(p(x,\xi)-\lambda)} \, d\xi \, dt$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}^n} \mathcal{F}_t^{-1}\{\rho \cdot r\}(p(x,\xi)-\lambda, x, \xi) \, d\xi$$

$$= \int O_N \left(1 + |p(x,\xi)-\lambda|\right)^{-N} \, d\xi = O(\lambda^{n-1}).$$

We also have $s(0, x, x, \xi) = 1 + s_{-1}(x, \xi)$, where s_{-1} is a symbol of order -1. Now if $\tilde{\chi}_\lambda = \chi_\lambda * \mathcal{F}_t^{-1}\{\rho\}$, then

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\rho(t)}{t+i0}s_{-1}(x,\xi)e^{2\pi it(p(x,\xi)-\lambda)} \, d\xi \, dt = \int_{\mathbb{R}^n} \tilde{\chi}_\lambda(p(x,\xi))s_{-1}(x,\xi) \, d\xi.$$

Then $\tilde{\chi}_\lambda(p(x,\xi)) \sim 1$ for $|\xi| \lesssim \lambda$, and for $|\xi| \gtrsim \lambda$,

$$\tilde{\chi}_\lambda(p(x,\xi)) \lesssim_N \langle|\xi|-\lambda\rangle^{-N}$$

This implies that

$$\left|\int_{\mathbb{R}^n} \tilde{\chi}_\lambda(p(x,\xi))s_{-1}(x,\xi) \, d\xi\right| \lesssim \int_0^{\lambda+O(1)} r^{n-2} + \int_{\lambda+O(1)}^{\infty} r^{n-1}(r-\lambda)^{-N} \lesssim \lambda^{n-1}.$$

Thus we have

$$\tilde{K}_\lambda(x,x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\rho(t)}{t+i0}e^{2\pi it(p(x,\xi)-\lambda)} + O(\lambda^{n-1})$$

$$= \int_{\mathbb{R}^n} \tilde{\chi}_\lambda(p(x,\xi)) \, d\xi + O(\lambda^{n-1})$$

$$= \int_{\mathbb{R}^n} \tilde{\chi}(p(x,\xi)-\lambda) \, d\xi + O(\lambda^{n-1})$$

$$= c(x)\lambda^n + \int_{\mathbb{R}^n} (\tilde{\chi}-\chi)(p(x,\xi)-\lambda) \, d\xi + O(\lambda^{n-1}).$$

Now $\tilde{\chi} - \chi = \chi * u$, where u is the inverse Fourier transform of $\rho(t) - 1$. But then

$$(\chi * u)(\tau) = \int_{-\infty}^{0} \hat{\rho}(\tau - \omega) \, d\omega - \mathbf{I}(\tau \geq 0),$$

from which we verify that

$$|(\chi * u)(\tau)| \lesssim_N (1 + |\tau|)^{-N}.$$

But we can use this to again show that

$$\left| \int_{\mathbb{R}^n} (\tilde{\chi} - \chi)(p(x, \xi) - \lambda) \, d\xi \right| \lesssim \int_0^{\lambda - O(1)} r^{n-1} |\xi|^{-N} \, dr$$
$$+ \int_{\lambda - O(1)}^{\lambda + O(1)} (\lambda - r)^{-N} r^{n-1} \, dr$$
$$+ \int_{\lambda + O(1)}^{\infty} r^{n-1} (r - \lambda)^{-N}$$
$$\lesssim \lambda^{n-1}.$$

Thus we have established the result for the kernel $K_{\lambda,\text{low}}$ of the operator $S_{\lambda,\text{low}}$. The proof of this theorem would be finished if we were able to justify that

$$|K_\lambda(x, x) - K_{\lambda,\text{low}}(x, x)| \lesssim \lambda^{n-1}.$$

Define the function

$$g(\lambda, x) = K_\lambda(x, x) - K_{\lambda,\text{low}}(x, x)$$
$$= \frac{1}{2\pi i} \int_{-\infty}^{\infty} (1 - \rho(t)) \{ e^{-2\pi i t (T - \lambda)}(x, x) \} \, dt.$$

Then the Fourier transform in the λ variable is

$$\hat{g}(t, x) = \frac{1}{2\pi i} (1 - \rho(t)) e^{-2\pi i t T}.$$

Thus the support of $\hat{g}$ is supported away from $|t| \leqslant \varepsilon/2$. A Lemma of Tauberian type following this proof shows that it suffices to show that for $\lambda > 0$, and $0 < \tau \leqslant 1$,

$$|g(\lambda + \tau, x) - g(\lambda + \tau, x)| \lesssim (1 + \lambda)^{n-1}$$

But this follows if we can show that

$$|K_{\lambda + \tau}(x, x) - K_\lambda(x, x)| \lesssim (1 + \lambda)^{n-1}$$

and that

$$|K_{\lambda + \tau,\text{low}}(x, x) - K_{\lambda,\text{low}}(x, x)| \lesssim (1 + \lambda)^{n-1}.$$

These inequalities are equivalent to the (L^1, L^2) *discrete restriction theorem* for eigenfunctions of the Laplacian, which we will analyze in the next section. Indeed, if we let $\chi_{\lambda,\tau}$ be the operator with kernel $K_{\lambda+\tau} - K_\lambda$, then the discrete restriction theorem tells us that

$$\|\chi_{\lambda,\tau}f\|_{L^2(M)} \lesssim (1+\lambda)^{(n-1)/2}\|f\|_{L^1(M)}$$

which by Schur's test, is equivalent to us having

$$\sup_{x\in M}\int_M |K_{\lambda+\tau}(x,y) - K_\lambda(x,y)|^2\,dy \lesssim (1+\lambda)^{n-1}$$

But the left hand side, by orthgonality, is equal to

$$\sup_{x\in M}\sum_j \mathbf{I}(\lambda \leqslant \lambda_j \leqslant \lambda+\tau)|e_j(x)|^2 = \sup_{x\in M}|K_{\lambda+\tau}(x,x) - K_\lambda(x,x)|,$$

which gives the required inequality. The analysis of $K_{\lambda,\text{low}}$ follows because we can write

$$K_{\lambda+\tau,\text{low}}(x,y) - K_{\lambda,\text{low}}(x,y) = \frac{1}{2\pi i}\int \frac{\rho(t)}{t+i0}(e^{-2\pi i\tau t} - 1)e^{2\pi i(T-\lambda)}\,dt.$$

The $e^{-2\pi i\tau t} - 1$ term annihilates the singularity at the origin, and one can now argue as above to show this quantity is $O(\lambda^{n-1})$. $\qquad\square$

Lemma 48.4. *Let $g(\lambda)$ be a piecewise continuous function on $\mathbb{R}$, such that for $\lambda > 0$, and $0 < \tau \leqslant 1$,*

$$|g(\lambda + \tau) - g(\lambda)| \lesssim (1+\lambda)^a.$$

If $\widehat{g}(t)$ vanishes for $|t| \leqslant 1$, then

$$|g(\lambda)| \lesssim (1+\lambda)^a.$$

Proof. If

$$G(\lambda) = \int_\lambda^{\lambda+1} g(\tau)\,d\tau,$$

then G is absolutely continuous, and for almost all λ, G is differentiable with

$$|G'(\lambda)| = |g(\lambda+1) - g(\lambda)| \lesssim (1+\lambda)^a.$$

743

The Fourier transform of G also vanishes for $|t| \leqslant 1$. But

$$|g(\lambda)| \leqslant |G(\lambda)| + O((1 + \lambda)^a)$$

so it suffices to prove the estimates for G. If $\eta(t) = 1$ for $|t| > 1$, and $\eta(t) = 0$ for $|t| \leqslant 1/2$, and if ψ has Fourier transform $\eta(t)/2\pi it$, then ψ is bounded and rapidly decreasing at infinity. But $G' * \psi = G$, which gives

$$|G(\lambda)| \lesssim (1 + \lambda)^a \int |\psi(s)|(1 + |s|)^a \, ds \lesssim (1 + \lambda)^a. \qquad \square$$

TODO: Compare to the proof of the lattice bounds $N(\lambda) = \text{Vol}(B) \cdot \lambda^n + O(\lambda^{n-2+2/(n+1)})$ that one can obtain in the setting $M = \mathbb{T}^d$, where B is the unit ball in $\mathbb{R}^d$.

Chapter 49

Nodal Sets

Let M^d be a compact manifold, and consider a classical, self-adjoint, elliptic pseudodifferential operator P with positive principal symbol. Then we have seen that one can decompose $L^2(M)$ into orthogonal eigenspaces $\oplus_\lambda E_\lambda$ of the Laplace-Beltrami operator on M, where λ ranges over a discrete subset of real numbers bounded from below, and E_λ is a finite dimensional subspace of $C^\infty(M)$ such that $PE_\lambda = \lambda E_\lambda$. The *node* of a real-valued eigenfunction e_λ is the zero set of e_λ, denoted N_λ or $N(e_\lambda)$. We are interested in the asymptotic behaviour of N_λ as $\lambda \to \infty$. In particular, one might ask questions such as:

- How does the number of connected components of $M - N_\lambda$, the *nodal count* of the eigenfunction, grow as $\lambda \to \infty$.

- What is the volume of N_λ. S.T. Yau's nodal size conjecture is that if $P = \sqrt{-\Delta}$, where Δ is the Laplace Beltrami operator, then

$$H^{n-1}(N_\lambda) \sim_M \lambda,$$

  where the implicit constant depends solely on the manifold M.

- One cannot 'hear the shape of a drum', i.e. there exists non-isomorphic compact Riemannian manifolds which are *isospectral*, i.e. whose Laplace-Beltrami operators have the same spectrum. However Smilansky has posed the question of whether one can 'count the shape of a drum', i.e. does the nodal count of N_λ for all λ determine the Riemannian manifold.

The analysis often breaks up into complex-variable techniques, which assume M is an analytic Riemannian manifold, and real-variable techniques, which depend

only on the fact that M is a smooth manifold. For concreteness, we will focus on the case where $P = \sqrt{-\Delta}$, which has been the scenario with the most focused study.

Example. *The only compact Riemannian manifold of dimension one, up to scaling, is $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. In this case, λ ranges over the non-negative integers, E_0 is one dimensional, the span of the constant function, and more generally, E_n is two dimensional, spanned by $e^{2\pi n i x}$ and $e^{-2\pi n i x}$. A general element of E_n can be written as*

$$e_n = A\cos(2\pi n x + \phi)$$

for some constants A and ϕ, and we see the nodal set $N(e_n)$ consists of $2n$ equidistributed points on the torus, and thus the nodal count is $NC(e_n) = 2n$.

More generally, we can consider a flat torus $T = \mathbb{R}^d/\Gamma$, where Γ is a lattice with some basis $\{z_1, \ldots, z_d\}$. We define the dual lattice Γ^* to be the space of all vectors w such that $w \cdot z \in \mathbb{Z}$ for any $z \in \Gamma$. If $w \in \Gamma^*$, then

$$e_w(x) = e^{2\pi i w \cdot x}$$

is seen to be an eigenfunction of the Laplacian on T with eigenvalue $2\pi|w|$. These form an orthogonal basis for $L^2(T)$, and therefore span the space of all eigenvalues. Due to the multiplicity of the eigenspaces, it seems very difficult to picture a general nodal set. But if we consider an eigenfunction of the form

$$e(x) = Ae^{2\pi i w \cdot x} + Be^{-2\pi i w \cdot x} = C\cos(2\pi w \cdot x + \phi)$$

then we see that the nodal set here is a union of $2v(w)$ hyperplanes, where $v(w)$ is the minimum positive value of $w \cdot \gamma$, where γ ranges over elements of Γ. If we consider the basis of Γ^* given by $\{w_1, \ldots, w_d\}$, where $w_i \cdot z_j = \delta_{i,j}$, and $w = \sum m_i w_i$, then $v(w)$ is the greatest common divisor of the integers $\{m_1, \ldots, m_n\}$.

Example. *The eigenfunctions of the sphere S^d consists of the spherical harmonics. The eigenvalues λ range over numbers of the form $n^{1/2}(n+d-1)^{1/2}$ for some $n \geqslant 0$, and for such λ, the eigenspace E_λ can be identified with the space of all harmonic polynomials in $d+1$ variables, homogeneous of degree n. The nodal sets of a general harmonic polynomial, however, is much more difficult to analyze than the torus, since the eigenspaces E_λ have high multiplicity for large λ.*

Knowledge of general nodal sets is very restricted. We know that the $d-1$ dimensional Hausdorff measure of a nodal set is finite, as well as the $d-2$ dimensional Hausdorff measure of the singularities of the nodal set, i.e. where e_λ and ∇e_λ vanishes.

746

49.1 Nodal Sets for Open Subsets of $\mathbb{R}^d$

Before we discuss the general setting, let us gain some intuition by studying the basic properties of the nodal sets of eigenfunctions of the Laplacian on bounded, open subsets of $\mathbb{R}^d$. We fix some open, bounded subset Ω of $\mathbb{R}^d$, and consider $C^2(\Omega)$, real-valued functions $e_\lambda : \Omega \to \mathbb{R}^d$ such that $\Delta e_\lambda = -\lambda^2 \phi$. In this case, we can reduce the study of e_λ to a harmonic function on $\Omega \times \mathbb{R}$, by setting $\phi(x,t) = e_\lambda(x)e^{\lambda t}$. Then

$$\Delta_{x,t}\phi = \Delta_x\phi + \Delta_t\phi = -\lambda^2\phi + \lambda^2\phi = 0.$$

Since $e^{\lambda t}$ never vanishes, we have $N(u) = N(\phi) \times \mathbb{R}$. Thus in this situation we can shift our analysis from the understanding of general eigenfunctions to the understanding of harmonic functions. In particular, we see that any eigenfunction lies in $C^\infty(\Omega)$.

Theorem 49.1. *Let B be the unit ball in $\mathbb{R}^d$. If $e_\lambda \in C^\infty(B)$, and $\Delta e_\lambda = -\lambda^2 e_\lambda$, then B/2 is contained in a $c\lambda^{-1}$ neighborhood of N_λ.*

Proof. We prove the result by contradiction. Suppose that $B/2$ was not contained in an $r < 1/4$ neighborhood of N_λ. Then e_λ does not change sign in some other ball B' with center lying in $B/2$ and with radius r, and we may assume without loss of generality that e_λ is positive here. Consider the harmonic function

$$\phi : B \times \mathbb{R} \to \mathbb{R}.$$

associated with e_λ by setting $\phi(x,t) = e_\lambda(x)e^{\lambda t}$. Harnack's inequality tells us that there exists a constant C such that

$$e^{\lambda r/2}\sup_{x\in B'} e_\lambda(x) = \sup_{|t|\leqslant r/2}\sup_{x\in B'}\phi(x,t) \leqslant C\inf_{|t|\leqslant r/2}\inf_{x\in B'}\phi(x,t) = Ce^{-\lambda r/2}\inf_{x\in B'} e_\lambda(x).$$

Thus

$$e^{\lambda r}\sup_{x\in B'} e_\lambda(x) \leqslant C\inf_{x\in B'} e_\lambda(x) \leqslant C\sup_{x\in B'} e_\lambda(x).$$

Thus, dividing out, we conclude that $e^{\lambda r} \leqslant C$, which implies that if $c = 2\log(C)$, then $r < c/\lambda$. □

The solution to the Cauchy uniqueness problem for harmonic functions tells us that non-zero harmonic functions can only vanish of order up to one on hyperplanes. This also remains true for eigenfunctions of the Laplacian by this extension trick.

49.2 Nodal Sets on the Sphere

To gain some intuition, let us restrict our analysis even further, to the study of *homogeneous harmonic polynomials* of degree k $\mathbf{H}_k$ on $\mathbb{R}^{d+1}$, which is equivalent to study eigenfunctions of the Laplacian on the sphere S^d. The space $\mathbf{H}_k$ then has dimension

$$\binom{d+k}{d} - \binom{d+k-2}{d}$$

General nodal sets are impossible to calculate, but the concreteness here should make a stronger analysis more possible.

Example. *The space $\mathbf{H}_k$ on $\mathbb{R}^2$ has dimension 2 for all k. The space therefore consists of functions expressible in polar coordinates in the form*

$$(r, \theta) \mapsto r^k \sin(k\theta + \phi).$$

Alternatively, we can get these results from the real and imaginary parts of the complex homogeneous polynomial $(x, y) \mapsto (x + iy)^k$. The nodal sets of these polynomials in $\mathbb{R}^2$ are precisely k lines through the origin, spaced out by a fixed angle. The nodal set on S^1 therefore has zero dimensional Hausdorff measure equal to k.

Example. *The space $\mathbf{H}_k$ on $\mathbb{R}^3$ has dimension $2k + 1$. One can find a basis of $\mathbf{H}_k$ using Legendre functions, but the nodal sets become hard to picture. But there are analogues here to the case of $\mathbb{R}^2$, e.g. the real and imaginary parts of the functions $(x + iy)^k$, $(x + iz)^k$, and $(y + iz)^k$, whose nodal sets are k planes with a common intersection line, and are space out at equal angles radially outward from this perpendicular line. Thus the one dimensional Hausdorff measure of this nodal set on S^2 is equal to $2k\pi$. One can perform a similar analysis in $\mathbb{R}^n$, with the functions $(x + iy)^k$ being formed from k hyperplanes in $\mathbb{R}^{d+1}$ with a common d dimensional plane lying at their intersection, and thus the nodal set on S^d has measure kc_d where c_d is the volume of an $d - 1$ dimensional great circle on S^d.*

To estimate the size of a general nodal set on a sphere, we utilize an integral geometric formula. Namely, if E is a smooth, n dimensional hypersurface in $\mathbb{R}^{n+1}$, and we define

$$A_i = \int_{\mathbb{R}^n} H^0(E \cap H_{i,y}) \, dy,$$

748

where $H_{i,y}$ is the hyperplane given by $\{x \in \mathbb{R}^{n+1} : (x_1, \ldots, \hat{x}_j, \ldots, x_{n+1}) = y\}$ for $y \in \mathbb{R}^n$, then

$$\left(\sum_{i=1}^{n+1} A_i^2 \right)^{1/2} \leqslant H^{n-1}(E) \leqslant \sum_{i=1}^{n+1} A_i.$$

This leads to a general bound.

Theorem 49.2. *If $f \in \mathbf{H}^k$, then $N(f) \subset S^d$ satisfies*

$$H^{d-1}(N(f)) \lesssim_d k.$$

Proof. Observe that the number of points in $N(f) = E \cap H_{i,y}$ is equal to the number of zeroes of the degree $\leqslant k$ polynomial $f(y_1, \ldots, t, \ldots, y_n)$. Since f is harmonic, and thus does not vanish to order greater than one on any hperplane, for almost every y, we conclude that $H^0(N(f) \cap H_{i,y}) \leqslant k$, and applying the formula above immediately yields the result. □

Remark. A lower bound here is immediately seen more difficult, since a general single variable degree $\leqslant k$ polynomial can have any number of zeroes on the real line, so this technique cannot work so well to get lower bounds unless we use analyticity in some way, and this probably depends on the underlying manifold we are working with to have some analytic structure. But one can obtain this result for the sphere, precisely because it has this analytic structure. TODO: See Han's notes on Nodal Sets.

49.3 Brownian Motion and Nodal Sets

Eigenfunctions e_λ to the Laplace-Beltrami operator on a compact manifold M^d behave well under the heat equation, i.e. if $e^{\Delta t}$ are the propogators for the heat equation $\partial_t = \Delta$ on M, and $e_\lambda u = -\lambda^2 u$, then

$$(e^{t\Delta} e_\lambda)(x) = e^{-\lambda^2 t} e_\lambda(x).$$

The heat equation is mathematically describing the distribution of a large number of particles, each diffusing through a medium in which they are subject to random molecular bombardments. The theory of *diffusions* in probability give us an alternate viewpoint through which to model this situation, so it makes sense this theory will bring light upon the theory of eigenfunctions.

49.3.1 Probabilistic Tools

To define diffusions, let us introduce the required machinery. We work over a fixed probability space Ω. A *continous stochastic process* valued in some space M is then a Borel-measurable function $X : \Omega \to C([0, \infty), M)$, which we also view as a family of M valued random variables $\{X_t : t \in [0, \infty)\}$. Intuitively, a stochastic process X is a random path formed by a particle travelling through space. Given a fixed process on M and $x \in M$, we let $\mathbb{E}^x[\cdot] = \mathbb{E}[\cdot | X_0 = x]$ denote the conditional expectation given that the process starts at the position x.

The most basic diffusion process is *Brownian motion*, and is used to generate the more general family. A one dimensional Brownian motion B is a continuous stochastic process such that for any interval $I = [t, s]$, the random variables $\Delta_I B = B_s - B_t$ are mean zero, variance $s - t$ Gaussians, and for any disjoint family of closed intervals $\{I_1, \ldots, I_n\}$ in $[0, \infty)$, the random variables $\{\Delta_{I_1} B, \ldots, \Delta_{I_n} B\}$ are independent of another. An n dimensional Brownian motion on $\mathbb{R}^d$ is precisely a stochastic process whose coordinates are independent Brownian motions. An *Itô diffusion* with *drift coefficient* $b(x)$ (valued in $\mathbb{R}^d$) and *diffusion coefficient* $A(x)$ (a $d \times d$ matrix) is a continuous random process X satisfying a *stochastic differential equation* of the form $dX = b(X)dt + A(X)dB$. The formal definition of this stochastic differential equation is quite difficult, but intuitvely, there exists a Brownian motion B such that

$$X_{t+\delta} \approx X_t + b(X_t)\delta + A(X_t)[B_{t+\delta} - B_t]$$

which the approximation becomes equality in a technical sense as $\delta \to 0$, i.e. the error $E_{t,\delta}$ between the two random variables is a random variable with mean $o(\delta)$ and third moment $O(\delta^3)$. As one might expect, one can define Itô diffusions on a compact Riemannian manifold M, given a vector field b on M operating as the drift coefficient, and a section of linear transformations on TM.

An important fact about the diffusions above is that they are *Markov*, i.e. if one observes the values of the Brownian motion $\{X_t\}$, up to a certain time T, the only information one observes which tells us about the evolution of the process past the time T is the position the diffusion arrives to at time T, i.e. the random variable X_T. More formally, a probabilist would write this by stating that for $S > T$, and for a function f on M,

$$\mathbb{E}[f(X_S)|\Sigma_T] = \mathbb{E}[f(X_S)|X_T],$$

where the conditional expectation is taken with respect to a σ algebra Σ_T, which models the information one sees when observing the process up to time T. One

can also let the times T and S be random *stopping times*, so a diffusion satisfies the *strong Markov property*.

Now we connect diffusions to the study of second order partial differential equations. For any $f \in C^2(M)$, if we define a differential operator L on M, the *generator of the diffusion*, by setting

$$Lf(x) = b(x) \cdot \nabla f(x) + \frac{A(x)A(x)^T \cdot Hf(x)}{2},$$

Then the derivative of the function $t \mapsto \mathbb{E}^x[f(X_t)]$ at $t = 0$ is precisely $Lf(x)$. Dynkin's formula follows by formally applying the fundamental theorem of calculus to this derivative, i.e. that

$$\mathbb{E}^x[f(X_T)] = \mathbb{E}^x[f(X_0)] + \mathbb{E}^x\left[\int_0^T Lf(X_s)\, ds\right] = f(x) + \mathbb{E}^x\left[\int_0^T Lf(X_s)\, ds\right],$$

and we can allow T to be a stopping time as well, provided that $\mathbb{E}^x[T] < \infty$. The *Feynman-Kac formula* reverses this process; if we define

$$u(x, t) = \mathbb{E}^x[f(X_t)]$$

then u is the solution to the partial differential equation $\partial_t u = Lu$, with initial conditions f. Thus one can convert problems about second order parabolic PDEs to problems about diffusions, and vice versa. In particular, we see that we can obtain a natural definition of Brownian motion on a manifold by letting B be a diffusion with generator given by the Laplace Beltrami operator.

TODO: Showing that it takes R^2/n time on average to leave a ball of radius R.

Next, we move onto the study of boundary value problems for semielliptic PDEs L via diffusions. Let us start with a bounded region D of $\mathbb{R}^d$. Consider a diffusion process X with generator L, and let $T_D = \inf\{t > 0 : X_t \notin D\}$. Then for sufficiently regular inputs, the Dynkin formula tells us that the unique solution to the problem $Lv = -h$ on D with boundary conditions ϕ on ∂D is given by

$$v(x) = \mathbb{E}^x[\phi(X_{T_D})] + \mathbb{E}^x\left[\int_0^{T_D} h(X_t)\, dt\right].$$

We can also solve the heat equation $\partial_t u = Lu$ with absorbing boundary conditions $u(x, t) = 0$ for $x \in \partial D$, and initial condition $u(x, 0) = f(x)$ by setting

$$u(x, t) = \mathbb{E}^x[f(X_t)\chi_t],$$

751

where $\chi_t = 1$ if $t < T_D$, and $\chi_t = 0$ otherwise (we kill paths that reach the boundary). Using linearity, we can now solve the problem $\partial_t u - Lu = h$ with a general boundary condition $u(x, t) = \phi(x)$ on ∂D and with initial conditions $u(x, 0) = f(x)$, namely, we have

$$u(x, t) = \mathbb{E}^x[\phi(X_{T_D})] + \mathbb{E}^x \left[\int_0^{T_D} h(X_t) \, dt \right] + \mathbb{E}^x[f(X_t)\chi_t].$$

Note that if D is bounded, and $f \in L^\infty(D)$, then dominated convergence implies that for each $x \in D$, $u(x, t) \to v(x)$ as $t \to \infty$, i.e. so the solution to Dirichlet's problem is really the long term distribution of the process.

49.4 Nodal Sets Via Brownian Motion

To exploit this relation to control nodal sets, we let Ω be a nodal domain of e_λ, i.e. a connected component of M. The Feynman-Kac formula tells us that solutions to the equation $\partial_t = \Delta$ on an open set $\Omega \subset M$, which are held equal to one on $\partial\Omega$ and vanish initially on the interior can be understood in terms of the probability that Brownian motion will hit $\partial\Omega$ in a given number of units of time.

We will get information about the nodal set by constructing diffusion processes which behave different near the boundary via the Feynman-Kac formula,

TODO: WHY IS NEUMANN VS DIRICHLET CONNECTED TO THE NODAL SET OF AN EIGENFUNCTION? NEUMANN CONDITIONS DON'T WORK, BECAUSE THE ASSOCIATED BROWNIAN MOTION MUST BE REFLECTED ABOUT THE SINGULAR SET OF e_λ, WHICH CAUSES PROBLEMS.

Instead of imposing boundary conditions, we instead construct a stochastic process which looks like a diffusion for small times, but converges back to initial data at large times. This leads to a bound of the form

$$H^{d-1}(N_\lambda) \gtrsim_M \lambda^{-\frac{n-3}{4}}$$

A related problem is to show that nodal domains can be 'concentrated' near flat surfaces. Call a smooth, embedded $d - 1$ dimensional hypersurface Σ in M *admissable up to distance r* if the geodesic flow outward from Σ is injective up to a distance r, i.e. any point in the r-thickening of Σ has a unique closest point on Σ. Roughly speaking, this implies that Σ is 'locally flat at a scale r'.

Theorem 49.3. *For any manifold M, there exists a constant $c > 0$ such that if Σ is admissable up to a distance $\lambda^{-1/2}$, then no nodal domain N_λ can be a contained in a $c\lambda^{-1/2}$ neighborhood of Σ.*

752

The proof will show one can generalize this result to Σ formed from a finite union of embedded manifolds admissable up to a distance $\lambda^{-1/2}$, provided these manifolds are sufficiently transversal to one another. This leads to a new proof of a class result due to Hayman.

Corollary 49.4. *There exists $c \geqslant 1/900$ such that if $\Omega \subset \mathbb{R}^2$ is simply connected with* inradius ρ *(ρ is the largest radius of a ball contained in Ω), $\lambda_1(\Omega) \geqslant c/\rho^2$.*

TODO: DO WE EXPECT DENSITY OF THE NODAL SET FOR LARGE λ, i.e. $\lambda^{-1/2}$ density.

(Citation: Green, Brown, and Probability) (Citation: Stochastic Differential Equations) (Citation: Tao Brownian Motion)

49.5   Discrete Restriction Theory

Let T be a classical, self-adjoint, elliptic pseudodifferential operator of order one on a compact manifold M, and consider an eigenfunction decomposition $\{e_j\}$ for T, corresponding to an increasing set of eigenvalues $\{\lambda_i\}$. We are concerned with the (L^p, L^q) bounds for the *spectral band projection operators*

$$\chi_{\lambda,\varepsilon} = \sum_{\lambda \leqslant \lambda_j \leqslant \lambda+\varepsilon} E_j.$$

These operators are closely related to restriction theory. The Stein-Tomas theorem studies the spherical projection operator, defined for functions on $\mathbb{R}^n$ by setting

$$\chi f(x) = \int_{|\xi|=1} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi.$$

We can also consider the *spherical projection operators*

$$\chi_\lambda f(x) = \int_{|\xi|=\lambda} \widehat{f}(\xi) e^{2\pi i \xi \cdot x} \, d\xi = \lambda^{-1} \text{Dil}_{1/\lambda} \chi \{\text{Dil}_\lambda f\}.$$

The Stein-Tomas theorem characterizes the mapping properties of the operator χ, showing that the only inequality the operator has are of the form

$$\|\chi f\|_{L^{p*}(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)}$$

for $1 \leqslant p \leqslant 2(n+1)/(n+3)$. Rescaled, one concludes that we have

$$\|\chi_\lambda f\|_{L^{p*}(\mathbb{R}^n)} \lesssim \lambda^{n(1/p-1/p*)-1} \|f\|_{L^p(\mathbb{R}^n)}.$$

753

This inequality implies that if we define the spectral band projection operators

$$\chi_{\lambda,\varepsilon}f(x) = \int_{\lambda \leqslant |\xi| \leqslant \lambda+\varepsilon} \widehat{f}(\xi)e^{2\pi i \xi \cdot x}\, d\xi = \int_{\lambda}^{\lambda+\varepsilon} S_{\tau}f(x)\, d\tau,$$

then,

$$\|\chi_{\lambda,\varepsilon}f\|_{L^{p^*}(\mathbb{R}^n)} \lesssim \int_{\lambda}^{\lambda+\varepsilon} \tau^{n(1/p-1/p^*)-1}\|f\|_{L^p(\mathbb{R}^n)}\, d\tau$$

$$\lesssim \varepsilon(1+\lambda)^{n(1/p-1/p^*)-1}\|f\|_{L^p(\mathbb{R}^n)}.$$

It is also clear that a proof of this inequality would prove that χ_{λ} was bounded, since $\chi_{\lambda} = \lim_{\varepsilon \to 0} \varepsilon^{-1}\chi_{\lambda,\varepsilon}$. Thus the theory of boundedness of spherical projections is closely related to the theory of boundedness for spectral band projections.

The original context for the spherical projection operators is to apply a TT^* argument, i.e. proving the (L^p, L^2) and (L^2, L^{p^*}) boundedness of the *restriction* and *extension* operators, which map functions on $\mathbb{R}^n$ to functions on λS^{n-1}, and vice-versa, by setting

$$R_{\lambda}f(\xi) = \widehat{f}(\xi) \quad \text{and} \quad E_{\lambda}f(x) = \int_{|\xi|=\lambda} f(\xi)e^{2\pi i \xi \cdot x}\, d\xi.$$

These operators are adjoints of one another, and $S_{\lambda} = E_{\lambda} \circ R_{\lambda}$. Thus a T^*T argument justifies that for $1 \leqslant p \leqslant 2(n+1)/(n+3)$,

$$\|R_{\lambda}f\|_{L^2(\lambda S^{n-1})} \lesssim \lambda^{n(1/p-1/2)-1/2}\|f\|_{L^p(\mathbb{R}^n)}$$

and

$$\|E_{\lambda}f\|_{L^{p^*}(\mathbb{R}^n)} \lesssim \lambda^{n(1/p-1/2)-1/2}\|f\|_{L^2(\lambda S^{n-1})}.$$

It is difficult to find direct analogues of the operators R_{λ} and E_{λ} on a manifold, just as we cannot directly find a direct analogue of χ_{λ}, because our eigenfunction is discrete. But we *can* consider analogues obtained by thickening the restriction set, i.e. considering the operators $R_{\lambda,\varepsilon}$ and $E_{\lambda,\varepsilon}$, mapping functions on $\mathbb{R}^n$ to functions on the annulus $A_{\lambda,\varepsilon} = \{\xi \in \mathbb{R}^n : \lambda \leqslant |\xi| \leqslant \lambda + \varepsilon\}$ and vice versa by setting

$$R_{\lambda,\varepsilon}f(\xi) = \widehat{f}(\xi) \quad \text{and} \quad E_{\lambda,\varepsilon}f(x) = \int_{\lambda \leqslant |\xi| \leqslant \lambda+\varepsilon} f(\xi)e^{2\pi i \xi \cdot x}\, d\xi.$$

We have $\chi_{\lambda,\varepsilon} = E_{\lambda,\varepsilon} \circ R_{\lambda,\varepsilon}$, and the Stein-Tomas theorem is then verified to be equivalent to bounds of the form

$$\|R_{\lambda,\varepsilon}f\|_{L^2(A_{\lambda,\varepsilon})} \lesssim \varepsilon\lambda^{n(1/p-1/2)-1/2}\|f\|_{L^p(\mathbb{R}^n)}$$

754

and

$$\|E_{\lambda,\varepsilon}f\|_{L^{p*}(\mathbb{R}^n)} \lesssim \varepsilon\lambda^{n(1/p-1/2)-1/2}\|f\|_{L^2(A_{\lambda,\varepsilon})}.$$

But the orthogonality of the Fourier transform, shows that the restriction bound is equivalent to the fact that

$$\|\chi_{\lambda,\varepsilon}f\|_{L^2(\mathbb{R}^n)} \lesssim \varepsilon\lambda^{n|1/p-1/2|-1/2}\|f\|_{L^p(\mathbb{R}^n)}$$

and

$$\|\chi_{\lambda,\varepsilon}f\|_{L^{p*}(\mathbb{R}^n)} \lesssim \varepsilon\lambda^{n|1/p-1/2|-1/2}\|f\|_{L^2(\mathbb{R}^n)}.$$

This bound is the one we will begin by focus on generalizing to arbitrary manifolds, and we will start with case $p = 1$. Since there is no ambiguity, in the sequel, we let $\chi_\lambda = \chi_{\lambda,1}$.

We note that in $\mathbb{R}^n$, one can prove the same result for any positive homogeneous function $p(\xi)$ of degree one such that the *cosphere* $\Sigma = \{\xi : p(\xi) = 1\}$ has non-vanishing curvature, Stein-Tomas gives results about

$$\chi_\lambda f(x) = \int_\Sigma \hat{f}(\xi)e^{2\pi i\xi\cdot x}\,d\xi.$$

Returning to the analysis of the operator T with principal symbol $p(x,\xi)$. We will obtain good results about the operator χ_λ provided that the cospheres $\Sigma(x) = \{\xi : p(x,\xi) = 1\}$ has non-vanishing curvature for all $x \in M$.

The easiest bound in the Stein-Tomas theorem are the equivalent bounds

$$\|\chi_\lambda f\|_{L^2(\mathbb{R}^n)} \lesssim \lambda^{(n-1)/2}\|f\|_{L^1(\mathbb{R}^n)}$$

and

$$\|\chi_\lambda f\|_{L^\infty(\mathbb{R}^n)} \lesssim \lambda)^{(n-1)/2}\|f\|_{L^2(\mathbb{R}^n)},$$

which does not even utilize the curvature of the domain of projection whatsoever, solely using the fact that the sphere of radius λ has surface measure $O(\lambda^{n-1})$. Let us see if we can justify the analogous result on a compact manifold. We will find we do not need to assume curvature here either.

Lemma 49.5. *Let T be a classical, elliptic, self-adjoint pseudodifferential operator of order one on a compact manifold M of dimension n, and consider the association spectral band projection operators $\{\chi_\lambda\}$. Then*

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim \lambda^{(n-1)/2}\|f\|_{L^1(M)},$$

and thus by self-adjointness, that

$$\|\chi_\lambda f\|_{L^\infty(M)} \lesssim \lambda^{(n-1)/2}\|f\|_{L^2(M)}.$$

755

Proof. We want to exploit the parametrix for the wave equation. Thus we start by replacing χ_λ with an operator $\tilde{\chi}_\lambda$ given by

$$\tilde{\chi}_\lambda f(x) = \sum_j \chi(\lambda_j - \lambda) E_j,$$

where χ is a non-negative Schwartz function with $\chi(0) > 0$, and such that $\hat{\chi}(t)$ has support on $|t| \leqslant \varepsilon/2$. Because $\mathbf{I}(\lambda \leqslant \lambda_j \leqslant \lambda + \varepsilon) \lesssim \chi(\lambda_j - \lambda)$, orthogonality shows that

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim \|\tilde{\chi}_\lambda f\|_{L^2(M)}$$

and so it suffices to prove the result for $\tilde{\chi}_\lambda$. But if $\tilde{\chi}_\lambda$ has kernel

$$\tilde{K}_\lambda(x, y) = \sum_j \chi(\lambda_j - \lambda) e_j(x)\overline{e_j(y)}.$$

This is equivalent to showing that

$$\|\tilde{K}_\lambda\|_{L^\infty_y L^2_x} \lesssim \lambda^{(n-1)/2}.$$

But orthogonality shows that

$$\begin{aligned}
\|\tilde{K}_\lambda\|_{L^\infty_y L^2_x} &= \left(\sup_{y \in M} \sum_j \chi(\lambda_j - \lambda)^2 |e_j(y)|^2 \right)^{1/2} \\
&\leqslant \|\chi\|_{L^\infty}^{1/2} \left(\sup_{y \in M} \sum_j \chi(\lambda_j - \lambda) |e_j(y)|^2 \right)^{1/2} \\
&\leqslant \|\chi\|_{L^\infty}^{1/2} \left(\sup_{y \in M} \tilde{\chi}_\lambda(y, y) \right)^{1/2}.
\end{aligned}$$

Thus we need to show that for all $y \in M$,

$$\tilde{\chi}_\lambda(y, y) \lesssim \lambda^{n-1}.$$

This follows by an argument similar to the proof of the Weyl law, e.g. applying the parametrix for the wave equation. $\square$

Now we move onto results that require a curvature assumption.

756

Theorem 49.6. *Let T be a classical, elliptic, self-adjoint pseudodifferential operator of order one on a compact manifold M of dimension n and with principal symbol $p(x,\xi)$, such that the cospheres*

$$\Sigma(x) = \{\xi \in T^*M : p(x,\xi) = 1\}$$

have non-vanishing curvature for each $x \in M$. If we consider the association spectral band projection operators $\{\chi_\lambda\}$, then for $p_c = 2(n+1)/(n+3)$, we have

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim (1+\lambda)^{(n-1)/2(n+1)} \|f\|_{L^{p_c}(M)}.$$

Remark. Combined with the (L^1, L^2) boundedness of the equation, and the fact that

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim \|f\|_{L^2(M)}$$

we conclude using interpolation that for $1 \leqslant p \leqslant 2(n+1)/(n+3)$,

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim (1+\lambda)^{n|1/p-1/2|-1/2} \|f\|_{L^p(M)}$$

and for $2(n+1)/(n+3) \leqslant p \leqslant 2$,

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim (1+\lambda)^{(\frac{n-1}{2})(1/p-1/2)} \|f\|_{L^p(M)}.$$

All these inequalities are sharp, as we will discuss after the proof.

Proof. We start by using a strategy we began the (L^1, L^2) boundedness result with, namely, to swap χ_λ out with a more well behaved operator. The operators $\tilde{\chi}_\lambda$ used in that proof become singular near the diagonal as $\lambda \to \infty$, i.e. the kernel is $\Theta(\lambda^n)$. Thus we modify their definition slightly. We fix a small quantity $\varepsilon_0 \leqslant \varepsilon/2$, and define a Schwartz function χ (not necessarily positive this time), with $\chi(0) = 1$ and with Fourier support on $\varepsilon_0/2 \leqslant t \leqslant \varepsilon_0$. Nonetheless, it remains true that $\chi(x) \gtrsim 1$ for $|x| \leqslant \varepsilon$, so that

$$\|\chi_\lambda f\|_{L^2(M)} \lesssim \|\tilde{\chi}_\lambda f\|_{L^2(M)}.$$

It thus suffices to obtain a bound for $\tilde{\chi}_\lambda$ at the critical exponent. Applying the parametrix, and seeing that the remainder term has operator norm $O_N((1+\lambda)^{-N})$ for all $N > 0$, it suffices to show that the operator $\tilde{\chi}_\lambda$ wih kernel

$$\int \int \hat{\chi}(t) s(t, x, y, \xi) e^{2\pi i[\phi(x,y,\xi)+t(p(y,\xi)-\lambda)]} \, dt \, d\xi$$

757

is well behaved. The kernel of this operator is an oscillatory integral, and on the support of the amplitude s, and if we set $\Phi(t, x, y, \xi) = \phi(x, y, \xi) + tp(y, \xi)$ has

$$|\nabla_\xi \Phi(t, x, y, \xi)| \gtrsim 1$$

on the support of the amplitude $\widehat{\chi}s$ for $\xi \neq 0$ except if $|x-y| \sim \varepsilon_0$. Thus, applying stationary phase, if $\eta \in C^\infty(M \times M)$ is a function which equals one for $|x-y| \sim \varepsilon_0$, and vanishes everywhere else, the difference between $\tilde{\chi}_\lambda$ and the operator χ'_λ with kernel

$$\int \int \widehat{\chi}(t) s(t, x, y, \xi) \eta(x, y) e^{2\pi i [\Phi(t,x,y,\xi) - \lambda t]} \, d\xi \, dt$$

is an operator of the form

$$\int \widehat{\chi}(t) A(t, x, y) e^{-2\pi i \lambda t},$$

where A is smooth, and it therefore follows this kernel is $O_N((1 + \lambda)^{-N})$. Let us set $a(t, x, y, \xi) = \widehat{\chi}(t) s(t, x, y, \xi) \eta(x, y)$. Our proof is now completed by applying some results about non-homogeneous oscillatory integral operators, once we show that

$$\int \int a(t, x, y, \xi) e^{2\pi i [\Phi(t,x,y,\xi) - \lambda t]} \, d\xi \, dt = \lambda^{(n-1)/2} a_\lambda(x, y) e^{2\pi i \lambda \psi(x,y)},$$

where ψ is non-homogeneous, and satisfies the $n \times n$ Carleson-Sjölin conditions on the support of a_λ, which is smooth, and satisfies $|\partial_x^\alpha \partial_y^\beta a_\lambda| \lesssim_{\alpha,\beta} 1$, uniformly in λ. $\qquad \square$

Part VII

Harmonic Analysis and Fractals

Chapter 50

Dimensions via Fourier Integrals

The key to applying Fourier analysis to the theory of fractal dimension begins with Frostman's lemma.

Lemma 50.1. *(Frostman) Let $E \subset \mathbb{R}^d$ be a Borel set. If $H^s(E) > 0$, then there exists a non-negative measure μ supported on E and a constant $c > 0$ such that*

$$\mu(B(x, r)) \leqslant cr^s$$

for all $x \in \mathbb{R}^d$ and $r > 0$.

A simple application of Frostman's lemma, is that $\dim_{\mathbb{H}}(A \times B) \geqslant \dim_{\mathbb{H}}(A) + \dim_{\mathbb{H}}(B)$, i.e. by taking product measures.

We call such a measure a *Frostman measure of dimension s*, and denote it by $\dim_{\mathbb{H}}(\mu)$. In particular, Frostman's lemma guarantees that for any Borel set E,

$$\dim_{\mathbb{H}}(E) = \sup\{\dim_{\mathbb{H}}(\mu) : \operatorname{supp}(\mu) \subset E\}.$$

A key operator associated with Frostman's lemma is the *s energy* of a measure μ, defined to be

$$I_s(\mu) = \int \int |x - y|^{-s} d\mu(x) d\mu(y) = \int k_s * \mu \, d\mu,$$

where $k_s(x) = |x|^{-s}$ is the *Riesz kernel*. If μ has compact support, then $I_s(\mu) < \infty$ implies that $I_t(\mu) < \infty$ for $0 < t < s$, so, from a local perspective, having finite s energy for larger s is a more restrictive condition. Energies immediately connect to Frostman type bounds via the identity

$$\int |x - y|^{-s} d\mu(y) = s \int_0^\infty \frac{\mu(B(x, r))}{r^s} \frac{dr}{r}.$$

Thus if $\mu(B(x, r)) \leqslant cr^s$, and is supported on a set with diameter at most D, then for any $\varepsilon > 0$,

$$
\begin{aligned}
\int |x - y|^{-(s-\varepsilon)} d\mu(y) &\leqslant (s - \varepsilon) \int_0^D \frac{cr^s}{r^{s-\varepsilon}} \frac{dr}{r} \\
&\leqslant (s - \varepsilon) \int_0^D cr^\varepsilon \frac{dr}{r} \\
&\leqslant c\frac{s - \varepsilon}{\varepsilon} D^\varepsilon.
\end{aligned}
$$

Thus in particular,

$$
I_s(\mu) \leqslant c\frac{s - \varepsilon}{\varepsilon} D^\varepsilon.
$$

In particular, if μ is a compactly supported Borel measure, we conclude that $\dim_{\mathbb{H}}(\mu) = \sup\{s : I_s(\mu) < \infty\}$. Conversely, if μ is a measure such that $I_s(\mu) < \infty$, then we conclude that for some $M > 0$, the set

$$
E_M = \{x : \int |x - y|^{-s} d\mu(y) \leqslant M\}
$$

has positive μ measure. If we consider $\nu = \mu \cdot \mathbf{I}_{E_M}$, then we see that

$$
\nu(B(x, r)) \leqslant (2^s M) r^s
$$

for all $x \in \mathbb{R}^d$ and $r > 0$, so that $\dim_{\mathbb{H}}(\nu) \geqslant s$. Thus we conclude that for any Borel set E,

$$
\dim_{\mathbb{H}}(E) = \sup\{s > 0 : \text{There is a probability measure } \mu \text{ supported on } E \text{ with } I_s(\mu) < \infty\},
$$

Thus Hausdorff dimension is closely connected to energy integrals.

Example. *If μ is the Lebesgue measure restricted to $[0, 1]^d$, then $I_s(\mu) < \infty$ for $s < d$. The Lebesgue measure on $\mathbb{R}^d$ has infinite s energy for every s. Similarily, if μ is the one-dimensional Hausdorff measure restricted to a rectifiable curve, then $I_s(\mu) < \infty$ for $s < 1$.*

Example. *If μ is the Cantor measure on the $1/3$ Cantor set, then $I_s(\mu) < \infty$ for $s < \log 2/\log 3$.*

Now let's allow Fourier analysis to enter the picture. Applying the multiplication formula for the Fourier transform, noting that for $0 < s < d$, $\hat{k}_s$ is a constant multiple of k_{d-s}, we conclude that

$$
\begin{aligned}
I_s(\mu) &= \langle k_s * \mu, \mu \rangle \\
&= \langle \widehat{k_s * \mu}, \hat{\mu} \rangle \\
&= \langle k_{d-s}\hat{\mu}, \hat{\mu} \rangle \\
&= \int |\xi|^{s-d} |\hat{\mu}(\xi)|^2 \, d\xi.
\end{aligned}
$$

To be more formally correct, we must apply a mollification to μ, so that we can interpret everything distributionally, and then apply Fatou's lemma to ensure everything works out at the end. But we'll leave that as an exercise.

For signed measures μ and ν, we can define the mutual energy

$$
I_s(\mu, \nu) = \int \int |x - y|^{-s} d\mu(x) \overline{d\nu(y)}.
$$

The function $(\mu, \nu) \mapsto I_s(\mu, \nu)$ is then a conjugate symmetric bilinear map, which is positive definite, and thus the set of signed measures μ such that $I_s(\mu) < \infty$ forms a Hilbert space. We then have a formula

$$
I_s(\mu, \nu) = \int |\xi|^{s-d} \hat{\mu}(\xi) \overline{\hat{\nu}(\xi)},
$$

which relates the mutual energy to the Fourier transforms.

If $I_s(\mu) < \infty$, then we must have $|\hat{\mu}(\xi)| \lesssim |\xi|^{-s/2}$ for *most frequencies ξ*. This manifests in various different ways:

- A weak bound together with the fact that

$$
\int_{|\xi| \leqslant R} |\xi|^s |\hat{\mu}(\xi)|^2 \leqslant R^d I_s(\mu)
$$

  implies that as $R \to \infty$,

$$
\{ |\xi| \leqslant R : |\hat{\mu}(\xi)| \geqslant |\xi|^{-s/2} \} = o(R^d).
$$

- If we define

$$
A_r(\mu) = \int_{|\xi|=1} |\hat{\mu}(r\xi)|^2 \, d\sigma(r),
$$

then $I_s(\mu) = \int_0^\infty A_r(\mu)r^{s-1}\,dr$, and we actually have

$$A_r(\mu) \lesssim I_s(\mu)r^{-s}.$$

To see this, we apply the inverse Fourier transform, writing

$$A_r(\mu) = \int \hat{\sigma}(r(x-y))d\mu(x)d\mu(y),$$

and then using the fact that $|\hat{\sigma}(r(x-y))| \lesssim r^{-s}|x-y|^{-s}$ and reverting back to energy integrals. Thus we see that L^2 spherical averages have the decay required.

Unfortunately, it is not true that we have a pointwise bound $|\hat{\mu}(\xi)| \lesssim |\xi|^{-s/2}$, which is a stricter condition. We define the *Fourier dimension* of a measure μ to be

$$\dim_{\mathbb{F}}(\mu) = \sup\{0 < s \leqslant d : \sup_\xi |\xi|^{s/2}\hat{\mu}(\xi) < \infty\}.$$

If $s < \dim_{\mathbb{F}}(\mu)$, then $I_s(\mu) < \infty$, so $\dim_{\mathbb{H}}(\mu) \leqslant \dim_{\mathbb{F}}(\mu)$. The *Fourier dimension* of a set E is equal to

$$\dim_{\mathbb{F}}(E) = \sup\{\dim_{\mathbb{F}}(\mu) : \operatorname{supp}(\mu) \subset E\}.$$

Then we see that $\dim_{\mathbb{F}}(E) \leqslant \dim_{\mathbb{H}}(E)$ for any Borel E. A set is *Salem* if it's Fourier dimension is equal to it's Hausdorff dimension.

Example. *A sphere in $\mathbb{R}^d$ is a Salem set of dimension $d - 1$. On the other hand, a plane is* not *a Salem set. In fact, the Fourier dimension of the plane is equal to* zero, *since the Fourier transform of any measure supported on a plane has no decay in the direction tangential to the plane.*

It is often true that Fourier dimension gives more information than the Hausdorff dimension. For instance, here is a result about sum set. For a set E, define $E^{\oplus n}$ to be $E + \cdots + E$, the n time set addition.

Theorem 50.2. *Suppose $E \subset \mathbb{R}$. If $\dim_{\mathbb{F}}(E) > 1/k$, then $|E^{\oplus k}| > 0$, and if $\dim_{\mathbb{F}}(E) > 2/k$, then $E^{\oplus k}$ contains an open interval.*

Proof. If μ is a measure, define μ^{*k} to be the k-fold convolution of μ with itself. Then if $|\hat{\mu}(\xi)| \lesssim |\xi|^{-s/2}$, then

$$|\widehat{\mu^{*k}}(\xi)| = |\hat{\mu}(\xi)|^k \lesssim |\xi|^{-sk/2}.$$

If $s > 1/k$, then we conclude that $\widehat{\mu^{*s}}$ is an L^2 function, and thus that μ^{*s} is an L^2 function, and thus is non-vanishing on a set of positive Lebesgue measure. If μ is supported on E, then μ^{*s} is supported on $E^{\oplus k}$, giving that $E^{\oplus k}$ contains a set of positive Lebesgue measure. If $s > 2/k$, then $\widehat{\mu^{*s}}$ is integrable, so μ^{*s} is a continuous function, which implies the points where it is non-vanishing are an open set, which is contained in $E^{\oplus k}$. $\qquad\square$

Chapter 51

Hausdorff Dimensions of Projections

Here we prove Marstrand's projection theorem, that if $E \subset \mathbb{R}^d$ is a Borel set, and $V \subset \mathbb{R}^d$ is a random subspace of $\mathbb{R}^d$ of dimension m, then with probability one, the projection $\pi_V : \mathbb{R}^d \to \mathbb{R}^d$ has the property that

$$\dim_{\mathbb{H}}(\pi_V(E)) = \min(\dim_{\mathbb{H}}(E), m).$$

We will show a proof using Fourier analysis, which also allows a more robust characterization of how many planes fail to have the right projection properties.

To do this, it is convenient to introduce the *Sobolev dimension* of a measure μ to be

$$\dim_{\mathbb{S}}(\mu) = \sup \left\{ s \in \mathbb{R} : \int |\hat{\mu}(\xi)|^2 (1 + |\xi|)^{s-d} \, d\xi < \infty \right\},$$

i.e. $\dim_{\mathbb{S}}(\mu)$ is the supremum of s such that $\mu \in H^{s/2 - d/2}(\mathbb{R}^d)$. We can consider the Sobolev energy $I_s(\mu)$, defined as above. It is equal to the energy we have defined in the last chapter for $0 < s < d$, but extends this definition to all $s \in \mathbb{R}$. The Sobolev dimension of a measure, roughly speaking, gives a measure of how smooth the measure is. For instance, if $\dim_{\mathbb{S}}(\mu) > d$, then $\mu \in L^2(\mathbb{R}^d)$, and if $\dim_{\mathbb{S}}(\mu) > 2d$, then μ is a continuous function.

Theorem 51.1. *Let $E \subset \mathbb{R}^d$ be a Borel set with $\dim_{\mathbb{H}}(E)$. If V is a random m dimensional subspace of $\mathbb{R}^d$ chosen from the Grassmannian manifold $G(d, m)$ with respect to the Haar measure on $G(d, m)$, then for almost every m dimensional subspace V,*

$$\dim_{\mathbb{H}}(P_V(E)) = \max(\dim_{\mathbb{H}}(E), m).$$

Moreover, if $\dim_{\mathbb{H}}(E) > m$, then $H^m(P_V(E)) > 0$ for almost every V.

Proof. We reduce the result to an integral argument about measures. Fourier analysis enters the picture because plane waves behave nicely under lifts. If $s < \dim_{\mathbb{H}}(E)$, then we can find a probability measure μ supported on E with $I_s(\mu) < \infty$. For each $V \in G(d, m)$, we consider the pushforward measures $\mu_V = (P_V)_* \mu$. We will show that

$$\int_{G(d,m)} I_s(\mu_V) < \infty,$$

from which it follows that $I_s(\mu_V) < \infty$ for almost every $V \in G(d, m)$, and, since μ_V is supported on $P_V(E)$, that $\dim_{\mathbb{H}}(P_V(E)) \geq s$ for almost every V. If we can choose $s > m$ as in the example above, then we conclude that for almost every V, by restricting μ_V to a smaller subset, we can choose a Frostman measure ν supported on $P_V(E)$ with dimension m. But this means that $H^m(P_V(E)) > 0$.

It now suffices to show that

$$\int_{G(d,m)} I_s(\mu_V) < \infty.$$

s $\hfill\square$

Chapter 52

Cantor Measures

Let us study the Fourier analytic properties of some Cantor like sets.

52.1 Middle Dissection Cantor Sets

For $0 < \alpha < 1/2$, we define a set $C_\alpha \subset [0,1]$, as the limit of a family of sets $C_{\alpha,n}$ constructed iteratively. We start by defining $C_{\alpha,0} = [0,1]$. For $k \geqslant 0$, the set $C_{\alpha,k}$ will be a union of disjoint intervals $\{I_{k,j}\}$ of length α^k. We define $C_{\alpha,k+1}$ as the union of the intervals obtaining by removing the middle portion of length $(1 - 2\alpha)2^{-k}\alpha^k$ from each interval $I_{k,j}$, and keeping the two length α^{k+1} intervals on the left and right that remain. Iterating this algorithm gives $C_\alpha = \bigcap_k C_{\alpha,k}$.

At each stage, $C_{\alpha,k}$ will be a union of 2^k intervals of length α^k. This gives an optimal cover of C_α at the scale α^k, from which one can prove the Minkowski dimension of C_α is equal to

$$s_\alpha = \frac{\log(2^k)}{\log(\alpha^{-k})} = \frac{\log 2}{\log(1/\alpha)}.$$

This is *also* the Hausdorff dimension of C_α. One can actually show that

$$H^{s_\alpha}(C_\alpha) = 1$$

The upper bound is immediately obtained by the covering argument above. Conversely, if $J_1, \ldots, J_N$ is any cover of C_α by intervals with length at most α^k (which we can assume is finite by compactness), then by a 'parent' argument one can show that

$$\sum |J_i|^{s_\alpha} \geqslant 2^k \alpha^{k s_k} = 1.$$

767

Thus taking $k \to \infty$ yields the right bound. Thus the Hausdorff dimension of C_α is also s_α.

We can construct a natural measure on C_α by taking a limit of the uniform probability measures on $C_{\alpha,k}$. Let us denote the limiting measure as μ_α. One can verify that

$$\mu(B(x,r)) \lesssim r^{s_\alpha}.$$

Moreover, one actually has that for $x \in C_\alpha$, $\mu(B(x,r)) \gtrsim r^{s_\alpha}$. We call such a measure *Ahlfors regular*.

Nowl et's compute the Fourier transform of μ_α.

Theorem 52.1. *We have*

$$\hat{\mu}_\alpha(\xi) = \prod_{j=0}^{\infty} \cos(\pi(1-\alpha)\alpha^j \xi)$$

Proof. It's easiest to do this by expressing μ_α as a limit of sums of Dirac deltas, and it is notationally best to do this using a binary sequence. We set

$$\mathcal{E}_k = \{0,1\}^k$$

and for $\varepsilon \in \mathcal{E}_k$, define $a(\varepsilon) = \sum_{j=1}^{k} \varepsilon_j (1-\alpha)\alpha^{j-1}$. For each ε, the point $a(\varepsilon)$ is the left hand endpoint of one of the 2^k intervals defining $C_{\alpha,k}$. If we define

$$\nu_k = 2^{-k} \sum_{\varepsilon \in \mathcal{E}_k} \delta_{\alpha(\varepsilon)}$$

then μ is the weak limit of the sequence of measures ν_k. We have

$$\hat{\nu}_k(\xi) = 2^{-k} \sum_{\varepsilon \in \mathcal{E}_k} e^{-2\pi i a(\varepsilon)\cdot \xi}$$

$$= 2^{-k} \sum_{\varepsilon \in \mathcal{E}_k} \prod_{j=1}^{k} e^{-2\pi i (1-\alpha)\alpha^{j-1}\xi \varepsilon_j}$$

$$= 2^{-k} \prod_{j=1}^{k} (1 + e^{-2\pi i (1-\alpha)\alpha^{j-1}\xi})$$

$$= 2^{-k} \prod_{j=1}^{k} e^{-i\pi(1-\alpha)\alpha^{j-1}\xi} \cos(\pi(1-\alpha)\alpha^{j-1}\xi)$$

$$= e^{i\pi(\alpha^k - 1)\xi} \prod_{j=1}^{k} \cos(\pi(1-\alpha)\alpha^{j-1}\xi).$$

The result follows now by taking $k \to \infty$. $\square$

Thus in particular, we get

$$|\widehat{\mu_{1/3}}(\xi)| = \prod_{j=1}^{\infty} \cos(2\pi 3^{-j}\xi).$$

If $\xi = 3^k$, then

$$\widehat{\mu_{1/3}}(\xi) = \prod_{j=1}^{\infty} \cos(2\pi/3^j) \geqslant \prod_{j=1}^{\infty} (1 - O(3^{-2j})) \gtrsim 1$$

Thus $\limsup_{\xi \to \infty} |\widehat{\mu_{1/3}}(\xi)| \neq 0$. Thus in particular, $\dim_{\mathbb{F}}(\mu_{1/3}) = 0$. More generally, if $\alpha \leqslant 1/3$, then there is no measure on C_α whose Fourier transform tends to zero as $\alpha \to \infty$ (such measures are called *Rachjman measures*).

Theorem 52.2. *If $n \geqslant 3$, and $\alpha = 1/n$, then C_α does not support a Rachjman measure.*

Proof. For any $x \in C_\alpha$, and $k \geqslant 1$, $[\alpha^{-k}x] \notin (\alpha, 1 - \alpha)$. Indeed, the map $x \mapsto [\alpha^{-k}x]$ maps intervals to their parents in the Cantor set construction. So let μ be supported on C_α. Choose a non-negative C_c^∞ function ϕ supported on $(\alpha, 1 - \alpha)$ with $\int \phi(x)\, dx = 1$. Consider the rescaled functions

$$\phi_j(x) = \phi([\alpha^{-k}x]).$$

Then ϕ_j has disjoint support from C_α for all j. Then

$$\phi_j(x) = \sum \hat{\phi}(n) e^{2\pi i \alpha^{-j}kx}.$$

Thus ϕ_j has Fourier support on integer multiplers of α^{-j}. Thus

$$\begin{aligned}
0 &= \int \phi_j(x) d\mu \\
&= \sum_n \hat{\phi}_j(n)\overline{\hat{\mu}(n)} \\
&= \sum_n \hat{\phi}(n)\overline{\hat{\mu}(\alpha^{-j}n)} \\
&= 1 + \sum_{0 < |n| < T} \hat{\phi}(n)\overline{\hat{\mu}(\alpha^{-j}n)} + \sum_{|n| \geqslant T} \hat{\phi}(n)\overline{\hat{\mu}(\alpha^{-j}n)}.
\end{aligned}$$

769

We have that for all $N > 0$,

$$\left| \sum_{|n| \geqslant T} \widehat{\phi}(n)\overline{\widehat{\mu}(\alpha^{-j}n)} \right| \leqslant \sum_{|n| \geqslant T} |\widehat{\phi}(n)| \lesssim_N T^{-N}.$$

On the other hand, we have

$$\left| \sum_{0 < |n| < T} \widehat{\phi}(n)\overline{\widehat{\mu}(\alpha^{-j}n)} \right| \leqslant 2T \sup_{0 < |n| < T} |\widehat{\mu}(\alpha^{-j}n)|.$$

Thus we conclude that for all $j > 0$,

$$\sup_{0 < |n| < T} |\widehat{\mu}(\alpha^{-j}n)| \geqslant 1/2T - O_N(T^{-N}) \gtrsim 1/T.$$

Thus it follows that there exists $|m| \geqslant \alpha^{-j}$ for all $j > 0$ such that $|\widehat{\mu}(m)| \gtrsim 1$, and thus μ cannot be a Rachjman measure. $\qquad \square$

All that we needed was that there was some interval $I \subset [0, 1]$ and an increasing sequence $\{k_j\}$ such that $[k_j x] \notin I$ for all $x \in C$. Thus if μ is a Rajchmann measure supported on some set E, it follows that for any interval I, and any sequence $\{k_j\}$, there exists $x \in E$ such that $[k_j x] \in I$. Thus points in E are 'recurrent' in some sense.

There are values α such that μ_α is a Rachjmann measure. We say a number $\theta > 1$ is *Pisot* if there exists $\lambda \neq 0$ such that

$$\sum_{k=0}^{\infty} \sin^2(\lambda\theta^k) < \infty.$$

This means that, roughly speaking, $\lambda\theta^k$ is roughly concentrated near integers values. The condition turns out to be equivalent to being an algebraic integer whose conjugates have modulus less than one.

Theorem 52.3. *The measure μ_α is Rachjmann if and only if $1/\alpha$ is not a Pisot number.*

Proof. Let $\theta = 1/\alpha$. Suppose that μ_α is not Rachjmann. Choose $\delta > 0$ and $\{u_k\}$ such that $|\widehat{\mu_\alpha}(u_k)| \geqslant \delta$ for all k. Write $\pi(1 - \alpha)u_k = \lambda_k\theta^{m_k}$ for $1 \leqslant \lambda_k < \theta$ and

770

some increasing sequence of integers $\{m_k\}$. Without loss of generality, we may assume λ_k converges to some λ as $k \to \infty$. Then

$$
\begin{aligned}
\delta &< |\widehat{\mu_\alpha}(u_k)| \\
&= |\prod_{j=0}^{\infty} \cos(\pi(1 - \alpha)\alpha^{j-1}u_k)| \qquad\qquad = |\prod_{j=0}^{\infty} \cos(\lambda_k\theta^{m_k-j+1})| \\
&\leqslant |\prod_{j=0}^{m_k} \cos(\lambda_k\theta^{m_k-j+1})|.
\end{aligned}
$$

TODO: Finish the argument.

Conversely, suppose θ is a Pisot number. Then there is λ such that

$$
\sum_{j=0}^{\infty} \sin^2(\lambda\theta^j) < \infty.
$$

Thus if $\pi(1 - \alpha)u_k = \lambda\theta^k$, then TODO

$$
|\widehat{\mu_\alpha}(u_k)| \gtrsim 1.
$$

Thus μ_α is not a Rachjmann measure. □

It is a harder (but true) result that C_α supports a Rachjmann measure if and only if $1/\alpha$ is Pisot.

52.2   Modified Cantor Sets

Fix $0 < N < M$, and consider a Cantor set $C_{N,M}$ where, at each stage, we take an interval with relative length 1, divide it into N subintervals with relative length $1/N$ and take the leftmost portion with relative length $1/M$. Thus after k iterations of the algorithm, we have N^k intervals with sidelength $1/M^k$. Thus we have a set with Minkowski dimension $s_{N,M} = \log N / \log M$. Similar arguments to before yield that this is also the Hausdorff dimension, and moreover, $H^{s_{N,M}}(C_{N,M}) = 1$. For each $\varepsilon \in \mathcal{E}_k = \{0, \ldots, N-1\}^k$, we can associate an interval of sidelength $1/M^k$ in the kth iteration of the algorithm, with left endpoint $\alpha(\varepsilon) = \sum_{j=0}^{\infty} \varepsilon_j M^{-j}/N$. Let us denote the uniform measure on $C_{N,M}$ by $\mu_{N,M}$. Similar arguments show that

$$
|\widehat{\mu}_{N,M}(\xi)| = \prod_{j=1}^{\infty} \frac{1 + e^{2\pi i\xi M^{-j}/N} + \cdots + e^{2\pi i\xi M^{-j}(N-1)/N}}{N}.
$$

This sequence is poorly behaved when ξ is an integer multiple of M^{-j} for any $j > 0$, i.e. the first m terms are equal to one if $\xi = NM^{-m}$.

52.3 Self Similar Measures

Consider a contractive similariy $S : \mathbb{R}^d \to \mathbb{R}^d$, i.e. such that $|S(x) - S(y)| = r|x-y|$ for some $0 < r < 1$. Then $Sx = rRx+a$ for some $R \in O_d$. A borel measure μ is *self similar* if there are similarity maps $S_1, \ldots, S_N$ and $0 < p_1, \ldots, p_N \leqslant 1$ such that $\sum p_i = 1$, and
$$\mu = \sum p_i(S_j)_* \mu.$$
Conversely, for any $S_1, \ldots, S_N$ and $p_1, \ldots, p_N$, a unique such measure μ exists. The support of μ is the unique compact set K which is invariant under S_j, i.e. such that $K = \bigcup S_j(K)$. If $p_j = r_j^s$, where r_j is the contraction ratio of S_j, and s is the unique value that makes these things sum to one, then s is called the *similarity dimension*, and one has $\dim_{\mathbb{H}}(K) = s$. The Cantor sets C_α fit into this scenario if $S_1(x) = \alpha x$ and $S_2(x) = \alpha x + 1 - \alpha$. If μ is self similar, then we have a simple formula for the Fourier transform, namely, if $\mathcal{J}_m = \{1, \ldots, N\}^m$, then

$$\hat{\mu}(\xi) = \lim_{m \to \infty} \sum_{J \in \mathcal{J}_m} (p_{J_1} \cdots p_{J_m}) e^{-2\pi i \xi \cdot a_J}$$

where $(S_{J_1} \circ \cdots \circ S_{J_N})x = r_J R_J x + a_J$.

Chapter 53

Riemann's 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 53.1. *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 53.2 (Cantor-Lebesgue Theorem). *If, for some pair of sequences $a_0, a_1, \ldots$ and $b_0, b_1, \ldots$ 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 $a_n \cos(2\pi n x) + b_n \sin(2\pi n x) = r_n \cos(nx + c_n)$. 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.

Part VIII

Abstract Harmonic Analysis

The main property of spaces where Fourier analysis applies is symmetry – for a function $\mathbb{R}$, we can translate and negate. On $\mathbb{R}^n$ 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 54

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 those groups which are locally compact.

Example. *There are a few groups we should keep in mind for intuition in the general topological group.*

- *The classical groups $\mathbb{R}^n$ and $\mathbf{T}^n$, from which Fourier analysis originated.*

- *The group μ of roots of unity, rational numbers $\mathbf{Q}$, and cyclic groups $\mathbf{Z}_n$, which occur in number theory.*

- *The matrix subgroups of the general linear group $GL(n)$, which occur in representation theory.*

- *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 $|p^{-m}q|_p = p^m$.*

54.1 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 54.1. *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 $\mathbb{R}$, take $U = \alpha\mathbb{Z}$, and $V = \mathbb{Z}$, where α is an irrational number. Then $U + V = \alpha\mathbb{Z} + \mathbb{Z}$ is dense in $\mathbb{R}$, and is hence not closed.*

There are useful ways we can construct neighbourhoods under the group operations, which we list below.

Lemma 54.2. *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 54.3. *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 54.4. *Every Kolmogorov topological group is Hausdorff.*

Theorem 54.5. *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 $\overline{A} \subset \bigcap AV$. $\square$

Theorem 54.6. *Every open subgroup of G is closed.*

Proof. Let H be an open subgroup of G. Then $\overline{H} = \bigcap_V HV$. If W is a neighbourhood of the origin contained in H, then we find $\overline{H} \subset HW \subset H$, so H is closed. $\square$

We see that open subgroups of a group therefore correspond to connected components of the group, so that connected groups have no proper open subgroups. This also tells us that a locally compact group is σ-compact on each of its components, for if V is a pre-compact neighbourhood of the origin, then $V^2, V^3, \ldots$ are all precompact, and $\bigcup_{k=1}^{\infty} V^k$ is an open subgroup of G, which therefore contains the component of e, and is σ-compact. Since the topology of a topological group is homogenous, we can conclude that all components of the group are σ compact.

54.2   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 54.7. *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. □

Theorem 54.8. *If G is locally compact, G/H is also.*

Proof. If $\{U_i\}$ is a basis of precompact neighbourhoods at the origin, then $U_i H$ 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 $U_i H \subset V$. □

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.

54.3   Uniform Continuity

An advantage of the real line $\mathbb{R}$ 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 $f : \mathbb{R} \to \mathbb{R}$ 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 $f : G \to \mathbb{C}$, 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.

781

Example. *Let G be any Hausdorff non-commutative topological group, with se-quences x_i and y_i for which $x_i y_i \to e$, $y_i x_i \to z \neq e$. Then the uniform structures on G are not equivalent.*

It is hopeless to express uniform continuity in terms of a new topology on G, because the topology only gives a local description of continuity, which prevents us from describing things uniformly across the whole group. However, we can express uniform continuity in terms of a new topology on $G \times G$. If $U \subset G$ is an open neighbourhood of the origin, let

$$L_U = \{(x,y) : yx^{-1} \in U\} \quad R_U = \{(x,y) : x^{-1}y \in U\}$$

The family of all L_U (resp. R_U) is known as the left (right) uniform structure on G, denoted $LU(G)$ and $RU(G)$. Fix a map $f : G \to H$, and consider the map

$$g(x,y) = (f(x), f(y))$$

from G^2 to H^2. Then f is left (right) uniformly continuous if and only if g is continuous with respect to $LU(G)$ and $LU(H)$ ($RU(G)$ and $RU(H)$). $LU(G)$ and $RU(G)$ are weaker than the product topologies on G and H, which reflects the fact that uniform continuity is a strong condition than normal continuity. We can also consider uniform maps with respect to $LU(G)$ and $RU(H)$, and so on and so forth. We can also consider uniform continuity on functions defined on an open subset of a group.

Example. *Here are a few examples of easily verified continuous maps.*

- *If the identity map on G is left-right uniformly continuous, then $LU(G) = RU(G)$, and so uniform continuity is invariant of the uniform structure cho-sen.*

- *Translation maps $x \mapsto axb$, for $a, b \in G$, are left and right uniform.*

- *Inversion is uniformly continuous.*

Theorem 54.9. *All continuous maps on compact subsets of topological groups are uniformly continuous.*

Proof. Let K be a compact subset of a group G, and let $f : K \to H$ be a continuous map into a topological group. We claim that f is then uniformly continuous. Fix

an open neighbourhood V of the origin, and let V' be a symmetric neighbourhood such that $V'^2 \subset V$. For any x, there is U_x such that

$$f(x)^{-1}f(xU_x) \subset V'$$

Choose U'_x such that $U'^2_x \subset U_x$. The xU'_x cover K, so there is a finite subcover corresponding to sets $U'_{x_1}, \ldots, U'_{x_n}$. Let $U = U'_{x_1} \cap \cdots \cap U'_{x_n}$. Fix $y \in G$, and suppose $y \in x_k U'_{x_k}$. Then

$$
\begin{aligned}
f(y)^{-1}f(yU) &= f(y)^{-1}f(x_k)f(x_k)^{-1}f(yU) \\
&\subset f(y)^{-1}f(x_k)f(x_k)^{-1}f(x_kUx_k) \\
&\subset f(y)^{-1}f(x_k)V' \\
&\subset V'^2 \subset V
\end{aligned}
$$

So that f is left uniformly continuous. Right uniform continuity is proven in the exact same way. $\qquad\square$

Corollary 54.10. *All maps with compact support are uniformly continuous.*

Corollary 54.11. *Uniform continuity on compact groups is invariant of the uniform structure chosen.*

54.4 Ordered Groups

In this section we describe a general class of groups which contain both interesting and pathological examples. Let G be a group with an ordering $<$ preserved by the group operations, so that $a < b$ implies both $ag < bg$ and $ga < gb$. We now prove that the order topology gives G the structure of a normal topological group (the normality follows because of general properties of order topologies).

First note, that $a < b$ implies $a^{-1} < b^{-1}$. This results from a simple algebraic trick, because
$$a^{-1} = a^{-1}bb^{-1} > a^{-1}ab^{-1} = b^{-1}$$
This implies that the inverse image of an interval (a, b) under inversion is (b^{-1}, a^{-1}), hence inversion is continuous.

Now let $e < b < a$. We claim that there is then $e < c$ such that $c^2 < a$. This follows because if $b^2 \geqslant a$, then $b \geqslant ab^{-1}$ and so

$$(ab^{-1})^2 = ab^{-1}ab^{-1} \leqslant ab^{-1}b = a$$

Now suppose $a < e < b$. If $\inf\{y : y > e\} = x > e$, then $(x^{-1}, x) = \{e\}$, and the topology on G is discrete, hence the continuity of operations is obvious. Otherwise, we may always find c such that $c^2 < b$, $a < c^{-2}$, and then if $c^{-1} < g, h < c$, then

$$a < c^{-2} < gh < c^2 < b$$

so multiplication is continuous at every pair (x, x^{-1}). In the general case, if $a < gh < b$, then $g^{-1}ah^{-1} < e < g^{-1}bh^{-1}$, so there is c such that if $c^{-1} < g', h' < c$, then $g^{-1}ah^{-1} < g'h' < g^{-1}bh^{-1}$, so $a < gg'h'h < b$. The set of gg', where $c^{-1} < g' < c$, is really just the set of $gc^{-1} < x < gc$, and the set of $h'h$ is really just the set of $c^{-1}h < x < ch$. Thus multiplication is continuous everywhere.

Example (Dieudonne). *For any well ordered set S, the dictionary ordering on $\mathbb{R}^S$ induces a linear ordering inducing a topological group structure on the set of maps from S to $\mathbb{R}$.*

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 $\mathbb{R}^S$ 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 $\mathbb{R}^S$ is isomorphic (topologically) to a discrete union of a certain number of copies of $\mathbb{R}$, one for each tuple in $S - \{w\}$.

If S has a countable cofinal subset $\{s_i\}$, the topology is no longer so simple, but $\mathbb{R}^S$ 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 $\mathbb{R}^S$ 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 $\mathbb{R}^S$, where S does not have a countable cofinal subset, then eventually $f_i = f$. We conclude all countable sets in $\mathbb{R}^S$ are closed, and this proof easily generalises to show that if S does not have a cofinal set of cardinality $\mathfrak{a}$, then every set of cardinality $\leqslant \mathfrak{a}$ is closed.

The simple corollary to this proof is that compact subsets are finite. Let $X = f_1, f_2, \ldots$ be a denumerable, compact set. Since all subsets of X are compact, we

may assume $f_1 < f_2 < \ldots$ (or $f_1 > f_2 > \ldots$, which does not change the proof in any interesting way). There is certainly $g \in \mathbb{R}^S$ such that $g < f_1$, and then the sets $(g, f_2), (f_1, f_3), (f_2, f_4), \ldots$ form an open cover of X with no finite subcover, hence X cannot be compact. We conclude that the only compact subsets of $\mathbb{R}^S$ are finite.

Furthermore, the class of open sets is closed under countable intersections. Consider a series of functions

$$f_1 \leqslant f_2 \leqslant \cdots < h < \cdots \leqslant g_2 \leqslant g_1$$

Suppose that $f_i \leqslant k < h < k' \leqslant g_j$. Then the intersection of the (f_i, g_i) contains an interval (k, k') around h, so that the intersection is open near h. The only other possiblity is that $f_i \to h$ or $g_i \to h$, which can only occur if $f_i = h$ or $g_i = h$ eventually, in which case we cannot have $f_i < h$, $h < g_i$. We conclude the intersection of countably many intervals is open, because we can always adjust any intersection to an intersection of this form without changing the resulting intersecting set (except if the set is empty, in which case the claim is trivial). The general case results from noting that any open set in an ordered group is a union of intervals.

54.5 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 $\mathcal{N} = \{\mathbf{Z}, 2\mathbf{Z}, 3\mathbf{Z}, \ldots \}$. 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 $1, 2, 3, \ldots$ 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, $\{p_1, p_2, \ldots, p_n\}$.

By the fundamental theorem of arithmetic,

$$\{-1, 1\} = (\mathbf{Z}p_1)^c \cap \cdots \cap (\mathbf{Z}p_n)^c$$

and is therefore an open set. But this is clearly not the case as open sets must contain infinite sequences.

Chapter 55

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 $\mathbb{R}$ (or $\mathbb{R}^n$). It commutes with translations because it is the measure induced by the linear functional corresponding to Riemann integration on $C_c^+(\mathbb{R}^n)$. 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 $\mathbb{R}^+$, dr/r, and the Haar measure on $\mathbf{T}$, $d\theta$.

Example. *The space $M_n(\mathbb{R})$ of all n by n real matrices under addition has a Haar measure dM, which is essentially the Lebesgue measure on $\mathbb{R}^{n^2}$. If we consider the measure on $GL_n(\mathbb{R})$, defined by*

$$\frac{dM}{det(M)^n}$$

To see this, note the determinant of the map $M \mapsto NM$ on $M_n(\mathbb{R})$ is $det(N)^n$, because we can view $M_n(\mathbb{R})$ as the product of $\mathbb{R}^n$ n times, multiplication operates on the space componentwise, and the volume of the image of the unit paralelliped in each $\mathbb{R}^n$ 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 $\mathbb{R}^n$, 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 $\phi : C_c(G) \to \mathbb{R}$ 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 55.1. *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 $\mathrm{supp}(f_1 + f_2)$, then the functions

$$h = f_1 + f_2 + \varepsilon g$$

$$h_1 = f_1/h \quad h_2 = f_2/h$$

are in $C_0^+(G)$, if we define $h_i(x) = 0$ if $f_i(x) = 0$. This implies that there is a neighbourhood V of e such that if $x \in V$, and y is arbitrary, then

$$|h_1(xy) - h_1(y)| \leqslant \varepsilon \quad |h_2(xy) - h_2(y)| < \varepsilon$$

If $\mathrm{supp}(\phi) \subset V$, and $h \leqslant \sum c_i L_{x_i}\phi$, then

$$f_j(x) = h(x)h_j(x) \leqslant \sum c_i \phi(x_i x)h_j(x) \leqslant \sum c_i \phi(x_i x) \left[h_j(x_i^{-1}) + \varepsilon\right]$$

since we may assume that $x_i x \in \mathrm{supp}(\phi) \subset V$. Then, because $h_1 + h_2 \leqslant 1$,

$$(f_1 : \phi) + (f_2 : \phi) \leqslant \sum c_j[h_1(x_j^{-1}) + \varepsilon] + \sum c_j[h_2(x_j^{-1}) + \varepsilon] \leqslant \sum c_j[1 + 2\varepsilon]$$

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 $\mathrm{supp}(\phi) \subset V$.

790

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 $F : C_c^+(G) \to \mathbb{R}$ 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 $I : C_c^+(G) \to \mathbb{R}$ contained in $\bigcap K(V)$. This means that every neighbourhood of I contains I_ϕ with $\mathrm{supp}(\phi) \subset V$, for all ϕ. This means that if $f_1, f_2 \in C_c^+(G)$, $\varepsilon > 0$, and V is arbitrary, there is ϕ with $\mathrm{supp}(\phi) \subset V$, and

$$|I(f_1) - I_\phi(f_1)| < \varepsilon \quad |I(f_2) - I_\phi(f_2)| < \varepsilon$$

$$|I(f_1 + f_2) - I_\phi(f_1 + f_2)| < \varepsilon$$

this implies that if V is chosen small enough, then

$$|I(f_1 + f_2) - (I(f_1) - I(f_2))| \leqslant 2\varepsilon + |I_\phi(f_1 + f_2) - (I_\phi(f_1) + I_\phi(f_2))| < 3\varepsilon$$

Taking $\varepsilon \to 0$, we conclude I is linear. Similar limiting arguments show that I is homogenous of degree 1, and commutes with all left translations. We conclude the extension of I to a linear functional on $C_0(G)$ is well defined, and the Radon measure obtained by the Riesz representation theorem is a Haar measure. $\qquad \square$

We shall prove that the Haar measure is unique, but first we show an incredibly useful regularity property.

Proposition 55.2. *If U is open, and μ is a Haar measure, then $\mu(U) > 0$. It follows that if f is in $C_c^+(G)$, then $\int f d\mu > 0$.*

Proof. If $\mu(U) = 0$, then for any $x_1, \ldots, x_n \in G$,

$$\mu\left(\bigcup_{i=1}^n x_i U\right) \leqslant \sum_{i=1}^n \mu(x_i U) = 0$$

If K is compact, then K can be covered by finitely many translates of U, so $\mu(K) = 0$. But then $\mu = 0$ by regularity, a contradiction. $\qquad \square$

Theorem 55.3. *Haar measures are unique up to a multiplicative constant.*

Proof. Let μ and ν be Haar measures. Fix a compact neighbourhood V of the identity. If $f, g \in C_c^+(G)$, consider the compact sets

$$A = \text{supp}(f)V \cup V\text{supp}(f) \quad B = \text{supp}(g)V \cup V\text{supp}(g)$$

Then the functions $F_y(x) = f(xy) - f(yx)$ and $G_y(x) = g(xy) - g(yx)$ are supported on A and B. There is a neighbourhood $W \subset V$ of the identity such that $\|F_y\|_\infty, \|G_y\|_\infty < \varepsilon$ if $y \in W$. Now find $h \in C_c^+(G)$ with $h(x) = h(x^{-1})$ and $\text{supp}(h) \subset W$ (take $h(x) = k(x)k(x^{-1})$ for some function $k \in C_c^+(G)$ with $\text{supp}(k) \subset W$, and $k = 1$ on a symmetric neighbourhood of the origin). Then

$$\left(\int h d\mu \right) \left(\int f d\lambda \right) = \int h(y)f(x)d\mu(y)d\lambda(x)$$

$$= \int h(y)f(yx)d\mu(y)d\lambda(x)$$

and

$$\left(\int h d\lambda \right) \left(\int f d\mu \right) = \int h(x)f(y)d\mu(y)d\lambda(x)$$

$$= \int h(y^{-1}x)f(y)d\mu(y)d\lambda(x)$$

$$= \int h(x^{-1}y)f(y)d\mu(y)d\lambda(x)$$

$$= \int h(y)f(xy)d\mu(y)d\lambda(x)$$

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

$$\left| \frac{\int f d\lambda}{\int f d\mu} - \frac{\int h d\lambda}{\int h d\mu} \right| \leqslant \frac{\varepsilon\lambda(A)}{\int f d\mu}$$

792

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 $\lambda = c\mu$. $\qquad\square$

The theorem can also be proven by looking at the translation invariant properties of the derivative $f = d\mu/d\nu$, where $\nu = \mu + \lambda$ (We assume our group is σ compact for now). Consider the function $g(x) = f(yx)$. Then

$$\int_A g(x) d\nu = \int_{yA} f(x) d\nu = \mu(yA) = \mu(A)$$

so g is derivative, and thus $f = g$ almost everywhere. Our interpretation is that for a fixed y, $f(yx) = f(x)$ almost everywhere with respect to ν. Then (applying a discrete version of Fubini's theorem), we find that for almost all x with respect to ν, $f(yx) = f(x)$ holds for almost all y. But this implies that there exists an x for which $f(yx) = f(x)$ holds almost everywhere. Thus for any measurable A,

$$\mu(A) = \int_A f(y) d\nu(y) = f(x)\nu(A) = f(x)\mu(A) + f(x)\nu(A)$$

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 $\mu(A) = \nu(A)$.

55.1 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 $Y_1 \subset Y_2 \subset \cdots \to Y$ be an increasing subsequence such that $\mu(Y_i) < \infty$, then $|f(x)| \leqslant \|f\|_\infty$ almost everywhere for $x \in Y_i$, and so by the monotone convergence theorem

$$\int |fg| d\mu = \lim_{Y_i \to \infty} \int_{Y_i} |fg| d\mu \leqslant \|f\|_\infty \int_{Y_i} |g| d\mu \leqslant \|f\|_\infty \|g\|_1$$

Thus the map $g \mapsto \int fg d\mu$ is a well defined, continuous linear functional with norm $\|f\|_\infty$. That $L^1(X)^* = L^\infty(X)$ follows from the decomposibility of the Carathéodory extension of μ, a fact we leave to the general measure theorists.

55.2 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 *modular function*, is a continuous homomorphism from G to the real numbers. Note that Δ is trivial if and only if G is unimodular.

Theorem 55.4. *Any compact group is unimodular.*

Proof. $\Delta : G \to \mathbb{R}^*$ is a continuous homomorphism, hence $\Delta(G)$ is compact. But the only compact subgroup of $\mathbb{R}$ 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 group theory.

Theorem 55.5. *If G/G^c is compact, then G is unimodular.*

Proof. Δ factors through G/G^c since it is abelian. But if Δ is trivial on G/G^c, it must also be trivial on G. $\square$

Corollary 55.6. *If G is a connected, semi-simple Lie group, then G is unimodular.*

Proof. The Lie algebra $\mathfrak{g}$ of G is a direct sum of simple Lie algebras, and so $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}$. But $[\mathfrak{g}, \mathfrak{g}]$ is the Lie algebra of $[G, G]$, so $G = [G, G]$ since G is connected. $\square$

More generally, if G is a connected Lie group, and Ad denotes the adjoint representation of G on $\mathfrak{g}$, then $\Delta(x) = \det(\mathrm{Ad}(x^{-1}))$. TODO

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.

55.3 Convolution

If G is a topological group, then $C(G)$ does not contain enough algebraic structure to identify G – for instance, if G is a discrete group, then $C(G)$ is defined solely by the cardinality of G. The algebras we wish to study over G is the space $M(G)$ of all complex valued Radon measures over G and the space $L^1(G)$ of integrable functions with respect to the Haar measure, because here we can place a Banach algebra structure with an involution. We note that $L^1(G)$ can be isometrically identified as the space of all measures $\mu \in M(G)$ which are absolutely continuous with respect to the Haar measure. Given $\mu, \nu \in M(G)$, we define the convolution measure

$$\int \phi d(\mu * \nu) = \int \phi(xy) d\mu(x) d\nu(y)$$

The measure is well defined, for if $\phi \in C_c^+(X)$ is supported on a compact set K, then

$$\left| \int \phi(xy) d\mu(x) d\nu(y) \right| \leqslant \int_G \int_G \phi(xy) d|\mu|(x) d|\nu|(y)$$
$$\leqslant \|\mu\| \|\nu\| \|\phi\|_\infty$$

This defines an operation on $M(G)$ which is associative, since, by applying the associativity of G and Fubini's theorem.

$$\int \phi d((\mu * \nu) * \lambda) = \int \int \phi(xz) d(\mu * \nu)(x) d\lambda(z)$$
$$= \int \int \int \phi((xy)z) d\mu(x) d\nu(y) d\lambda(z)$$
$$= \int \int \int \phi(x(yz)) d\mu(x) d\nu(y) d\lambda(z)$$
$$= \int \int \phi(xz) d\mu(x) d(\nu * \lambda)(z)$$
$$= \int \phi d(\mu * (\nu * \lambda))$$

Thus we begin to see how the structure of G gives us structure on $M(G)$. Another example is that convolution is commutative if and only if G is commutative. We have the estimate $\|\mu * \nu\| \leqslant \|\mu\| \|\nu\|$, because of the bound we placed on the integrals above. $M(G)$ is therefore an involutive Banach algebra, which has a unit, the dirac delta measure at the identity.

As a remark, we note that involutive Banach algebras have nowhere as near a nice of a theory than that of C^* algebras. $M(G)$ cannot be renormed to be a C^* algebra, since every weakly convergent Cauchy sequence converges, which is impossible in a C^* algebra, except in the finite dimensional case.

A **discrete measure** on G is a measure in $M(G)$ which vanishes outside a countable set of points, and the set of all such measures is denoted $M_d(G)$. A **continuous measure** on G is a measure μ such that $\mu(\{x\}) = 0$ for all $x \in G$. We then have a decomposition $M(G) = M_d(G) \oplus M_c(G)$, for if μ is any measure, then $\mu(\{x\}) \neq 0$ for at most countably many points x, for

$$\|\mu\| \geqslant \sum_{x \in G} |\mu|(x)$$

This gives rise to a discrete measure ν, and $\mu - \nu$ is continuous. If we had another decomposition, $\mu = \psi + \phi$, then $\mu(\{x\}) = \psi(\{x\}) = \nu(\{x\})$, so $\psi = \nu$ by discreteness, and we then conclude $\phi = \mu - \nu$. $M_c(G)$ is actually a closed subspace of $M(G)$, since if $\mu_i \to \mu$, and $\mu_i \in M_c(G)$, and $\|\mu_i - \mu\| < \varepsilon$, then for any $x \in G$,

$$\varepsilon > \|\mu - \mu_i\| \geqslant |(\mu_i - \mu)(\{x\})| = |\mu(\{x\})|$$

Letting $\varepsilon \to 0$ shows continuity.

The convolution on $M(G)$ gives rise to a convolution on $L^1(G)$, where

$$(f * g)(x) = \int f(y)g(y^{-1}x)dy$$

which satisfies $\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$. This is induced by the identification of f with $f(x)dx$, because then

$$\int \phi(f(x)dx * g(x)dx) = \int \int \phi(yx)f(y)g(x)dydx$$
$$= \int \phi(y) \left(\int f(y)g(y^{-1}x)dx \right) dy$$

Hence $fd\mu * gd\mu = (f * g)d\mu$. What's more,

$$\|f\|_1 = \|fd\mu\|$$

If $\nu \in M(G)$, then we can still define $\nu * f \in L^1(G)$

$$(\nu * f)(x) = \int f(y^{-1}x)d\mu(y)$$

798

which holds since

$$\int \phi d(v * f\mu) = \int \phi(yx)f(x)dv(y)d\mu(x) = \int \phi(x)f(y^{-1}x)dv(y)d\mu(x)$$

If G is unimodular, then we also find

$$\int \phi d(f\mu * v) = \int \phi(yx)f(y)d\mu(y)dv(x) = \int \phi(x)f(y)d\mu(y)dv(y^{-1}x)$$

So we let $f * \mu(x) = \int f(y)d\mu(y^{-1}x)$.

Theorem 55.7. $L^1(G)$ 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 v is an arbitrary measure, and μ is continuous, then

$$(\mu * v)(\{x\}) = \int_G \mu(\{y\})dv(y^{-1}x) = 0$$

$$(v * \mu)(\{x\}) = \int_G \mu(\{y\})dv(xy^{-1}) = 0$$

so $M_c(G)$ is an ideal. Finally, we verify $L^1(G)$ is closed, because it is complete, and if $v \in M(G)$ is arbitrary, and if U has null Haar measure, then

$$(fdx * v)(U) = \int \chi_U(xy)f(x)dx \, dv(y) = \int_G \int_{y^{-1}U} f(x)dxdv(y) = 0$$

$$(v * fdx)(U) = \int \chi_U(xy)dv(x)f(y)dy = \int_G \int_{Ux^{-1}} f(y)dydv(x) = 0$$

So $L^1(G)$ is a two-sided ideal. $\qquad\qquad\square$

If we wish to integrate by right multiplication instead of left multiplication, we find by the substitution $y \mapsto xy$ that

$$(f * g)(x) = \int f(y)g(y^{-1}x)dy$$

$$= \int\int f(xy)g(y^{-1})dy$$

$$= \int\int \frac{f(xy^{-1})g(y)}{\Delta(y)}dy$$

Observe that

$$f * g = \int f(y)L_{y^{-1}}g\ dy$$

which can be interpreted as a vector valued integral, since for $\phi \in L^\infty(\mu)$,

$$\int (f * g)(x)\phi(x)dx = \int f(y)g(y^{-1}x)\phi(x)dxdy$$

so we can see convolution as a generalized 'averaging' of translate of g with respect to the values of f. If G is commutative, this is the same as the averaging of translates of f, but not in the noncommutative case. It then easily follows from operator computations $L_z(f * g) = (L_z f) * g$, and $R_z(f * g) = f * (R_z g)$, or from the fact that

$$(f * g)(zx) = \int f(y)g(y^{-1}zx)dy = \int f(zy)g(y^{-1}x)dy = [(L_z f) * g](x)$$

$$(f * g)(xz) = \int f(y)g(y^{-1}xz)dy = [f * (R_z g)](x)$$

Convolution can also be applied to the other L^p spaces, but we have to be a bit more careful with our integration.

Theorem 55.8. *If $f \in L^1(G)$ and $g \in L^p(G)$, then $f * g$ is defined for almost all x, $f * g \in L^p(G)$, and $\|f * g\|_p \leqslant \|f\|\|g\|_p$. If G is unimodular, then the same results hold for $g * f$, or if G is not unimodular and f has compact support.*

Proof. We use Minkowski's inequality to find

$$\|f * g\|_p = \left(\int \left|\int f(y)|g(y^{-1}x)dy\right|^p dx\right)^{1/p}$$

$$\leqslant \int |f(y)| \left(\int |g(y^{-1}x)|^p dx\right)^{1/p} dy$$

$$= \|f\|_1 \|g\|_p$$

800

If G is unimodular, then

$$\|g * f\|_p = \left(\int \left| \int g(xy^{-1}) f(y) dy \right|^p dx \right)^{1/p}$$

and we may apply the same trick as used before.

If f has compact support K, then $1/\Delta$ is bounded above by $M > 0$ on K and

$$\|g * f\|_p = \left(\int \left| \int \frac{g(xy^{-1}) f(y)}{\Delta(y)} dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int \left(\int \left| \frac{g(xy^{-1}) f(y)}{\Delta(y)} \right|^p dx \right)^{1/p} dy$$

$$= \|g\|_p \int_K \frac{|f(y)|}{\Delta(y)} d\mu(y)$$

$$\leqslant M \|g\|_p \|f\|_1$$

which shows that $g * f$ is defined almost everywhere. $\square$

Theorem 55.9. *If G is unimodular, $f \in L^p(G)$, $g \in L^q(G)$, and $p = q^*$, then $f * g \in C_0(G)$ and $\|f * g\|_\infty \leqslant \|f\|_p \|g\|_q$.*

Proof. First, note that

$$|(f * g)(x)| \leqslant \int |f(y)| |g(y^{-1}x)| dy$$

$$\leqslant \|f\|_p \left(\int |g(y^{-1}x)|^q dy \right)^{1/q}$$

$$= \|f\|_p \|g\|_q$$

For each x and y, applying Hölder's inequality, we find

$$|(f * g)(x) - (f * g)(y)| \leqslant \int |f(z)| |g(z^{-1}x) - g(z^{-1}y)| dz$$

$$\leqslant \|f\|_p \left(\int |g(z^{-1}x) - g(z^{-1}y)|^q dz \right)^{1/q}$$

$$= \|f\|_p \left(\int |g(z) - g(zx^{-1}y)|^q dz \right)^{1/q}$$

$$= \|f\|_p \|g - R_{x^{-1}y} g\|_q$$

801

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 55.10. *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 55.11. *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 $z \to 1$. $\square$

802

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 55.12. *For each neighbourhood U of the origin, pick a function $f_U \in (L^1)^+(G)$, with $\int \phi_U = 1$, $supp(f_U) \subset U$. Then if g is any function in $L^p(G)$,*

$$\|f_U * g - g\|_p \to 0$$

where we assume g is left uniformly continuous if $p = \infty$, and if f_U is viewed as a net with neighbourhoods ordered by inclusion. If in addition $f_U(x) = f_U(x^{-1})$, then $\|g * f_U - g\|_p \to 0$, where g is right uniformly continuous for $p = \infty$.

Proof. Let us first prove the theorem for $p \neq \infty$. If $g \in C_c(G)$ is supported on a compact K, and if U is small enough that $|g(y^{-1}x) - g(x)| < \varepsilon$ for $y \in U$, then because $\int_U f_U(y) = 1$, and by applying Minkowski's inequality, we find

$$\|f_U * g - g\|_p = \left(\int \left| \int f_U(y)[g(y^{-1}x) - g(x)]dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int f_U(y) \left(\int |g(y^{-1}x) - g(x)|^p dx \right)^{1/p} dy$$

$$\leqslant 2\mu(K)\varepsilon \int f_U(y)dy \leqslant 2\mu(K)\varepsilon$$

Results are then found for all of L^p by taking limits. If g is left uniformly continuous, then we may find U such that $|g(y^{-1}x) - g(x)| < \varepsilon$ for $y \in U$ then

$$|(f_U * g - g)(x)| = \left| \int f_U(y)[g(y^{-1}x) - g(x)] \right| \leqslant \varepsilon$$

For right convolution, we find that for $g \in C_c(G)$, where $|g(xy) - g(x)| < \varepsilon$ for

$y \in U$, then

$$\|g * f_U - g\|_p = \left(\int \left| \int g(y) f_U(y^{-1}x) - g(x) dy \right|^p dx \right)^{1/p}$$

$$= \left(\int \left| \int [g(xy) - g(x)] f_U(y) dy \right|^p dx \right)^{1/p}$$

$$\leqslant \int \left(\int |g(xy) - g(x)|^p dx \right)^{1/p} f_U(y) dy$$

$$\leqslant \mu(K)\varepsilon \int f_U(y)(1 + \Delta(y)) dy$$

$$= \mu(K)\varepsilon + \mu(K)\varepsilon \int f_U(y)\Delta(y) dy$$

We may always choose U small enough that $\Delta(y) < \varepsilon$ for $y \in U$, so we obtain a complete estimate $\mu(K)(\varepsilon + \varepsilon^2)$. If g is right uniformly continuous, then choosing U for which $|g(xy) - g(x)| < \varepsilon$, then

$$|(g * f_U - g)(x)| = \left| \int [g(xy) - g(x)] f_U(y) dy \right| \leqslant \varepsilon$$

We will always assume from hereon out that the approximate identities in $L^1(G)$ are of this form. $\qquad\square$

We have already obtained enough information to characterize the closed ideals of $L^1(G)$.

Theorem 55.13. *If V is a closed subspace of $L^1(G)$, then V is a left ideal if and only if it is closed under left translations, and a right ideal if and only if it is closed under right translations.*

Proof. If V is a closed left ideal, and f_U is an approximate identity at the origin, then for any g,

$$\|(L_z f_U) * g - L_z g\|_1 = \|L_z(f_U * g - g)\|_1 = \|f_U * g - g\| \to 0$$

so $L_z g \in V$. Conversely, if V is closed under left translations, $g \in L^1(G)$, and $f \in V$, then

$$g * f = \int g(y) L_{y^{-1}} f \, dy$$

which is in the closed linear space of the translates of f. Right translation is verified very similarily. $\qquad\square$

55.4 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 55.14. *For any* $0 < \theta < 1$, *and* $0 < p, q \leqslant \infty$. *If we define*

$$1/r_\theta = (1 - \theta)/p + \theta/q$$

to be the inverse interpolation of the two numbers. Then

$$\|f\|_{r_\theta} \leqslant \|f\|_p^{1-\theta} \|f\|_q^\theta$$

Proof. We apply Hölder's inequality to find

$$\|f\|_{r_\theta} \leqslant \|f\|_{p/(1-\theta)} \|f\|_{q/\theta} = \left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \left(\int |f|^{q/\theta} \right)^{\theta/q}$$

so it suffices to prove $\|f\|_{p/(1-\theta)} \leqslant \|f\|_p^{1-\theta}$, $\|f\|_{q/\theta} \leqslant \|f\|_q^\theta$.

The map $x \mapsto x^p$ is concave for $0 < p < 1$, so we may apply Jensen's inequality in reverse to conclude

$$\left(\int |f|^{p/(1-\theta)} \right)^{(1-\theta)/p} \leqslant \left(\int |f|^p \right)^{1/p}$$

$\square$

The Riesz Thorin interpolation theorem then implies $L^p * L^q \subset L^r$, for $p^{-1} + q^{-1} = 1 + r^{-1}$. However, these estimates only guarantee $L^1(G)$ is closed under convolution. If G is compact, then $L_p(G)$ is closed under convolution for all p (TODO). The L_p conjecture says that this is true if and only if G is compact. This was only resolved in 1990.

55.5 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 55.15. *If a σ compact topological group G has a transitive topological action on X, and $x \in X$, then the continuous bijection from G/G_x to X is a homeomorphism.*

Proof. It suffices to show that the map $\phi : G \to X$ is open, and we need only verify this for the neighbourhood basis of compact neighbourhoods V of the origin by properties of the action. G is covered by countably many translates $y_1 V, y_2 V, \dots$, and since each $\phi(y_k V) = y_k \phi(V)$ is closed (compactness), we conclude that $y_k \phi(V)$ has non-empty interior for some y_k, and hence $\phi(V)$ has a non-empty interior point $\phi(y_0)$. But then for any $y \in V$, y is in the interior of $\phi(y V y_0^{-1}) \subset \phi(V V y_0^{-1})$, so if we fix a compact U, and find V with $V^3 \subset U$, we have shown $\phi(U)$ is open in X. $\square$

We shall say a space X is homogenous if it is homeomorphic to G/H for some group action of G over X. The H depends on our choice of basepoint x, but only up to conjugation, for if if we switch to a new basepoint y, and c maps x to y, then $ax = x$ holds if and only if $cac^{-1} y = y$. The question here is to determine whether we have a G-invariant measure on X. This is certainly not always possible. If we had a measure on $\mathbb{R}$ invariant under the affine maps $ax + b$, then it would be equal to the Haar measure by uniqueness, but the Haar measure is not invariant under dilation $x \mapsto ax$.

Let G and H have left Haar measures μ and ν respectively, denote the projection of G onto G/H as $\pi : G \to G/H$, and let Δ_G and Δ_H be the respective modular functions. Define a map $P : C_c(G) \to C_c(G/H)$ by

$$(Pf)(Hx) = \int_H f(xy) d\nu(y) = \int_H$$

this is well defined by the invariance properties of ν. Pf is obviously continuous, and $\operatorname{supp}(Pf) \subset \pi(\operatorname{supp}(f))$. Moreover, if $\phi \in C(G/H)$ we have

$$P((\phi \circ \pi) \cdot f)(Hx) = \phi(xH) \int_H f(xy) d\nu(y)$$

so $P((\phi \circ \pi) \cdot f) = \phi P(f)$.

Lemma 55.16. *If E is a compact subset of G/H, there is a compact $K \subset G$ with $\pi(K) = E$.*

Proof. Let V be a compact neighbourhood of the origin, and cover E by finitely many translates of $\pi(V)$. We conclude that $\pi^{-1}(E)$ is covered by finitely many of the translates, and taking the intersections of these translates with $\pi^{-1}(E)$ gives us the desired K. $\square$

806

Lemma 55.17. *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 55.18. *If $\phi \in C_c(G/H)$, there is $f \in C_c(G)$ with $Pf = \phi$, and $\pi(\operatorname{supp} f) = \operatorname{supp}(\phi)$, and also $f \geqslant 0$ if $\phi \geqslant 0$.*

Proof. There exists $g \geqslant 0$ in $C_c(G/H)$ with $Pg = 1$ on $\operatorname{supp}(\phi)$, 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 55.19. *There is a G invariant measure ψ on G/H if and only if $\Delta_G = \Delta_H$ when restricted to H. In this case, the measure is unique up to a common factor, and if the factor is chosen, we have*

$$\int_G f d\mu = \int_{G/H} Pf d\psi = \int_{G/H} \int_H f(xy) d\nu(y) d\psi(xH)$$

Proof. Suppose ψ existed. Then $f \mapsto \int Pf d\psi$ is a non-zero left invariant positive linear functional on G/H, so $\int Pf d\psi = c \int f d\mu$ for some $c > 0$. Since $P(C_c(G)) = C_c(G/H)$, we find that ψ is determined up to a constant factor. We then compute, for $y \in H$,

$$\Delta_G(y) \int f(x) d\mu(x) = \int f(xy^{-1}) d\mu(x)$$

$$= \int_{G/H} \int_H f(xzy^{-1}) d\nu(z) d\psi(xH)$$

$$= \Delta_H(y) \int_{G/H} \int_H f(xz) d\nu(z) d\psi(xH)$$

$$= \Delta_H(y) \int f(x) d\mu(x)$$

807

Hence $\Delta_G = \Delta_H$. Conversely, suppose $\Delta_G = \Delta_H$. First, we claim if $f \in C_c(G)$ and $Pf = 0$, then $\int f d\mu = 0$. Indeed if $P\phi = 1$ on $\pi(\text{supp} f)$ then

$$0 = Pf(xH) = \int_H f(xy) d\nu(y) = \Delta_G(y^{-1}) \int_H f(xy^{-1}) d\nu(y)$$

so

$$\begin{aligned}
0 &= \int_G \int_H \Delta_G(y^{-1}) \phi(x) f(xy^{-1}) d\nu(y) d\mu(x) \\
&= \int_H \int_G \phi(xy) f(x) d\mu(x) d\nu(y) \\
&= \int_G P\phi(xH) f(x) d\mu(x) \\
&= \int_G f(x) d\mu(x)
\end{aligned}$$

This implies that if $Pf = Pg$, then $\int_G f = \int_G g$. Thus the map $Pf \mapsto \int_G f$ is a well defined G invariant positive linear functional on $C_c(G/H)$, and we obtain a Radon measure from the Riesz representation theorem. $\square$

If H is compact, then Δ_G and Δ_H are both continuous homomorphisms from H to $\mathbb{R}^+$, so Δ_G and Δ_H are both trivial, and we conclude a G invariant measure exists on G/H.

Chapter 56

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 $\mathbb{R}$ are of the form $t \mapsto e(t\xi)$, for $\xi \in \mathbb{R}$. To see this, let $e : \mathbb{R} \to \mathbf{T}$ 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 $\xi \in \mathbb{R}$.

Example. *Consider the group $\mathbb{R}^+$ of positive real numbers under multiplication. The map $x \mapsto \log x$ is an isomorphism from $\mathbb{R}^+$ and $\mathbb{R}$, so that every character on $\mathbb{R}^+$ is of the form $e(s \log x) = x^{is}$, for some $s \in \mathbb{R}$. The character group is then $\mathbb{R}$, 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 56.1. *For any character $\phi : G \to \mathbf{C}$, the map*

$$\varphi(f) = \int \frac{f(x)}{\phi(x)}dx$$

is a non-zero character on the convolution algebra $L^1(G)$, and all characters arise this way.

Proof. The induced map is certainly linear, and

$$\begin{aligned} \varphi(f * g) &= \int \int \frac{f(y)g(y^{-1}x)}{\phi(x)}dydx \\ &= \int \int \frac{f(y)g(x)}{\phi(y)\phi(x)}dydx \\ &= \int \frac{f(y)}{\phi(y)}dy \int \frac{g(x)}{\phi(x)}dx \end{aligned}$$

810

Since ϕ is continuous, there is a compact subset K of G where $\phi > \varepsilon$ for some $\varepsilon > 0$, and we may then choose a positive f supported on K in such a way that $\varphi(f)$ is non-zero.

The converse results from applying the duality theory of the L^p spaces. Any character on $L^1(G)$ is a linear functional, hence is of the form

$$f \mapsto \int f(x)\phi(x)dx$$

for some $\phi \in L^\infty(G)$. Now

$$\int\int f(y)g(x)\phi(yx)dydx = \int\int f(y)g(y^{-1}x)\phi(x)dydx$$

$$= \int f(x)\phi(x)dx \int g(y)\phi(y)dy$$

$$= \int f(x)g(y)\phi(x)\phi(y)dxdy$$

Since this holds for all functions f and g in $L^1(G)$, we must have $\phi(yx) = \phi(x)\phi(y)$ almost everywhere. Also

$$\int \varphi(f)g(y)\phi(y)dy = \varphi(f * g)$$

$$= \int\int g(y)f(y^{-1}x)\phi(x)dydx$$

$$= \int\int (L_{y^{-1}}f)(x)g(y)\phi(x)dydx$$

$$= \int \varphi(L_{y^{-1}}f)g(y)dy$$

which implies $\varphi(f)\phi(y) = \varphi(L_{y^{-1}}f)$ almost everywhere. Since the map $\varphi(L_{y^{-1}}f)/\varphi(f)$ is a uniformly continuous function of y, ϕ is continuous almost everywhere, and we might as well assume ϕ is continuous. We then conclude $\phi(xy) = \phi(x)\phi(y)$. Since $\|\phi\|_\infty = 1$ (this is the norm of any character operator on $L^1(G)$), we find ϕ maps into **T**, for if $\|\phi(x)\| < 1$ for any particular x, $\|\phi(x^{-1})\| > 1$. □

Thus there is a one-to-one correspondence with $\Gamma(G)$ and $\Gamma(L^1(G))$, which implies a connection with the Gelfand theory and the character theory of locally compact groups. This also gives us a locally compact topological structure on

$\Gamma(G)$, induced by the Gelfand representation on $\Gamma(L^1(G))$. A sequence $\phi_i \to \phi$ if and only if

$$\int \frac{f(x)}{\phi_i(x)} dx \to \int \frac{f(x)}{\phi(x)} dx$$

for all functions $f \in L^1(G)$. This actually makes the map

$$(f, \phi) \mapsto \int \frac{f(x)}{\phi(x)} dx$$

a jointly continuous map, because as we verified in the proof above,

$$\hat{f}(\phi)\phi(y) = \widehat{L_y f}(\phi)$$

And the map $y \mapsto L_y f$ is a continuous map into $L^1(G)$. If $K \subset G$ and $C \subset \Gamma(G)$ are compact, this allows us to find open sets in G and $\Gamma(G)$ of the form

$$\{\gamma : \|1 - \gamma(x)\| < \varepsilon \text{ for all } x \in K\} \quad \{x : \|1 - \gamma(x)\| < \varepsilon \text{ for all } \gamma \in C\}$$

And these sets actually form a base for the topology on $\Gamma(G)$.

Theorem 56.2. *If G is discrete, $\Gamma(G)$ is compact, and if G is compact, $\Gamma(G)$ is discrete.*

Proof. If G is discrete, then $L^1(G)$ contains an identity, so $\Gamma(G) = \Gamma(L^1(G))$ is compact. Conversely, if G is compact, then it contains the constant 1 function, and

$$\hat{1}(\phi) = \int \frac{dx}{\phi(x)}$$

And

$$\frac{1}{\phi(y)}\hat{1}(\phi) = \int \frac{dx}{\phi(yx)} = \int \frac{dx}{\phi(x)} = \hat{1}(\phi)$$

So either $\phi(y) = 1$ for all y, and it is then verified by calculation that $\hat{1}(\phi) = 1$, or $\hat{1}(\phi) = 0$. Since $\hat{1}$ is continuous, the trivial character must be an open set by itself, and hence $\Gamma(G)$ is discrete. $\square$

Given a function $f \in L^1(G)$, we may take the Gelfand transform, obtaining a function on $C_0(\Gamma(L^1(G)))$. The identification then gives us a function on $C_0(\Gamma(G))$, if we give $\Gamma(G)$ the topology induced by the correspondence (which also makes $\Gamma(G)$ into a topological group). The formula is

$$\hat{f}(\phi) = \phi(f) = \int \frac{f(x)}{\phi(x)}$$

812

This gives us the classical correspondence between $L^1(\mathbf{T})$ and $C_0(\mathbf{Z})$, and $L^1(\mathbb{R})$ and $C_0(\mathbb{R})$, 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

- $\widehat{f * g} = \hat{f}\hat{g}$ (The transform is a homomorphism).

- $\mathcal{F}$ is norm decreasing and therefore continuous: $\|\hat{f}\|_\infty \leqslant \|f\|_1$.

- If G is unimodular, and $\gamma \in \Gamma(G)$, then $(f * \gamma)(x) = \gamma(x)\hat{f}(\gamma)$.

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

$$\hat{\mu}(\phi) = \int \frac{dx}{\phi(x)}$$

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 $\hat{\mu}$ 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

$$(L_\nu\hat{\mu} - \hat{\mu})(\phi) = \int \frac{1 - \nu(x)}{\nu(x)\phi(x)} d\mu(x)$$

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

$$|(L_\nu\hat{\mu} - \hat{\mu})(\phi)| \leqslant 2|\mu|(K^c) + \varepsilon\|\mu\| \leqslant \varepsilon(2 + \|\mu\|)$$

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 $M(\mathbb{R}^3)$. 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 $\mathbb{R}^3$, 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 56.3. $L^1(G)$ *is weak-$*$ dense in* $M(G)$.

Proof. First, note that the Dirac delta function can be weak-$*$ approximated by elements of $L^1(G)$, since we have an approximate identity in the space.

First, note that if $\mu_i \to \mu$, then $\mu_i * \nu \to \mu * \nu$, because

$$\int f d(\mu_i * \nu) = \int \int f(xy) d\mu_i(x) d\nu(y)$$

The functions $y \mapsto \int f(xy) d\mu_i(x)$ converge pointwise to $\int f(xy) d\mu(y)$. Since

$$\left| \int f(xy) d\mu_i(x) \right| \leqslant \|f\|_1 \|\mu_i\|$$

If i is taken large enough that $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

814

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

$$\int \left| \int \frac{d\mu(x)}{\phi(x)} \right| d\phi \leq \int \int \|\mu\|$$

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

$$\mathcal{F}^{-1}(\{a_k\}) = \sum a_k e_k(t) \quad \mathcal{F}^{-1}(f)(x) = \int f(t)e(xt)$$

which reverses the fourier transform on $\mathbf{T}$ and $\mathbb{R}$ respectively, on a certain subclass of L^1. One of the challenges of Harmonic analysis is trying to find where this holds for the general class of measurable functions.

The first problem is to determine surjectivity. We denote by $A(G)$ the space of all continuous functions which can be represented as the fourier transform of some function in $L^1(G)$. It is to even determine $A(\mathbf{T})$, the most basic example. $A(G)$ always separates the points of $\Gamma(G)$, by Gelfand theory, and if G is unimdoular, then it is closed under conjugation. If we let $g(x) = \overline{f(x^{-1})}$, we find

$$\mathcal{F}(g)(\phi) = \int \frac{g(x)}{\phi(x)}dx = \overline{\int \frac{f(x^{-1})}{\phi(x^{-1})}dx} = \overline{\int \frac{f(x)}{\phi(x)}dx} = \overline{\mathcal{F}(f)(\phi)}$$

so that by the Stone Weirstrass theorem $A(G)$ is dense in $C_0(\Gamma(L^1(G)))$.

Chapter 57

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 $L^1(\mathbb{R})$ are the maps $f \mapsto \widehat{f}(x)$, so that the Gelfand space of $\mathbb{R}$ is $\mathbb{R}$, 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 58

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 59

Interpolation of Besov and Sobolev spaces

An important class of operators arise as singular integrals, that is, they arise as convolution operators T given by $T(f) = f * K$, where K is an appropriate distribution. Taking Fourier transforms, these operators can also be defined by $\widehat{T(f)} = \hat{f}\hat{K}$. The function $\hat{K}$ is known as a **Fourier multiplier**, because it operates by multiplication on the frequencies of the function f. We say $\hat{K}$ is a **Fourier multiplier on** $L^p(\mathbb{R}^n)$ if T is a bounded map from $S(\mathbb{R}^n)$ to $L^p(\mathbb{R}^n)$, under the L^p norms. Such maps clearly extend uniquely to maps from $L^p(\mathbb{R}^n)$ to $L^p(\mathbb{R}^n)$, 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*

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbb{R}^n)} \frac{|(f * K)(0)|}{\|f\|_\infty}$$

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

$$(f * K)(0) = \int f(-y) \, d\mu(y)$$

818

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

$$\|K\|_\infty = \sup_{f \in L^\infty(\mathbb{R}^n)} \frac{\left|\int f(y)\, d\mu(y)\right|}{\|f\|_\infty} = \|\mu\|_1$$

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 59.1. *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 $\|K\|_p = \|K\|_q$. $\qquad\square$

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 $L^\infty(\mathbb{R}^n)$, 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 59.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a surjective affine transformation. Then the endomorphism T^* on $M_p(\mathbb{R}^n)$ 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 59.3. *There exists a Schwartz function φ on $\mathbb{R}^n$ 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 59.4 (Calderon-Zygmund Decomposition). *Let $f \in L^1(\mathbb{R}^n)$, and $\sigma > 0$. Then there are pairwise almost disjoint cubes $I_1, I_2, \ldots$ with edges parallel to the coordinate axis and*

$$\sigma < \frac{1}{|I_n|} \int_{I_n} |f(x)|\, dx \leqslant 2^n \sigma$$

and with $|f(x)| \leqslant \sigma$ for almost all x outside these cubes.

Theorem 59.5 (The Mihlin Multiplier Theorem). *Let m be a bounded function on $\mathbb{R}^n$ which is smooth except possibly at the origin, such that*

$$\sup_{\substack{\xi \in \mathbb{R}^n \\ |\alpha| \leqslant L}} |\xi|^{|\alpha|} |(D^\alpha m)(x)| < \infty$$

Then m is an L^p Fourier multiplier for $1 < p < \infty$.

59.1   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 $s \in \mathbb{R}$, 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 $2^{n-1} \leqslant |\xi| \leqslant 2^{n+1}$

59.2 Proof of The Projection Result

As with Marstrand's projection theorem, we require an energy integral variant. Rather than considering the Riesz kernel on $\mathbb{R}^n$, 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 $\mathbb{R}^n$, then $\mu * K_\alpha \in L^\infty(\mathbb{R}^d)$ because μ cancels out the singular part of K_α. Assuming $\beta < d$, we conclude $\mu * K_\alpha \in L^1(\mathbb{R}^d)$. Applying interpolation (TODO: Which interpolation), we conclude that $\nu * K_\rho$

Chapter 60

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 60.1 (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 60.2 (Roth - 1956). *If A is a set of integers in $[1, N]$ which is free of three term arithmetic progressions, then $|A| = O(N/ \log \log N)$.*

Szemerédi proved that if A is free of k term arithmetic progressions, $|A| = o(N)$. If Erdös Turan, if $\sum_{x \in X} 1/x$ diverges, then X contains arbitrarily long arithmetic progressions. For now, we'll restrict our attention to three term arithmetic progressions. Heath and Brown showed that three term arithmetic progresisons

are $O(N/(\log N)^c)$ for some constant c. In 2016, the best known bound was given by Bloom, given $O(N(\log \log N)^4/ \log N)$.

One way we can simplify our problem is to note that avoiding three term arithmetic progressions is a local issue, so we can embed $[1, N]$ in $\mathbf{Z}/M\mathbf{Z}$ for suitably large M, and we lose none of the problems we had over the integers. A heuristic is that it is easier to solve these kind of problems in $\mathbf{F}_p^n$, where p is small and n is large, which should behave like $\{1, \ldots, p^n\}$. This leads naturally to the cap set problem.

Theorem 60.3 (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 60.4 (Meschulam - 1995). *Let $A \subset \mathbf{F}_3^n$ be a cap set. Then $|A| = O(3^n/n)$. This is analogous to a $N/\log N$ case over the integers, giving evidence that the finite field case is easier.*

In 2012, Bateman and Katz showed $|A| = O(3^n/n^{1+\varepsilon})$ for some $c > 0$. This was a difficult proof. In 2016, there was a more significant breakthrough, which gave an easy proof using the polynomial method of an exponentially small bound of c^n, where $c < 4$, over $\mathbf{Z}/4\mathbf{Z}$, and a week later Ellenberg-Gijswijt used this argument in the $\mathbf{F}_3$ case to prove that if A is a capset in $\mathbf{F}_3$, then $|A| = O(c^n)$, for $c = 2.7551 \ldots$.

The idea of the polynomial method is to take combinatorial information about some set, encode it as some algebraic structura information, and then apply the theory of polynomials to encode this algebraic information and use it to limit and enable certain properties to occur.

If V is the space of polynomials of degree d vanishing on a set A, then we know $\dim V \geqslant \dim \mathcal{P}_d - |A|$. This gives a lower bound on the size of A, whereas we want a lower bound. To get an upper bound, we take $|A|^c$ instead, which shows

$$\dim V \geqslant \dim \mathcal{P}_d + |A| - 3^n$$

whichs gives $|A| \leqslant 3^n + \dim V - \dim \mathcal{P}_d$. Now using linear algebra, we can find a polynomial P vanishing on A^c with support of cardinality greater than or equal to $\dim V$, hence

$$|A| \leqslant 3^n - \dim \mathcal{P}_d + \max |\mathrm{supp}(P)|$$

It follows that A is a cap set if and only if $x + y = 2z$, or $x + y + z = 0$ holds if and only if $x = y = z$. This is an algebraic property which says directly that A has no nontrivial three term arithmetic progressions. Thus for any $a_1, \ldots, a_m \in A$, $P(-a_i - a_j) = 0$ when $i \neq j$. Equivalently, this means $P(-a_i - a_j) \neq 0$ when $i = j$. This suggests we consider the $|A|$ by $|A|$ matrix M with $M_{ij} = P(-a_i - a_j)$. This is a diagonal matrix, with $M_{ii} = P(a_i)$. Thus the rank of this matrix is the dimension of the support of P, so it suffices to upper bound the rank of M. The key observation, where we now explicitly employ the fact that P is a polynomial, is that $P(-x - y)$ is a polynomial in $2n$ variables $x, y \in \mathbf{F}_3^n$,

Part IX

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. The theory of decoupling was initiated by Wolff in the early 2000s, and Laba-Wolff, Laba-Pramanik, Pramanik-Seeger, and Garrigos and Seeger in the 2000s. But the theory was brought to the forefront of harmonic analysis by Bourgain, Demeter, and Guth in the mid 2010s.

Chapter 61

The General Framework

In any norm space X, given $f = f_1 + \cdots + f_N$, let's say, with $\|f_1\|_X, \ldots, \|f_N\|_X \leqslant 1$, then the triangle inequality implies that

$$\|x\|_X \leqslant N.$$

In general, such a result is sharp, i.e. if all the $f_1, \ldots, f_N \in X$ are equal to one another, in which case we see *constructive interference*. On the other hand, one can often significantly improve this bound if one knows that the functions $f_1, \ldots, f_N \in X$ do not interact with one another to a reasonable extent, i.e. if we assume some *orthogonality*.

Classically, this situation comes about when X is a Hilbert space. We can then write

$$\|f\|_X^2 = \sum_i \|f_i\|_X^2 + 2\sum_{i<j} \langle f_i, f_j \rangle.$$

If $f_1, \ldots, f_N$ are pairwise orthogonal, i.e. $\langle f_i, f_j \rangle = 0$ for $i \neq j$, then we obtain that

$$\|f\|_X = N^{1/2}.$$

Thus we see *square root cancellation*. For general values $f_1, \ldots, f_N \in X$, this calculation shows us that

$$\|f\|_X = \left(\sum_i \|f_i\|_X^2 \right)^{1/2}.$$

More generally, if one assumes an *almost orthogonality condition*. For instance, if we assume that for each i, there are at most $O(1)$ indices j with $\langle f_i, f_j \rangle \neq 0$, and

if we rearrange our vectors so that

$$\|f_1\|_X \geqslant \cdots \geqslant \|f_N\|_X,$$

then for any i and j, we can use the trivial Cauchy-Schwartz bound

$$|\langle f_i, f_j \rangle| \leqslant \|f_i\|_X \|f_j\|_X \leqslant \|f_i\|_X^2$$

to conclude that

$$\begin{aligned}
\|f\|_X^2 &= \sum_i \|f_i\|_X^2 + 2 \sum_{i<j} \|f_i\|_X \|f_j\|_X \\
&\leqslant \sum_i \|f_i\|_X^2 + O(1) \sum_i \|f_i\|_X^2 \\
&\lesssim \sum_i \|f_i\|_X^2,
\end{aligned}$$

and thus

$$\|f\|_X \lesssim \left(\sum_i \|f_i\|_X^2 \right)^{1/2}.$$

This situation, in which uncorrelated vectors have a *square root cancellation inequality*, occurs often in harmonic analysis. For instance, in $X = L^2(\mathbb{R}_x^d)$, harmonic analysts often decompose $\mathbb{R}_\xi^d$ down into a union of finitely overlapping pieces $\{\theta\}$, consider an associated partition of unity $\{\chi_\theta\}$ adapted to this cover, and then performing a frequency decomposition, writing a general $f \in L^2(\mathbb{R}_x^d)$ as

$$f = \sum_\theta f_\theta,$$

where $\widehat{f_\theta} = \chi_\theta \widehat{f}$. The fact that these pieces are finitely overlapping implies they will almost orthogonal, so that

$$\|f\|_X \lesssim \left(\sum_\theta \|f_\theta\|_X^2 \right)^{1/2}.$$

Decoupling theory is the study of when one can obtain square root cancellation bounds for more general norm spaces X, most importantly, when X is an L^p space for $p > 2$. Such estimates are called *decoupling estimates*

One might expect that decoupling is connected to reverse square function estimates of the form

$$\|f\|_X \lesssim \left\| \left(\sum_i |f_i|^2 \right)^{1/2} \right\|_X .$$

A classical example of such an estimate is Littlewood-Paley theory. In the region $p > 2$, reverse square function estimates are stronger than decoupling estimates. But the advantage of decoupling is that they same to be easier to *iterate*. If we decompose $f = \sum f_i$, and then perform a further decomposition $f_i = \sum f_{ij}$, and we are able to obtain inequalities of the form

$$\|f\|_X \lesssim \left(\sum_i \|f_i\|_X^2 \right)^{1/2}$$

and

$$\|f_i\|_X \lesssim \left(\sum_j \|f_{ij}\|_X^2 \right)^{1/2},$$

then by combining these bounds we automatically obtain a bound of the form

$$\|f\|_X \lesssim \left(\sum_{i,j} \|f_{ij}\|_X^2 \right)^{1/2}.$$

Thus decoupling often works well with *induction on scales*.

Remark. We are primarily interested in studying decoupling in $L^p(\Omega)$. However, in this case we should only expect decoupling to occur when $p \geqslant 2$. Functions in L^p spaces are 'most orthogonal' when their supports are disjoint. If a family of functions $\{f_i\}$ have disjoint supports, then we actually have

$$\|f_1 + \cdots + f_N\|_{L^p(\Omega)} = \left(\|f_1\|_{L^p(\Omega)}^p + \cdots + \|f_N\|_{L^p(\Omega)}^p \right)^{1/p}.$$

For $p < 2$, we have a sharp inequality

$$\left(\|f_1\|_{L^p(\Omega)}^p + \cdots + \|f_N\|_{L^p(\Omega)}^p \right)^{1/p} \leqslant N^{1/p-1/2} \left(\|f_1\|_{L^p(\Omega)}^2 + \cdots + \|f_N\|_{L^2(\Omega)}^2 \right)^{1/2},$$

so we do not have square root cancelation in this setting. On the other hand, when $p \geqslant 2$ we have

$$\left(\|f_1\|_{L^p(\Omega)}^p + \cdots + \|f_N\|_{L^p(\Omega)}^p \right)^{1/p} \leqslant \left(\|f_1\|_{L^p(\Omega)}^2 + \cdots + \|f_N\|_{L^2(\Omega)}^2 \right)^{1/2}$$

and so we obtain a decoupling inequality for functions with disjoint support.

The most basic case where we can obtain decoupling in $L^p(\Omega)$ for $p > 2$ is when p is an even integer. As a basic example, say a family of functions $f_1, \ldots, f_N \in L^4(\Omega)$ are *biorthogonal* if $\{f_i f_j : i < j\}$ forms an orthogonal family in $L^2(\Omega)$.

Theorem 61.1. *If $f_1, \ldots, f_N$ are a biorthogonal family, 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. We write

$$
\begin{aligned}
\|f_1 + \cdots + f_N\|_{L^4(\Omega)}^2 &= \left\| (f_1 + \cdots + f_N)^2 \right\|_{L^2(\Omega)} \\
&= \left\| \sum_{1 \leqslant i, j \leqslant N} f_i f_j \right\|_{L^2(\Omega)} \\
&\lesssim \sum_{i=1}^{N} \|f_i^2\|_{L^2(\Omega)} + \left\| \sum_{1 \leqslant i < j \leqslant N} f_i f_j \right\|_{L^2(\Omega)}
\end{aligned}
$$

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 in square roots completes the claim. $\square$

Remark. The calculation above shows a decoupling inequality still holds if the family $f_i f_j$ is almost biorthogonal, e.g. if each of the elements of the family $\{f_i f_j\}$ are orthogonal to all but $O_\varepsilon(N^\varepsilon)$ of the other elements of the family.

Remark. Similarily, if $f_1, \ldots, f_N \in L^6(\Omega)$ are chosen to be *triorthogonal*, in the sense that the family of functions $\{f_i f_j f_k\}$ are 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 $f_1, \ldots, f_N$ with disjoint Fourier supports in $L^p(\mathbb{R}^d)$, where $p \geqslant 2$. It is then certainly true that

$$\|f_1 + \cdots + f_N\|_{L^2(\mathbb{R}^d)} \leqslant \left(\|f_1\|_{L^2(\mathbb{R}^d)}^2 + \cdots + \|f_N\|_{L^2(\mathbb{R}^d)}^2 \right)^{1/2}.$$

However, for $p > 2$ having disjoint Fourier supports does not immediately imply decoupling, because constructive interference can still ocur in the L^p norm in a way not detected in the L^2 norm, and we require addition features of the family in order to guarantee that this constructive interference does not occur.

Theorem 61.2. *If $f_1, \ldots, f_N \in L^p(\mathbb{R}^d)$, have disjoint Fourier support, and $2 \leqslant p \leqslant \infty$, then*

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R}^d)} \leqslant N^{1/2 - 1/p} \left(\|f_1\|_{L^p(\mathbb{R}^d)}^2 + \cdots + \|f_N\|_{L^p(\mathbb{R}^d)}^2 \right)^{1/2}.$$

Proof. For $p = \infty$, we have the trivial inequality

$$\|f_1 + \cdots + f_N\|_{L^\infty(\mathbb{R}^d)} \leqslant \|f_1\|_{L^\infty(\mathbb{R}^d)} + \cdots + \|f_N\|_{L^\infty(\mathbb{R}^d)}$$
$$\leqslant N^{1/2} \left(\|f_1\|_{L^\infty(\mathbb{R}^d)}^2 + \cdots + \|f_N\|_{L^\infty(\mathbb{R}^d)}^2 \right)^{1/2}.$$

By a density argument (applying a cutoff in frequency space), we may assume without loss of generality that the functions $\{f_i\}$ are Schwartz. But then by orthogonality of the Fourier transform implies the functions themselves are orthogonal, so we have

$$\|f_1 + \cdots + f_N\|_{L^2(\mathbb{R}^d)} \leqslant \left(\|f_1\|_{L^2(\mathbb{R}^d)}^2 + \cdots + \|f_N\|_{L^2(\mathbb{R}^d)}^2 \right)^{1/2}.$$

We can then interpolate with the case $p = \infty$. $\qquad\square$

In general, this result is optimal, as the next result shows one can have significant constructive interference for $p > 2$ if our functions are spaced about on a periodic family of frequencies.

Example. *Fix $\phi \in \mathcal{S}(\mathbb{R}^d)$ with $\phi(0) = 1$, and with Fourier support in $[0, 1]$. For each $k \in \{1, \ldots, N\}$, set $f_k = e^{2\pi i k x} \phi(x)$. Then f_k has Fourier support in $[2k, 2k + 1]$, and $f_k(0) = 1$. The uncertainty principle thus implies that the functions f_k are all roughly constant at a scale $O(1/N)$, which implies that in a ball of radius*

$O(1/N)$ around the origin, we have $f_1(x) + \cdots + f_n(x) \approx 1$. Thus we conclude that

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R})} \gtrsim N^{1-1/p}.$$

On the other hand, we have

$$\left(\|f_1\|^2_{L^p(\mathbb{R})} + \cdots + \|f_N\|^2_{L^p(\mathbb{R})} \right)^{1/2} \lesssim N^{1/2},$$

Thus

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R})} \gtrsim N^{1/2-1/p} \left(\|f_1\|^2_{L^p(\mathbb{R})} + \cdots + \|f_N\|^2_{L^p(\mathbb{R})} \right)^{1/2},$$

which shows our result is tight up to constants.

In the face of this result, we are interested in knowning, for a given family $\mathcal{S}$ of disjoint sets in $\mathbb{R}^d$, what the smallest constant $\mathrm{Dec}(\mathcal{S}, p)$ is for which it is true that if $f_1, \ldots, f_N$ have Fourier support on distinct regions $S_1, \ldots, S_N \in \mathcal{S}$, we have

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R}^d)} \leqslant \mathrm{Dec}(\mathcal{S}, p) \left(\|f_1\|^2_{L^p(\mathbb{R}^d)} + \cdots + \|f_N\|^2_{L^p(\mathbb{R}^d)} \right)^{1/2}.$$

Pure orthogonality gives $\mathrm{Dec}(\mathcal{S}, 2) = 1$. The triangle inequality gives $\mathrm{Dec}(\mathcal{S}, \infty) \leqslant \#(S)^{1/2}$, and this is actually sharp: we have $\mathrm{Dec}(\mathcal{S}, \infty)$

If each element of $\mathcal{S}$ contains a ball of radius δ, then by summing up modulated bump functions on these balls, we can find functions such that $\mathrm{Dec}(\mathcal{S}, \infty) \gtrsim \delta^d \#(S)^{1/2}$. Thus non-trivial decoupling in the framework we are considering is impossible in $L^\infty(\mathbb{R}^d)$, i.e. $\mathrm{Dec}(\mathcal{S}, \infty) \sim_\delta \#(S)^{1/2}$. In between, we can interpolate to get $\mathrm{Dec}(\mathcal{S}, p) \leqslant \#(S)^{1/2-1/p}$.

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 above. Below is a positive result for a particular family $\mathcal{S}$, easily proved using the biorthogonality arguments established above.

Theorem 61.3. *Suppose $\mathcal{S}$ is a family of sets in $\mathbb{R}^d$ such that for any four sets $S_1, S_2, S_3, S_4 \in \mathcal{S}$ with $\{S_1, S_2\} \neq \{S_3, S_4\}$, the sets $S_1 + S_2$ are disjoint from $S_3 + S_4$. Then if distinct sets $S_1, \ldots, S_N \in \mathcal{S}$ are selected from $\mathcal{S}$, and $f_1, \ldots, f_N$ are a family of Schwartz functions in $\mathbb{R}^d$ such that f_i has Fourier support in S_i for each i, we find''*

$$\|f_1 + \cdots + f_N\|_{L^4(\Omega)} \lesssim \left(\|f_1\|^2_{L^4(\Omega)} + \cdots + \|f_N\|^2_{L^4(\Omega)} \right)^{1/2}.$$

Remark. We say a set of integers $A \subset \{1, \ldots N\}$ 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 $S = \{[2k, 2k+1] : k \in A\}$ satisfies the constraints of the result above, and so we obtain 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

$$\| \sum_{k \in A} f_k \|_{L^4(\mathbb{R})} \lesssim \left(\sum_{k \in A} \|f_k\|_{L^4(\mathbb{R})}^2 \right)^{1/2}.$$

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

$$\left\| \sum_{k \in A} f_k \right\|_{L^4(\mathbb{R})} \gtrsim \frac{\#(A)^{1/2}}{N^{1/4}} \left(\sum_{k \in A} \|f_k\|_{L^4(\mathbb{R})}^2 \right)^{1/2}.$$

Combining this inequality with the decoupling inequality, we obtain an interesting number theory result: any Sidon set A, must satisfy the bound $\#(A) \lesssim N^{1/2}$. More generally, we can extend this result to show that any set $A \subset \{0, \ldots, N-1\}$ having no nontrivial solutions to the equation

$$x_1 + \cdots + x_m = y_1 + \cdots + y_m$$

should satisfy $\#(A) \lesssim N^{1/m}$.

Another example family of sets S where Decoupling occurs occurs in Littlewood-Paley theory.

Theorem 61.4. *Let S be the collection of all boxes in $\mathbb{R}^d$ of the form*

$$I_k = I_{\pm k_1} \times \cdots \times I_{\pm k_d}$$

such that $I_{+k_1} = [2^{k_1}, 2^{k_1+1}]$ and $I_{-k_1} = [-2^{k_1+1}, -2^{k_1}]$. Littlewood-Paley theory implies that if $S_1, \ldots, S_N \in S$ and $f_1, \ldots, f_N$ are Schwartz functions with f_i having Fourier support on S_i for each i, then for each $2 \leqslant p < \infty$,

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R}^d)} \sim_{p,d} \|f_i\|_{L^p(\mathbb{R}^d) l_N^2} \lesssim \|f_i\|_{l_N^2 L^p(\mathbb{R}^d)}.$$

This is precisely a decoupling inequality. More generally, whenever we have a reverse square function estimate

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R}^d)} \lesssim \|f_i\|_{L_x^p l_N^2},$$

we get a decoupling inequality by changing norms.

We say a set $\omega \subset \mathbb{R}^d$ is an *almost rectangular box* with sidelengths $L_1, \ldots, L_d$ if there is a rectangular box R with sidelengths $L_1, \ldots, L_d$ centered at the origin and $\theta \in \mathbb{R}^d$ such that $\theta + C^{-1}R \subset \theta \subset \theta + C \cdot R$ for some universal constant C. As a special case, we have *almost cubes* as well. For each $R \geqslant 1$, partition $[-1, 1]^{d-1}$ into $\sim R^{(d-1)/2}$ almost rectangular boxes of sidelength $R^{-1/2}$. If we consider the paraboloid $\mathbf{P}^{n-1} = \{\xi, |\xi|^2\}$, then we obtain a partition of $N(\mathbf{P}^{n-1} \cap [0, 1]^d, R^{-1})$ by a family $\Theta(1/R)$ of $R^{-1/2} \times R^{-1}$ almost rectangular boxes, by setting

$$\Theta(1/R) = \{(\omega \times \mathbb{R}) \cap N(\mathbf{P}^{n-1} \cap [0, 1]^d)\}$$

It is conjectured that for $p = 2n/(n-1)$, and for Schwartz functions $f_1, \ldots, f_N$ with Fourier support on distinct elements of $\Theta(1/R)$, we have

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R}^d)} \gtrsim_R \|f_i\|_{L^p(\mathbb{R}^d) \ell_N^2},$$

This reverse square function estimates is strong. In particular, it implies the restriction conjecture. But it also implies the decoupling estimate

$$\|f_1 + \cdots + f_N\|_{L^p(\mathbb{R}^d)} \lesssim_\varepsilon R^\varepsilon \left(\|f_1\|_{L^p(\mathbb{R}^d)} + \cdots + \|f_N\|_{L^p(\mathbb{R}^d)} \right)^{1/2}$$

. However, unlike the restriction conjecture, this problem is closed: we have a proof of this decoupling estimate not only when $p = 2d/(d-1)$, but even when $2 \leqslant p \leqslant 2(d+1)/(d-1)$.

61.1   Localized Estimates

Suppose $f_1, \ldots, f_N$ are Schwartz functions in $\mathbb{R}^d$ with disjoint Fourier supports, and $\Omega \subset \mathbb{R}^d$. 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 $\Omega = \{x \in \mathbb{R} : |x| \lesssim 1/N\}$, 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 $f_1, \ldots, f_N$, 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

$$w_I(x) = \left(1 + \frac{|x - x_0|}{R} \right)^{-M}$$

834

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 $\mathbb{R}^d$, 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

61.2   Vinogradov Systems

One long standing number theoretic conjecture that has been particularly amenable to the use of decoupling techniques is *Vinogradov's Conjecture*. Let $J_{s,k}(N)$ denote the number of solutions to the *system* of equations

$$x_1^i + \cdots + x_s^i = y_1^i + \cdots + y_s^i$$

for $1 \leqslant i \leqslant k$, where $x_1, \ldots, x_s, y_1, \ldots, y_s \in \{1, \ldots, N\}$. Our primary interest is determining the asymptotic growth in this quantity as $N \to \infty$. Setting $y_i = x_i$ gives a family of N^s solutions. Thus $J_{s,k}(N) \geqslant N^s$. On the other hand, dyadic pidgeonholing implies that there exists some family of integer tuples $I \subset [1, N] \times \cdots \times [1, N^k]$ and some integer M with $N^s = M\#(I)$ such that for each tuple $(r_1, \ldots, r_k) \in I$, there are $\Omega(M)$ tuples of values $(x_1, \ldots, x_s)$ such that for each $1 \leqslant i \leqslant k$,

$$x_1^i + \cdots + x_s^i = r_i.$$

For any i, $x_1^i + x_s^i$ takes on values between 1 and N^i. Since $1 \leqslant \#(I) \leqslant N^{k(k+1)/2}$, we have $N^{s-k(k+1)/2} \lesssim M \lesssim N^s$, and so

$$J_{s,k}(N) \gtrsim \#(I)M^2 \gtrsim N^{2s-k(k+1)/2},$$

a bound that is much tighter if the number of variables is large, i.e. $s \geqslant k(k+1)/2$. Vinogradov's mean value conjecture is that

$$J_{s,k}(N) \lesssim_{s,k,\varepsilon} N^\varepsilon (N^{2s-k(k+1)/2} + N^s),$$

for any $\varepsilon > 0$. One result of decoupling is a proof of this conjecture.

Remark. Another approach to the Vinogradov mean value conjecture is to use more number theoretic techniques, for instance, efficient congruencing. There has even been some feedback from this alternate approach in the field of decoupling, e.g. using efficient congruencing to obtain decoupling inequalities (see Guo-Li-Yung-Zorin Kranich, 2021).

Similar types of number theoretic problems had been studied using harmonic analysis techniques. Classically, Hardy and Littlewood studied the quantity $\mathrm{HL}_{k,s}(N)$, which counts the number of solutions to the *single equation*

$$x_1^k + \cdots + x_s^k = y_1^k + \cdots + y_s^k$$

with $x_1, \ldots, x_s, y_1, \ldots, y_s \in \{1, \ldots, N\}$. To study this quantity, Hardy and Littlewood introduced the function

$$f_{k,N}(x) = \sum_{n=1}^{N} e^{2\pi i n^k x}.$$

Orthogonality then implies that

$$\mathrm{HL}_{k,s}(N) = \|h_{k,N}\|_{L^{2s}[0,1]}.$$

Bounding the nature of the function $h_{k,N}$ then involves Hardy and Littlewood's circle method. On the other hand, for Vinogradov's conjecture we are lead to study a higher dimensional function

$$j_{k,N}(x) = \sum_{n=1}^{N} e^{2\pi i (n x_1 + n^2 x_2 + \cdots + n^k x_k)},$$

in which case

$$J_{k,s}(N) = \|j_{k,N}\|_{L^{2s}[0,1]^k}.$$

Though in higher dimensions, Vinogradov initially expected his result to be *easier* than Hardy and Littlewood's problem, because of the *rescalable* nature of the problem, which we will get to later. Indeed, decoupling has lead to a solution of Vinogradov's conjecture for almost all parameters, but a complete solution to the conjectured bound

$$\mathrm{HL}_{k,s}(N) \lesssim_\varepsilon N^\varepsilon (N^s + N^{2s-k})$$

has not yet been resolved.

To do our analysis, it will be helpful to rescale our quantities. Thus we let

$$f_{k,N}(x) = \sum_{n=1}^{N} e^{2\pi i((n/N)x_1 + \cdots + (n/N)^k x_k)}.$$

The function $f_{k,N}$ is then N-periodic, and so Vinogradov's conjecture will follow from showing that

$$\|f_{k,N}\|_{L^{2s}[0,N^k]^k} = N^{k^2/2s}\|j_{k,N}\|_{L^{2s}[0,1]^k}$$

satisfies a bound

$$\|f_{k,N}\|_{L^{2s}[0,N^k]^k} \lesssim_\varepsilon N^{k^2/2s+\varepsilon}(N^s + N^{2s-k}).$$

This is what we will address using the theory of decoupling.

From the perspective of decoupling, it will be more handy to study the quantities

$$\tilde{j}_{k,N}(x) = \sum_{n=1}^{N} e^{2\pi i(nx_1 + \cdots + n^k x_k)}\phi(x),$$

and

$$\tilde{f}_{k,N}(x) = \sum_{n=1}^{N} e^{2\pi i((n/N)x_1 + \cdots + (n/N)^k x_k)}\phi(x_1/N, \ldots, x_k/N^k),$$

where $\phi \in \mathcal{S}(\mathbb{R}^d)$ has Fourier support in $|\xi| \leqslant 1/10$ and $\phi(x) \geqslant 1$ in a neighborhood of the origin. Bounding the L^{2s} norm of $\tilde{j}_{k,N}$ and $\tilde{f}_{k,N}$ is essentially equivalent to bounding $j_{k,N}$ since we can upper bound $j_{k,N}$ and $f_{k,N}$ pointwise by $O(1)$ sums of this form on the domain upon which we are taking the L^{2s} norm. Now if we let

$$f_i = e^{2\pi i((n/N)x_1 + \cdots + (n/N)^k x_k)}\phi(x_1/N, \ldots, x_k/N^k),$$

then $\hat{f}_i$ is supported on a rectangle with sidelengths $O(N) \times \cdots \times O(N^k)$ centered at the point $(n/N, \ldots, (n/N)^k)$ of the moment curve. The l^2 decoupling inequality for the moment curve thus implies that

$$\|f\|_{L^{2s}(\mathbb{R}^d)} \lesssim_\varepsilon N^{1/2+\varepsilon},$$

which proves Vinogradov's mean value result for $s \geqslant k^3/(2k^2 - 1)$, i.e. for $s \gtrsim k$, which, up to constants, is the current state of the art for the mean-value theorem.

61.3 Multilinear Kakeya

If $f_1, \ldots, f_N$ are functions, then Hölder's inequality implies that

$$\|f_1 \ldots f_N\|_{L^p(\mathbb{R}^d)} \leqslant \|f_1\|_{L^{Np}(\mathbb{R}^d)} \cdots \|f_N\|_{L^{Np}(\mathbb{R}^d)}.$$

This is tight if the $\{f_i\}$ are large in the same place. In the case of decoupling into caps on a parabola, we end up with functions that are locally constant on transverse tubes. In this case, we can significantly improve upon Hölder's inequality. In two dimensions, if we have two functions f_1 and f_2 constant in orthogonal directions, then by a rotation, we can assume that f_1 is a function of the x variable, and f_2 a function of the y variable. We then compute using Fubini's theorem that

$$\|f_1 f_2\|_{L^p(\mathbb{R}^2)} = \|f_1\|_{L^p(\mathbb{R})} \|f_2\|_{L^p(\mathbb{R})}.$$

This is a significant improvement upon an application of Hölder's inequality. The Loomis-Whitney inequality generalizes this to more general functions locally constant in orthogonal directions.

Theorem 61.5 (Loomis-Whitney)**.** *If $\pi_i(x) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$ are projections onto orthogonal hyperplanes, and $f_1, \ldots, f_n : \mathbb{R}^{n-1} \to [0, \infty]$ are functions, then*

$$\int_{\mathbb{R}^d} \prod_{j=1}^{d} f_j(\pi_j(x))^{\frac{1}{d-1}} \, dx \leqslant \prod_{j=1}^{d} \|f_j\|_{L^1(\mathbb{R}^{d-1})}^{\frac{1}{n-1}}.$$

Proof. Prove $d = 2$ case as above, and then apply induction TODO. $\qquad\square$

Changing coordinates gives an affine-invariant version of the ienquality.

Theorem 61.6. *If $v_1, \ldots, v_d$ are unit normal vectors to hyperplanes $H_1, \ldots, H_d$ through the origin with associated projection maps $\pi_j : \mathbb{R}^d \to H_j$, then for any functions $f_j : H_j \to [0, \infty]$,*

$$\int_{\mathbb{R}^d} \prod_{j=1}^{d} f_j(\pi_j(x))^{\frac{1}{d-1}} \, dx \leqslant |v_1 \wedge \cdots \wedge v_d|^{-\frac{1}{d-1}} \prod_{j=1}^{d} \|f_j\|_{L^1(H_j)}^{\frac{1}{d-1}}.$$

The multilinear Kakeya inequality follows by replacing the d unit vectors $\{v_j\}$ by d families of unit vectors $\mathcal{N}_1, \ldots, \mathcal{N}_d$ such that any choice of d from this familiy is α transverse, in the sense that

$$|v_1 \wedge \cdots \wedge v_j| \geqslant \alpha.$$

These unit vectors generate a family of radius r tubes $\mathcal{T}_1, \ldots, \mathcal{T}_j$. We then consider a ball B of radius $R \geqslant r$.

Theorem 61.7 (Multilinear Kakeya). *We have*

$$\int_B \prod_{j=1}^d \left(\sum_{T \in \mathcal{T}_j} a_T \mathbf{I}_T \right)^{\frac{1}{d-1}} \lesssim_\varepsilon \alpha^{-O(1)} (R/r)^\varepsilon r^d \prod_{j=1}^d \left(\sum_{T \in \mathcal{T}_j} a_T \right)^{\frac{1}{d-1}}$$

Proof. SEE ZORIN KRANICH NOTES: (Guth 10) shows $(R/r)^\varepsilon$ can be removed, and the power of α can be taken to be $1/(d-1)$. $\square$