Harmonic Analysis

Spring 2016

Ioannis Parissis

IKERBASQUE
research fellow

University of the
Basque Country
(UPV/EHU)

July 12, 2016

Harmonic Analysis

Ioannis Parissis

I.P.: Departamento de Matemáticas, Universidad del Pais Vasco, Aptdo. 644, 48080 Bilbao, Spain and Ikerbasque, Basque Foundation for Science, Bilbao, Spain

E-mail address: ioannis.parissis@ehu.es

E-mail address: ioannis.parissis@gmail.com

URL: https://sites.google.com/site/ioannisparissis/

Ioannis Parissis is supported by grant MTM2014-53850 of the Ministerio de Economía y Competitividad (Spain), grant IT-641-13 of the Basque Gouvernment, and IKERBASQUE.

Contents

Course Overview vii
- Prerequisites vii
- Syllabus viii

Preface xi
- Hardy notation; a constant $c > 0$ that has an unspecified value. xii
- The Vinogradov notation. xiii
- The Landau - big O -notation. xiii

Chapter 1. A brief overview of measure and integral 1
- 1.1. Basic notions from measure theory 1
- 1.2. L^p-spaces 2
- 1.3. The dual space of L^p 11
- 1.4. Weak L^p-spaces 12

Chapter 2. Convolution, Dense subspaces and interpolation of operators 15
- 2.1. Convolutions and approximations to the identity 15
- 2.2. Some dense classes of functions 20
- 2.3. Operators on L^p spaces; boundedness and interpolation 23

Chapter 3. The Fourier transform and the space of tempered distributions 33
- 3.1. The Fourier transform on $L^1(\mathbb{R}^n)$. 33
- 3.2. The Fourier transform of the Schwartz class 47
- 3.3. The Fourier transform on $L^2(\mathbb{R}^n)$ 51
- 3.4. The Fourier transform on $L^p(\mathbb{R}^n)$ and Hausdorff-Young 52
- 3.5. The space of tempered distributions $\mathcal{S}'(\mathbb{R}^n)$ 54
- 3.6. The Uncertainty principle 64
- 3.7. Translation invariant operators 65

Chapter 4. The Hardy-Littlewood maximal function 69
- 4.1. Averages and maximal operators 69
- 4.2. The Hardy-Littlewood maximal theorem 73
- 4.3. Consequences of the maximal theorem 76
- 4.4. Variants of the Hardy-Littlewood maximal function 80
- 4.5. The Calderón-Zygmund decomposition 88

Chapter 5. The Hilbert transform 95
- 5.1. The definition of the Hilbert transform on $\mathcal{S}(\mathbb{R}^n)$. 96
- 5.2. The Hilbert transform on $L^2(\mathbb{R})$ 99
- 5.3. The Hilbert transform on $L^p(\mathbb{R})$ 102

Chapter 6. Calderón-Zygmund operators 111
6.1. Singular kernels and Calderón-Zygmund operators 111
6.2. Pointwise convergence and maximal truncations 115
6.3. Vector valued Calderón-Zygmund singular integral operators 120

Chapter 7. The space of functions of bounded mean oscillation, BMO. 123
7.1. Singular integral operators on L^∞ and BMO. 123
7.2. The John-Nirenberg Inequality 128
7.3. Interpolation and BMO. 131
7.4. Commutators of singular integrals 134

Chapter 8. Littlewood-Paley inequalities and multiplier operators 137
8.1. The Littlewood-Paley decomposition 137
8.2. Littlewood-Paley Projections and derivatives 139
8.3. The Littlewood-Paley inequalities 141
8.4. Two theorems on multipliers 146
8.5. The Sobolev spaces $W^{s,p}(\mathbb{R}^n)$ 154

Course Overview

During the spring semester of 2013 there will be an advanced course in Harmonic Analysis at Aalto University, taught by Ioannis Parissis. The target group of this course is advanced undergraduate, Master's and PhD students. The main goal of the course is to discuss a wide class of operators acting on L^p spaces for which an axiomatic theory can be established. These operators are usually called Calderón-Zygmund operators and the relevant notions, tools, and results can be summed up with the term "Calderón-Zygmund theory". The main prerequisite for this course is a solid knowledge of Real Analysis, while some knowledge of basic Functional Analysis will be helpful, but not critical. Several tools will be introduced in order to discuss mapping properties of operators acting on L^p spaces. These include, interpolation theorems, the Fourier transform, the discussion of tempered distributions, weak derivatives, and the action of the Fourier transform on these objects, the Hardy-Littlewood maximal function and its variants, dyadic decompositions and model operators. These tools will be combined to provide us with general theorems that will assure the boundedness of a wide class of operators on L^p-spaces. Together with these tools we will also discuss in some detail the space of Schwartz functions (smooth functions that together with their derivatives of every order decay faster than any polynomial power at infinity). This space consists of so nice functions that it allows its dual to contain pretty rough objects. This dual is the space of tempered distributions which we will also discuss in some detail. Finally we will introduce the space BMO consisting of functions of bounded mean oscillation. This function space is larger than the space of bounded functions and usually serves as a replacement endpoint for the boundedness of singular integral operators, which fail to be bound on L^∞. There will be two two-hour lectures every week and an extra two-hour exercise session every other week.

Prerequisites

This course will assume a prior knowledge of real analysis and in particular, the notions of Lebesgue measure and integration, measurable sets and functions, convergence and approximation theorems (e.g. Lebesgue dominated convergence, approximation of integrable functions by simple functions), L^p-spaces, basic inequalities such as Hölder and Minkowski, and a mild familiarity with using these notions. Furthermore, some knowledge of basic Functional Analysis such as classical Banach spaces and their duals, Hahn-Banach theorem and its consequences, and so on, will be useful but not crucial. Some of these notions will be reviewed (but not rigorously defined nor discussed) in the beginning of the course.

Syllabus

Although there might be small changes, the main plan for the course is the following:

Introduction: In the first part of the course we will recall some basic notions from real analysis and add some new elements.

We will start by setting up the main environment for our studies, that is, the appropriate function spaces where our functions will live and our operators will act. There will always be an underlying measure space $(X, \mathcal{B}, \mu)$. As a typical example you should think of X as the Euclidean space $\mathbb{R}^n$, $\mathcal{B}$ as the σ-algebra of Borel, or Lebesgue measurable sets, and μ as the Lebesgue measure on $\mathbb{R}^n$. We will however put things in a more general context whenever it is useful or necessary. We will usually consider appropriate spaces of functions $f : X \to \mathbb{C}$. The most typical example here would be the space of functions whose p-th powers are integrable with respect to the measure μ, that is the spaces $L^p(d\mu)$ and p will usually lie in the interval $[1, \infty]$. Another relevant space of importance is the space of functions that marginally fail to be in L^p, that is the weak-L^p spaces. These, as we will see, are defined in terms of the measure of the distribution function of the function f. We will also extensively use the spaces of infinitely differentiable functions with compact support, the space of Schwartz functions, that is the space of infinitely differentiable functions whose partial derivatives of every order (including the 0-order derivative, that is the function itself) decay faster than any polynomial power at infinity, the space of continuous functions that tend to zero at infinity and so on. I will assume that most of the audience is familiar with these notions on some level or another. However, this will be our starting point; we will recall these notions from measure theory (or real analysis if you want) and take them one step further. A recurring theme in this course will be the study of operators acting on these function spaces and, in particular, their boundedness and mapping properties. For this we will oftentimes use classical inequalities in measure spaces as for example Hölder's inequality, Minkowski's inequality and Young's inequality, as well as slightly more sophisticated tools, that is, different forms of interpolation of operators (e.g. Marcinkiewicz interpolation theorem, Riesz-Thorin interpolation theorem), Schur's test, convolution inequalities and duality arguments. We will review the classical inequalities and introduce the more sophisticated tools just mentioned.

The Fourier transform: We will introduce the Fourier transform of appropriate functions $f : \mathbb{R}^n \to \mathbb{C}$ and study its main properties on the corresponding spaces. Special mention will be made on the Fourier transform on the space of finite measures on $\mathbb{R}^n$, on $L^1(\mathbb{R}^n)$, on $L^2(\mathbb{R}^n)$ as well as on the Schwartz space $\mathcal{S}(\mathbb{R}^n)$. Although the latter function space seems pretty limited, its dual, the space of tempered distributions, is rich enough to allow us to extend the definition of the Fourier transform (in a weak sense) to a wide variety of objects, including L^p spaces for $p > 2$. We will rely on the space of tempered distributions in order to define operators (as for example the Fourier transform, or the derivative) on functions that do not possess the necessary regularity. We will give examples of classical Fourier transforms, like the Fourier transform of the Gaussian, and discuss how one can reconstruct the original function from its Fourier transform, that is we will see when, how, and in what sense we can 'invert' the Fourier transform. Some time will be given to the discussion of bounded linear operators that *commute with translations*. We will see that these operators are convolution operators with an appropriate distribution.

The Hardy-Littlewood Maximal function. We will introduce (or recall) the Maximal function of Hardy and Littlewood and prove its main boundedness properties. This will be done in different ways; we will use the classical approach that is prove the L^1 to weak L^1 inequality by means of a covering lemma and then interpolate between this bound and the trivial $L^\infty \to L^\infty$ bound. We will also study the relevance of the maximal function to the standard Calderón-Zygmund decomposition. In parallel, we will study the dyadic maximal function and see how it relates to the usual one.

Singular Integrals: We will introduce singular integral operators initially acting on "nice" (say Schwartz) functions on $\mathbb{R}^n$. The purpose here is to show apriori boundedness of these operators, which will automatically allow us to extend them to the spaces L^p for example. Our starting point will be the Hilbert transform which is the primordial example of a Calderón-Zygmund operator and whose properties we will examine in detail. We will build in this section the basic hypotheses of Calderón-Zygmund theory and show how we can deduce the boundedness properties of general Calderón-Zygmund operators from some size and regularity assumptions on the kernel of the operator together with an initial boundedness hypothesis (for example that the operator is already bounded on some L^p space. Given time, we will discuss how this apriori boundedness assumption can be replaced by suitable testing conditions by means of the famous $T(1)$ theorem. We will also hint at some recent developments in Harmonic Analysis involving the representation of general Calderón-Zygmund operators by appropriate averages of dyadic model operators.

Littlewood-Paley theory and multiplier operators. This concluding section of the course aims mainly at introducing the Littlewood-Paley decomposition of a function and prove the Littlewood-Paley inequalities. Roughly speaking, these inequalities allow us to decompose a function to different pieces which have localized frequencies in dyadic annuli, and behave almost orthogonally to each other. In the Hilbert space L^2 this is precise. The Littlewood-Paley inequalities provide us with a certain substitute of in L^p, $p \neq 2$. Given time we will discuss multiplier operators and give two fundamental theorems: the Mikhlin-Hörmander multiplier theorem and the Marcinkiewicz multiplier theorem.

The preceding description gives the main topics I would like to cover in the course. On the other hand I plan to touch upon some special subjects as for example, oscillatory integral estimates, Sobolev inequalities and relation to PDE's, weighted norm inequalities, Fourier transform on different groups, Fourier series and so on. There will be relevant exercises in your homework giving you a flavor of these subjects (with appropriate guidance of course!) as well as examples in the classroom. There will also be a home assignment on a special self-contained subject to be presented in the classroom. The choice has to be compatible with the material of the course. To give a flavor, a (not exhaustive) list of possible subjects is:

- Weighted Inequalities for Maximal functions and Singular Integrals.
- Dyadic representation of Singular Integral Operators.
- Sharp weighted inequalities for Maximal functions and Singular Integrals.
- Sobolev Embedding Theorems.
- An instance of the Stein-Tomas restriction Theorem.
- Interpolation theorems not covered in the course.
- $T(1)$ and $T(b)$ theorems.
- Oscillatory integral estimates and the method of stationary phase.
- Singular Integral operators outside the scope of Calderón-Zygmund theory.
- Three term Arithmetic Progressions via Fourier transform.

Schedule. There will be two two-hour lectures every week, normally on Monday and Tuesday. There will be a third two-hour meeting (on Thursday) which will serve as an exercise session.

Grading, Homework and Exams. There will be a set of exercises given to you as homework (approximately every two-three weeks). You will have to hand in your solutions usually within 2-3 weeks. We will discuss these problems (and others) in the Thursday problem session.

Communication. I expect you to check your e-mails on a regular basis for course related issues. Check also my web site where all my contact information is available.

Literature. I will suggest some books that I think will be of great help throughout the course. This list however is neither restrictive nor exhaustive. I would encourage you to use any book or online resource that you feel can help you.

[F] G. Folland, "Real Analysis: Modern Techniques and Applications", Wiley, 1984.

[D] J. Duoandikoetxea, "Fourier Analysis", AMS, 2001.

[K] Y. Katznelson, "An Introduction to Harmonic Analysis" 2nd edition, Cambridge, 2004.

[R] W. Rudin, "Real and Complex Analysis", 3rd ed., McGraw-Hill, 1987.

[SW] E. Stein, G. Weiss, "Introduction to Fourier Analysis on Euclidean Spaces", Princeton Univ. Press, 1971.

[S] E. Stein, "Singular Integrals and Differentiability Properties of Functions", Princeton Univ. Press, 1970.

[S2] E. Stein, "Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals", Princeton Univ. Press, 1993.

[WZ] R. L. Wheeden.; A. Zygmund, "Measure and integral: An introduction to real analysis. Pure and Applied Mathematics", Marcel Dekker, 1977.

Preface

As mentioned in the overview of the course, we will be mainly concerned with operators acting on certain function spaces, or even spaces of more rough "objects" such as measures or distributions. Typically we will want to study the mapping properties of such an operator, that is whether it maps one function space to another and so on. A typical estimate in this context is of the form

$$\|Tf\|_Y \leq C\|f\|_X,$$

where X, Y are certain (usually Banach) spaces of functions or measures or distributions and $\|\cdot\|_X, \|\cdot\|_Y$ are norms, or semi-norms or, in general, norm-like quantities. Thus such an estimate states that the operator T takes functions (or "objects") from the space X to the space Y in a continuous way. Already such an estimate can reveal quite a lot for the nature and the properties of the operator T.

When studying the mapping properties of an operator it is often useful to restrict our attention to a "nice" subclass V inside X. If X for example consists of integrable functions, a good idea is to first consider the action of T on the class of smooth functions with compact support, or on the class of simple functions. These subclasses are nice or explicit enough to allow us to overcome many technical difficulties in trying to define $T(f)$ for a general object $f \in X$. Furthermore, when these classes are dense in the original space there is a very natural candidate for the extension of T to the whole class. It turns out that this extension goes through whenever T is bounded on the dense subclass. Another useful technique is to decompose a general function $f \in X$ into different pieces. Since T is usually linear, we can then examine the effect of T on each piece and sum the pieces together. Likewise, we can decompose the operator T to different components, each component being easier to control than the 'whole' operator T. Finally, we can combine these two ideas and decompose both the function and the operator into different pieces. Usually good control on the different pieces is expected to imply a good control on the original operator and/or function. There are however technical difficulties in putting the pieces together, understanding how they interact with each other and, most importantly, justifying how the individual estimates sum up to a "global" estimate.

Overall this course is all about estimates: Estimating the norm of a function, the norm of an operator, the norms of the different pieces of a decomposition of a function and so on. It is very useful to introduce some notation:

Hardy notation; a constant $c > 0$ that has an unspecified value.

Such a constant $c, c_1, c_2, \ldots$, or C, A, B and so on, usually represents a numerical constant that does not depend on any of the parameters of the problem. Using this notation we will many times use a letter, say c, to denote a generic numerical constant. Different appearances of the letter c will not necessarily denote the same numerical constant, even in the same line of text. For example a very useful estimate is the following

$$\frac{2}{\pi}|x| \leq |\sin(x)| \leq |x|, \quad |x| \leq \frac{\pi}{2}.$$

We will use the Hardy notation in order to write estimates likes this in the form

$$c_1|x| \leq |\sin x| \leq c_2|x|, \quad |x| \leq \frac{\pi}{2},$$

which is just the statement the fact that the function $\sin x$ behaves linearly close to 0. The precise values of the constants, that is, the precise slopes of the linear functions appearing in the estimate, are rarely of any importance and the do not depend on anything interesting. Taking this one step further we would write for example

$$|2\sin(x)/(1+x)| \leq c|x/(1+x)| \leq c|x|$$

when x is close to 0 and

$$|2\sin(x)/(1+x)| \leq c/|1+x|,$$

when $|x| \to \infty$.

A variation of this notation is useful when a constant actually depends on one of the parameters of the problem. Thus we could write

$$\|Tf\|_Y \leq c_{X,Y,T}\|f\|_X,$$

which means that the constants $c_{X,Y,T}$ may depend on X, Y and T but *not* on the function f. One should be careful with estimates like this. For example, the notation

$$2^n \leq c_n$$

is correct though it might be confusing as it "hides" the dependence of the constant c_n on n (for example whether it is bounded in n, whether it grows to infinity in n and so on). On the other hand, the notation

$$2^n \leq c$$

is wrong though the estimate is actually true for *fixed* n. Such a notation would imply that the sequence 2^n is uniformly bounded in n which is of course not true. Such a notation is true for example in the case

$$|\sin(2\pi n)| \leq c.$$

The Vinogradov notation.

Suppose that we have an estimate of the form $Y \leq cX$ where X, Y could be norms of functions, or operators and so on. We will write this estimate in the form

$$Y \lesssim X.$$

Similarly we write $Y \gtrsim X$ whenever $Y \geq cX$. If we have that $Y \lesssim X$ and $Y \gtrsim X$ then we will use the notation $X \simeq Y$. This latter notation states that the quantities X,Y are equivalent up to numerical constants. For example, we could write $2\sin(2\pi n) \lesssim 1$ and also $\sin x \simeq x$ for x close to 0. If we want to state a dependence on a parameter we use a subscript. For example we write

$$\|Tf\|_Y \lesssim_{X,Y,T} \|f\|_X,$$

to denote the dependence of the implied constant on X, Y and T.

A lot of attention should be given when iterating this notation. While this is legitimate for a finite number of steps, an infinite number of steps can create many problems. Beware of this situation especially in inductive arguments: never hide the dependence on the induction parameter in the Vinogradov notation!

The Landau - big O -notation.

In this notation, writing $Y = O(X)$ means that there exists a numerical constant $C > 0$ such that $|Y| \leq CX$. The big O notation however is mostly useful when we want to denote a main term and an error term, and keep track of everything in a nice way. Imagine for example that we want to study the function $\sin x$ for x close to zero, say $|x| < \frac{\pi}{2}$. The Taylor expansion of $\sin x$ around zero is of the form

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots .$$

While it is correct that $\sin x = O(|x|)$ as $x \to 0$, what happens if we want to keep track of lower order terms? Well, we could use the big-O notation to write

$$\sin x = x + O(|x|^3), \quad x \to 0.$$

Note that this is correct since the higher order terms x^5, x^7 and so on, are always controlled by x^3 when $x \to 0$. This is a very useful device if we want to "carry" the lower order terms in our calculations. For example, since

$$\sin x = x + O(|x|^3), \quad \cos x = 1 - \frac{x^2}{4} + O(x^6), \quad x \to 0,$$

we can write

$$\sin x \cos x = (x + O(|x|^3))(1 - \frac{x^2}{4} + O(|x|^4)) = x + O(|x|^3).$$

If we want to state the dependence on some parameter we use subscripts again. Thus we could write $Y = O_n(X)$ meaning that $|Y| \leq c_n X$. Also note that the bound $\|Tf\|_Y \lesssim_{X,Y,T} \|f\|_X$ can be written in the form $\|Tf\|_Y = O_{X,Y,T}(\|f\|_X)$.

CHAPTER 1

A brief overview of measure and integral

1.1. Basic notions from measure theory

We begin this introductory chapter by recalling some basic facts from the theory of measure and integration. As mentioned in the description of the course, our first task will be to recall all the basic notions and tools from integration theory and L^p spaces, thus defining our main setup. Our basic environment is a *measure space* $(X, \mathcal{X}, \mu)$, that is a set X together with a σ-algebra $\mathcal{X}$ of sets in X and a non-negative measure μ on X. The measure μ will always assumed to be σ-finite (X can be decomposed as a countable union of sets of finite μ-measure). Recall that our subject is *Euclidean* harmonic analysis so, in most cases, the underlying space X will be the n-dimensional Euclidean space, μ will be the Lebesgue measure on $\mathbb{R}^n$ and $\mathcal{B}$ will be either the σ-algebra of *Lebesgue*-measurable sets, or the σ-algebra of *Borel*-measurable sets in $\mathbb{R}^n$.

Typically we will consider measurable functions $f : (X, \mathcal{X}, \mu) \to (Z, \mathcal{Z}, \nu)$; remember here that measurable means that the pre-image of every measurable set (thus of every set in $\mathcal{Z}$) is a measurable set (that it is a set in $\mathcal{X}$). However, we will mostly consider functions $f : X \to \mathbb{C}$, where it is understood that $\mathbb{C}$ is equipped with the Borel σ-algebra[1]. Again, the special case of Lebesgue-measurable complex valued functions on $\mathbb{R}^n$ is of particular importance. Thus the main example to keep in mind is a *Lebesgue-measurable*, complex valued function

$$f : \mathbb{R}^n \to \mathbb{C},$$

where $\mathbb{R}^n$ is equipped with the Lebesgue σ-algebra and $\mathbb{C}$ is equipped with the Borel σ-algebra of sets in $\mathbb{C}$. Note these definitions and conventions here since we won't repeat them every time we consider measurable functions.

Let us go back to the case of a general measure space $(X, \mathcal{X}, \mu)$. If not otherwise stated, a *set* in X will mean a *measurable* set in $\mathcal{X}$ and a function f will mean a *measurable* complex valued function. For a set E in X, the indicator function of E will be denoted by $\mathbf{1}_E(x)$ or $\mathbf{1}_E(x)$:

$$\mathbf{1}_E(x) = \mathbf{1}_E(x) = \begin{cases} 1, & \text{if} \quad x \in E, \\ 0, & \text{if} \quad x \notin E. \end{cases}$$

[1] Recall that the Borel σ-algebra on X is the smallest σ-algebra containing all open sets (or, equivalently, all closed sets).

A *simple* function is then a *finite linear combination of indicator functions*, that is a function $g(x)$ defined as

$$g(x) = \sum_{j=1}^{N} c_j \mathbf{1}_{E_j},$$

where $c_1, \dots, c_N \in \mathbb{C}$ and $E_1, \dots, E_N$ are (measurable) sets. A standard way to identify a set in X with a measurable function on X is via the map $E \mapsto \mathbf{1}_E$.

Two functions (or sets) in a measure space will be considered one and the same object if they agree *almost everywhere*. For example, consider a set E in X and a subset $E' \subset E$ with $\mu(E \setminus E') = 0$. For the purposes of this course, the functions $\mathbf{1}_E$ and $\mathbf{1}_{E'}$ are one and the same function. If you want to be more rigorous, you have to think of a measurable function as an *equivalence class* of functions, where two measurable functions f, f' are equivalent if and only if $f = f'$, μ-*almost everywhere* (μ-a.e.). That is, $f = f'$ everywhere on X except maybe on a set of measure zero. We will however abuse language a bit and just refer to f as a function, arbitrarily choosing a representative from every equivalence class. Moreover, we can choose the member of the class that is more convenient for our purposes. To give an example of the usefulness of this principle, think of the equivalence class of functions f, say on $\mathbb{R}$, that agree with 0 almost everywhere. One can think of functions that behave very erratically and are equal to 0 almost everywhere. However, the function f that is identically equal to 0 *everywhere* still belongs to the same equivalence class and is continuous, thus it qualifies as a "nice" representative of this equivalence class. For continuous functions however, there is no ambiguity.

EXERCISE 1.1. Let X, Y be two topological spaces and suppose that Y is Hausdorff. Assume that μ is a Borel measure on X such that $\mu(U) > 0$ for every open set $U \subset X$. Prove that if $f, g : X \to Y$ are continuous and $f = g$ μ-a.e. on X, then $f = g$ on X.

Hint: Since the space Y is Hausdorff, "open sets separate points": for every $y_1, y_2 \in Y$ with $y_1 \neq y_2$ there exist disjoint open neighborhoods V_{y_1}, V_{y_2} of y_1, y_2, respectively.

1.2. L^p-spaces

Let us begin by fixing a measure space $(X, \mathcal{X}, \mu)$. We assume as usual that μ is a non-negative σ-finite measure on X. The most important spaces of functions in this course will be the spaces $L^p = L^p(X, \mu)$, $0 < p < \infty$, defined as the spaces of measurable functions $f : X \to \mathbb{C}$ such that

$$\|f\|_{L^p(X,\mu)} = \left(\int_X |f(x)|^p d\mu(x) \right)^{\frac{1}{p}} < +\infty.$$

For $p = \infty$ we define the space of essentially bounded functions $f : X \to \mathbb{C}$, that is the space of measurable functions f such that

$$\|f\|_{L^\infty(X,\mu)} = \operatorname*{ess\,sup}_{x \in X} |f(x)| < +\infty.$$

Recall here that the essential supremum of a function f is the smallest positive number c such that $|f(x)| \le c$, μ-almost everywhere:

$$\operatorname*{ess\,sup}_{x\in X} |f(x)| = \inf\Big\{c > 0 : \mu(\{x \in X : |f(x)| > c\}) = 0\Big\}.$$

We will alternatively use the notations $\|f\|_{L^p}$ or even $\|f\|_p$ whenever the underlying measure space is clear from context or not relevant for a statement.

EXERCISE 1.2. Let f be a simple function of finite measure support, that is, a finite linear combination of indicator functions of sets of finite measure. Show that

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty,$$

and that

$$\lim_{p\to 0} \|f\|_p^p = \mu(\operatorname{supp}(f)),$$

where

$$\operatorname{supp}(f) = \{x \in X : f(x) \ne 0\}.$$

As we shall shortly see, for $1 \le p \le \infty$, the quantities $\|\cdot\|_{L^p(X,\mu)}$ are Banach-space norms. In order to show this, the only difficulty is the triangle (or Minkowski, in this case) inequality. For $0 < p < 1$ these quantities are not norms any more but we have a quasi-triangle inequality, that is a triangle inequality with a constant strictly greater than 1.

LEMMA 1.3. *Let $(X, \mathcal{X}, \mu)$ be a measure space. For $1 \le p < \infty$, the quantity $\|\cdot\|_{L^p(X,\mu)}$ is a norm. In particular we have the following, for all functions $f, g \in L^p(X, \mu)$:*

(i) *(Point Separation)*

$$\|f\|_{L^p(X,\mu)} = 0 \Leftrightarrow f = 0.$$

(ii) *(Positive Homogeneity) For all $c \in \mathbb{C}$ we have*

$$\|cf\|_{L^p(X,\mu)} = |c|\|f\|_{L^p(X,\mu)}.$$

(iii) *(Triangle inequality)*

$$\|f + g\|_{L^p(X,\mu)} \le \|f\|_{L^p(X,\mu)} + \|g\|_{L^p(X,\mu)}.$$

For $0 < p < 1$, (i) and (ii) still hold true. Triangle inequality is replaced by

(iii′) *(Quasi-triangle inequality)*

$$\|f + g\|_{L^p(X,\mu)} \lesssim_p \|f\|_{L^p(X,\mu)} + \|g\|_{L^p(X,\mu)}.$$

PROOF. The statements (i) and (ii) are trivial, given the fact that we identify functions that agree μ-a.e. For (iii) we can assume that f, g are non-zero because of (i), otherwise there is nothing to prove. The case $p = \infty$ of (iii) is trivial so we assume that $1 \le p < \infty$. Because of the homogeneity property (ii) it is enough to prove that $\|f + g\|_p \le 1$ whenever $\|f\|_p + \|g\|_p = 1$. Since f, g are non-zero this means that there exists $\theta \in (0, 1)$ such that $\|f\|_p = \theta$ and $\|g\|_p = 1 - \theta$. Setting $F = f/\theta$ and $G = g/(1-\theta)$ the problem reduces to showing that

$$\int |\theta F(x) + (1-\theta)G(x)|^p d\mu(x) \le 1, \tag{1.1}$$

whenever

$$\|F\|_p = \|G\|_p = 1.$$

We will show (1.1) by using a basic convexity estimate. For $s \in (0, \infty)$ we consider the function given by the formula $h(s) = s^p$ where $1 \le p < \infty$. Then the function h is convex. This means in particular that for $s_1, s_2 > 0$ and $0 < \theta < 1$ we have $h(\theta s_1 + (1-\theta)s_2) \le \theta h(s_1) + (1-\theta)h(s_2)$. Using the complex triangle inequality and the convexity of h we can thus write

$$\begin{aligned}\int |\theta F(x) + (1-\theta)G(x)|^p d\mu(x) &\le \int (\theta|F(x)| + (1-\theta)|G(x)|)^p d\mu(x)\\ &\le \theta \int |F(x)|^p d\mu(x) + (1-\theta)\int |G(x)|^p d\mu(x)\\ &= \theta + (1-\theta) = 1,\end{aligned} \tag{1.2}$$

because of the normalization $\|F\|_p = \|G\|_p = 1$.

The quasi-triangle inequality (iii′) is any easy consequence of the basic estimate $(a+b)^p \le a^p + b^p$, for $a, b > 0$ and $0 < p \le 1$, and is left as an exercise. □

EXERCISE 1.4. Show that the triangle inequality is an equality if and only if $f = g = 0$ or $f = cg$ for some $c > 0$.

Hint: Check carefully when the inequalities in the previous proof become equalities. Use the fact that for $f \ge 0$ we have $\int f = 0 \Leftrightarrow f = 0$ a.e.

For $1 \le p \le \infty$, the spaces $L^p(\mathbb{R}^n)$ are *Banach* spaces, that is they are normed vector spaces which are complete with respect to the corresponding norm $\|\cdot\|_{L^p(\mathbb{R}^n)}$. For $0 < p < 1$ we don't have a triangle inequality. However, the quasi-triangle inequality allows us to show that the spaces $L^p(X, \mu)$ are (quasi-normed) complete vector spaces,or, *quasi-Banach spaces*.

PROPOSITION 1.5. *For $1 \le p \le \infty$ the space $L^p(X, \mu)$ is a Banach space. For $0 < p < 1$ the space $L^p(X, \mu)$ is a quasi-normed complete vector space (quasi-Banach space). Furthermore, for $0 < p < \infty$ the preceding spaces are* separable. *Separability fails however for $p = \infty$.*

EXERCISE 1.6. Show that $L^\infty(\mathbb{R})$ is not separable. The underlying measure here is the Lebesgue measure.

A very useful variation of Minkowski's inequality is one where we "replace" the sum by an integral (which, roughly speaking, is also a sum!). Minkowski's (integral) inequality is the statement that "the norm of a sum (integral) is always smaller or equal to the sum (integral) of the norms."

PROPOSITION 1.7 (Minkowski's integral inequality). *Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be two measure spaces where the measures μ, ν are σ-finite non-negative measures. Let f be a $\mathcal{X} \otimes \mathcal{Y}$-measurable function on the product space $X \times Y$.*

(i) *If $f \ge 0$ and $1 \le p < \infty$, then*

$$\left(\int_X \left|\int_Y f(x,y) d\nu(y)\right|^p d\mu(x)\right)^{\frac{1}{p}} \le \int_Y \left(\int_X |f(x,y)|^p d\mu(x)\right)^{\frac{1}{p}} d\nu(y).$$

(ii) *If $1 \le p \le \infty$, $f(\cdot, y) \in L^p(X, \mu)$ for ν-a.e. $y \in Y$, and the function $y \mapsto \|f(\cdot, y)\|_{L^p(X,\mu)}$ is in $L^1(Y, \nu)$ for μ-a.e. $x \in X$, then $f(x, \cdot) \in L^1(Y, \nu)$ for μ-a.e.*

x, the function $x \mapsto \int_Y f(x,y)d\nu(y)$ *is in* $L^p(X,\mu)$ *and*

$$\Big\| \int_Y f(\cdot,y)d\nu(y)\Big\|_{L^p(X,\mu)} \leq \int_Y \|f(\cdot,y)\|_{L^p(X,\mu)}d\nu(y).$$

Writing (ii) of Minkowski's integral inequality also highlights the similarity to the classical triangle inequality, where one just has to think of the integral as a "generalized sum." This is also a good trick to help you memorize the inequality. Observe that the triangle inequality is just a special case of the integral version of Minkowski's integral inequality where the measure ν is the counting measure. You can find the proof of this inequality in most textbooks of real analysis. See for example [F].

After the triangle inequality the next most important inequality in the spaces $L^p(X,\mu)$ is Hölder's inequality.

LEMMA 1.8. *Let* $f \in L^p(X,\mu)$ *and* $g \in L^q(X,\mu)$ *for some* $0 < p,q \leq \infty$. *Define the exponent* r *by means of the "Hölder relationship"*

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q}.$$

Then the function $fg \in L^r(X,\mu)$ *and we have the norm estimate*

$$\|fg\|_{L^r(X,\mu)} \leq \|f\|_{L^p(X,\mu)}\|g\|_{L^q(X,\mu)}.$$

EXERCISE 1.9. Prove Lemma 1.8 above.

Hint: Note that the case $p = q = r = \infty$ is trivial. Assuming that $p,q,r < \infty$ homogeneity allows us to normalize $\|f\|_{L^p(X,\mu)} = \|g\|_{L^q(X,\mu)} = 1$, the case $f = 0$ or $g = 0$ being trivial. Normalizing and setting $F(x) = |f(x)|^p, G(x) = |g(x)|^q$, it is enough to prove that $\int_X F^\theta G^{1-\theta} \leq 1$ whenever $\int G = \int F = 1$, for suitable $\theta \in (0,1)$. Complete the proof using the fact that the function $\theta \mapsto a^\theta \beta^{1-\theta}$ is convex, where a,β are positive real numbers. To show this you can use the convexity of the function $x \mapsto e^x$.

REMARK 1.10. Observe that Hölder's inequality is invariant under the transformation $f \mapsto c_1 f$ and $g \mapsto c_2 g$ for any constants $c_1, c_2 > 0$. Note also that this inequality refers to a general measure space (X,μ). Replacing the measure μ by the measure $\tilde{\mu} = \lambda\mu$ for some constant $\lambda > 0$ observe that $f \in L^p(\mu) \Leftrightarrow f \in L^p(\tilde{\mu})$. Using these invariances and applying Hölder's inequality with $f = g = \mathbf{1}_A$ with $\mu(A) = 1$, we get

$$\lambda^{\frac{1}{r}} \leq \lambda^{\frac{1}{p}+\frac{1}{q}},$$

for all $\lambda > 0$. We conclude that we must have the Hölder relation between the exponents r,p,q,

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q},$$

whenever Hölder's inequality holds true.

1.2.1. Log-convexity of the L^p-norms. We will now study a characteristic of the L^p-norms which is implicit in many parts of the discussion on L^p-spaces, and especially in interpolation theorems. This convexity property is already hidden in the proof of Hölder's inequality above.

Let us start with a function $F:\mathbb{R}\to\mathbb{R}$. The function F is called *convex* if for every $x,y\in\mathbb{R}$ and any $0\le\theta\le 1$ we have that

$$F((1-\theta)x+\theta y)\le(1-\theta)F(x)+\theta F(y).$$

The same definition makes perfect sense whenever the function F is defined on some interval of the real line or, in fact, on any convex subset of a vector space. Observe that the definition states that the line connecting the points $(x,F(x))$ and $(y,F(y))$ of the graph of F always lies "above" the graph of the function itself. Now if a function F is positive, we will say that F is *log-convex* if the function $x\to\log F(x)$ is convex. In this case we must have

$$F((1-\theta)x+\theta y)\le F(x)^{1-\theta}F(y)^{\theta}.$$

PROPOSITION 1.11 (Log-convexity of the L^p-norms). *Let $0<p_1<p_2\le\infty$ and define p_2, $p_1\le p_2\le p_3$, as*

$$\frac{1}{p_2}=\frac{1-\theta}{p_1}+\frac{\theta}{p_3},$$

where $0<\theta<1$. Thus $1/p_2$ is a convex combination *of $1/p_1$ and $1/p_3$. Then*

$$\|f\|_{L^{p_2}(X,\mu)}\le\|f\|^{1-\theta}_{L^{p_1}(X,\mu)}\|f\|^{\theta}_{L^{p_3}(X,\mu)}.$$

Note that this means that the function $\frac{1}{p}\mapsto\|f\|_{L^p(X,\mu)}$ is log-convex.

PROOF OF PROPOSITION 1.11 VIA HÖLDER. Observing that $\frac{(1-\theta)p_2}{p_1}+\frac{\theta p_2}{p_3}=1$, we apply Hölder's inequality to $|f|^{p_2}=|f|^{(1-\theta)p_2+\theta p_2}$ to get

$$\int|f|^{p_2}=\int|f|^{(1-\theta)p_2}|f|^{\theta p_2}\le\Big(\int|f|^{p_1}\Big)^{\frac{(1-\theta)p_2}{p_1}}\Big(\int|f|^{p_3}\Big)^{\frac{\theta p_2}{p_3}},$$

which proves the desired estimate. □

The proof of the log-convexity of the L^p-norms via Hölder's inequality is quite elegant but not very illuminating. We will give another proof that employs a notion of convexity in complex analysis and, in particular, the maximum principle. We state the following lemma which will also be useful in the rest of the course.

LEMMA 1.12 (Three lines lemma). *Suppose that F is a bounded continuous complex-valued function on the closed strip $S=\{x+iy=z\in\mathbb{C}:0\le x\le 1\}$, that is analytic in the interior of S. Suppose that F obeys the bounds $|F(iy)|\le A$ and $|F(1+iy)|\le B$ for all $y\in\mathbb{R}$. Then we have that $|F(x+iy)|\le A^{1-x}B^x$ for all $z=x+iy\in S$.*

PROOF. First of all we can assume that $A,B>0$ otherwise there is nothing to prove. Now, consider the function $G(z)=F(z)/A^{1-z}B^z$ for $z\in\bar{S}$. Thus it suffices to show that $|G(z)|\le 1$ for all $z\in S$, whenever $|G(iy)|\le 1$ and $|G(1+iy)|\le 1$. First consider the case that $\lim_{|y|\to+\infty}|G(x+iy)|=0$ uniformly in $0\le x\le 1$. Then the result follows from the maximum principle. Indeed, there is some $y_o>0$ such that $|G(x+iy)|\le 1$ for all $|y|\ge y_o$. Now G is bounded by 1 on the boundary of the rectangle $[0,1]\times[-iy_o,iy_o]$ and the maximum principle implies that G is also bounded by 1 in the interior of the rectangle as well. Thus, in this case, G is bounded by 1 throughout the strip S.

To get rid of the condition $\lim_{|y|\to+\infty}|G(x+iy)|=0$ consider the sequence of functions $G_n(z)=G(z)e^{(z^2-1)/n}$, for $n\in\mathbb{N}$. Since G is bounded, say $|G(z)|\leq M$, we have that

$$|G_n(z)|=|G(z)|e^{-y^2/n}e^{(x^2-1)/n}\leq Me^{-y^2/n}\to 0,$$

as $|y|\to+\infty$, uniformly in $0\leq x\leq 1$. Observe that we still have the bounds $|G_n(iy)|\leq 1$ and $|G_n(x+iy)|\leq 1$ for $y\in\mathbb{R}$, uniformly in $n\in\mathbb{N}$ and every G_n is analytic in S and continuous in the interior of S. Thus we also conclude that $|G_n(z)|\leq 1$ for all n. Letting $n\to+\infty$ we get that $|G(z)|\leq 1$. □

REMARK 1.13. Observe that if we define the function $\phi:[0,1]\to\mathbb{C}$ as $\phi(x)=\sup\{|F(x+iy)|:y\in\mathbb{R}\}$, then under the hypothesis of the three lines lemma, we get that ϕ is log-convex in $(0,1)$. Another point to observe here is that the hypothesis we have stated here is not quite optimal. Indeed, we can actually relax the condition that F is bounded with the growth condition $|F(z)|\lesssim_F e^{O_F(e^{(\pi-\delta)|z|})}$ for some $\delta>0$ when $z\in S$. The idea of the proof is exactly the same. One first proves the result in the case that $\lim_{|y|\to+\infty}F(x+iy)=0$ uniformly in $x\in[0,1]$. Then we apply this for the sequence of functions $F_n(z)=e^{\frac{1}{n}e^{i[(\pi-\frac{1}{n})z+\frac{1}{2n}]}}F(z)$.

PROOF OF PROPOSITION 1.11 VIA COMPLEX ANALYSIS. We begin by making some reductions. Observe that the inequality we want to prove is invariant under the transformations $f\mapsto cf$ and $\mu\mapsto\lambda\mu$ for any constants $c,\lambda>0$. Using these invariances it is enough to show that if $\|f\|_{L^{p_1}}=\|f\|_{L^{p_3}}=1$ then we have that $\int|f|^{p_2}\leq 1$, for all p_2 with $0<p_1<p_2<p_3<\infty$. To do this, consider the entire function

$$\mathbb{C}\ni z\mapsto F(z)=\int_X|f|^{(1-z)p_1+zp_3}d\mu$$

Assuming that f is a simple function it is easy to see that F is bounded throughout the strip $S=\{x+iy:0\leq x\leq 1,y\in\mathbb{R}\}$. Observe also that we have the bounds $|F(0+iy)|\leq\|f\|_{p_1}$ and $|F(1+iy)|\leq\|f\|_{p_3}$. Using the three lines lemma we conclude that

$$|F(x+iy)|\leq 1,$$

for all $y\in\mathbb{R}$ and $0\leq x\leq 1$. Applying this bound for $y=0$ gives the log-convexity of the L^p-norms for simple functions. A limiting argument gives the log convexity for general functions. □

REMARK 1.14. In fact, one can follow the opposite direction and prove Hölder's inequality by means of the log-convexity of the L^p norms. Also, as in the case of Hölder's inequality, it is not hard to verify that whenever such an estimate is true, the indices p_1,p_2,p_3 must be related as

$$\frac{1}{p_2}=\frac{1-\theta}{p_3}+\frac{\theta}{p_1}.$$

To see this apply the inequality replacing the measure μ by $\lambda\mu$, where $\lambda>0$.

EXERCISE 1.15. Use the three lines lemma to give a different proof of Hölder's inequality.

Hint: Show Hölder's inequality initially for simple functions with finite measure support. For this, apply the three lines lemma to the function

$$F(z) = \int_X |f|^{p(1-z)}|g|^{qz} d\mu,$$

for $z \in S = \{x + iy = z \in \mathbb{C} : 0 \leq x \leq 1, y \in \mathbb{R}\}$. You can take for granted that simple functions with finite measure support are dense in $L^p(X, \mu)$, $1 \leq p < \infty$. Fill in the details of the limiting argument (omitted in the previous proof).

1.2.2. Heuristic discussion and examples of L^p-spaces.

Let us now see a couple of specific examples of L^p spaces which will come up often in this course.

EXAMPLE 1.16. The most common setting for this course will be the Euclidean setting, that is the measure space $(\mathbb{R}^n, \mathcal{L}, dx)$, where $\mathcal{L}$ denotes the σ-algebra of Lebesgue measurable sets in $\mathbb{R}^n$ and which typically will be omitted from the notation. A typical point in $\mathbb{R}^n$ will be denoted by $x = (x_1, \ldots, x_n)$ and $dx = dx_1 \cdots dx_n$ denotes the n-dimensional Lebesgue measure. For a set E in $\mathbb{R}^n$ we will many times write $|E|$ for its Lebesgue measure.

EXAMPLE 1.17. Consider the measure space $(\mathbb{Z}, \mathcal{D}, \nu)$, where $\mathcal{D}$ is the σ-algebra of all subsets of $\mathbb{Z}$. Here ν is the *counting measure*. Recall that for $E \subset \mathbb{Z}$, the counting measure of E is the cardinality of E, typically denoted by $|E|$, if E is finite, and $\nu(E)$ is defined to be $+\infty$ if E is infinite. Every subset of $\mathbb{Z}$ is clearly measurable with respect to ν. With these definitions taken as understood observe that the space $L^p(\mathbb{Z}, \mathcal{D}, \nu)$ is just the space of sequences on $\mathbb{Z}$ whose p-th powers are summable, that is, the space of all sequences $a = \{a_k\}_{k\in\mathbb{Z}}$ such that

$$\|a\|_p = \Big(\sum_{k\in\mathbb{Z}} |a_k|^p\Big)^{\frac{1}{p}} < +\infty.$$

These spaces come up so often in analysis that they deserve to have a special notation; we usually denote them by $\ell^p(\mathbb{Z})$. Maybe this seems like an unnecessary complication to state a very simple definition. Observe however that once we put things in this language we automatically have all the tools from measure theory at our disposal.

EXERCISE 1.18. Let $\{a^{(n)}\}_{n\in\mathbb{N}}$ be a sequence of elements in $(\mathbb{Z}, \mathcal{D}, \nu)$, that is, a sequence of sequences. For each positive integer $n \in \mathbb{N}$ we write $a^{(n)} = \{a_k^{(n)}\}_{k\in\mathbb{Z}}$. Assume that for each fixed $k \in \mathbb{Z}$, there is a complex number a_k such that $\lim_{n\to+\infty} a_k^{(n)} = a_k$, that is, the sequence $\{a^{(n)}\}_{n\in\mathbb{N}}$ converges *pointwise* to some sequence $a = \{a_k\}_{k\in\mathbb{Z}}$. State Lebesgue's dominated convergence theorem in this setup. When can we interchange the limit with summation?

EXAMPLE 1.19. We denote by $\mathbb{T}$ the *torus*, that is the quotient space $\mathbb{R}/2\pi\mathbb{Z}$ where $2\pi\mathbb{Z}$ is the group of integral multiples of 2π. Thus two points of $\mathbb{R}$ are identified if the differ by an integral multiple of 2π. There is a natural identification of functions on $\mathbb{T}$ and 2π-periodic functions on $\mathbb{R}$. The Lebesgue measure dt on $\mathbb{T}$ can also be identified with the restriction of the Lebesgue measure of $\mathbb{R}$ on the interval $[0, 2\pi)$, or in fact, any interval in $\mathbb{R}$ of length 2π. Remember that the Lebesgue measure on $\mathbb{R}$ is translation invariant. We equip $\mathbb{T}$ with the

Lebesgue σ-algebra. The integral of a function $f : \mathbb{T} \to \mathbb{C}$ can thus be written as

$$\int_{\mathbb{T}} f(t)dt = \int_0^{2\pi} f(x)dx,$$

where f is considered as a 2π-periodic function on $\mathbb{R}$. The preceding definitions imply that the measure dt on $\mathbb{T}$ is translation invariant. The Lebesgue spaces $L^p(\mathbb{T})$, $1 \leq p \leq \infty$, are defined in the obvious way. Since the total measure of $\mathbb{T}$ is finite, an important feature of the spaces $L^p(\mathbb{T})$ is that they are nested; for $1 \leq p_1 \leq p_2 \leq \infty$ we have that $L^{p_2}(\mathbb{T}) \subset L^{p_1}(\mathbb{T})$, $L^\infty(\mathbb{T})$ being the "smaller" space. Furthermore this embedding is continuous. See also Exercise 1.23.

We now briefly discuss why a function may fail to belong to $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$. For simplicity, let us focus on the real line and consider candidature to the spaces $L^p(\mathbb{R})$. Very similar conclusions hold in the n-dimensional Euclidean space. Roughly speaking, there are two main obstructions:

The decay of the function at infinity. Simply put, the function might not decay fast enough as $|x| \to +\infty$ for the integral of $|f(x)|^p$ to be finite. The most naive example one can think of is a constant function $f(x) = c$, $x \in \mathbb{R}$, for some complex number $c \in \mathbb{C}$. Obviously this function raised to any power cannot be integrable close to infinity. A slightly more subtle example is the function f which agrees with $1/x$ for $x \to +\infty$, i.e. $f(x) = \frac{1}{x}\mathbf{1}_{\{x\geq 1\}}(x)$ This function fails logarithmically to be in $L^1(\mathbb{R})$ but belongs to $L^p(\mathbb{R})$ for any $p > 1$. Of course we can similarly construct functions that decay even slower at infinity so that they fail to be in L^p for $p > 1$ as well. Thus, whenever a function f belongs to some L^p space for some $1 \leq p < +\infty$ this imposes a control on the decay of f at infinity. Increasing p will only make things better at infinity, provided that the function already has some decay. Observe that this obstruction does not exist on a finite measure space. This is the case for the spaces $L^p(\mathbb{T})$ for example.

Blow up at local singularities. Here it is enough to consider any compact set and study the behavior of the function locally. If the function is bounded on compact sets, i.e. if it is locally bounded, then the local behavior will not be an obstruction for the function to belong to some L^p space. Things become more interesting when there is a local singularity around a point. Here we can consider again the function $f(x) = \frac{1}{x}\mathbf{1}_{\{|x|\leq 1\}}(x)$, close to zero this time. This function has a logarithmic singularity at zero, and thus it does not belong to $L^1(\mathbb{R})$. Observe here that we have forced the function to be zero away from the origin in order to isolate the obstruction. As p increases to values $p > 1$, this function fails more and more dramatically to belong to $L^p(\mathbb{R})$ since we raise this singularity to higher powers, thus $|f|^p$ presents a more severe singularity at the origin. The 'solution' here would be to consider the L^p spaces for $p < 1$. Thus local singularities may also prevent a function from belonging to some L^p space. Unlike the behavior at infinity, the local behavior of $|f|^p$ improves as we decrease p. For example, the function $f(x) = \frac{1}{\sqrt{x}}\mathbf{1}_{\{|x|\leq 1\}}$ fails to be in $L^2(\mathbb{R})$ but clearly belongs to all $L^p(\mathbb{R})$ spaces for $p < 2$.

REMARK 1.20. A function f is in some L^p space if and only if the function $|f|$ belongs to the L^p space. Thus, there is no cancellation involved in the L^p-integrability of a function. This is an essential difference between the Lebesgue integral and the Riemann integral. The typical example here is to consider the function

$$f(x)=\sum_{n=0}^{\infty}\frac{(-1)^n}{n}\mathbf{1}_{[n,n+1)}(x).$$

Since $\int|f|$ is the harmonic series, f is not Lebesgue integrable. However, f is Riemann integrable since $\int f=\sum_{n=1}^{\infty}\frac{(-1)^n}{n}$ and the last series converges. Thus, whenever a function oscillates, we expect some cancellation in its integral that *will not* be reflected in the Lebesgue integrability of the function.

EXERCISE 1.21. Based on the previous discussion, answer the following questions (it is a simple calculation):

(i) Let $q\in(0,+\infty)$ be a given number. Based on the previous discussion, construct a function that belongs to $L^p(\mathbb{R})$ for all $p<q$ but does not belong to $L^q(\mathbb{R})$. For example, for $q=1$, a possible answer is the function $f(x)=\frac{1}{x}\mathbf{1}_{\{|x|\leq 100\}}$. Also, construct a function that belongs to $L^p(\mathbb{R})$ for all $p>q$ but does not belong to $L^q(\mathbb{R})$.

(ii) For $x\in\mathbb{R}^n$ and $\delta>0$, consider the function $f(x)=\frac{1}{|x|^\delta}\mathbf{1}_{\{|x|\leq 1\}}$. Characterize the values of $\delta>0$ as a function of n,p so that the function f belongs to $L^p(\mathbb{R}^n)$. Consider all the range $0<p<+\infty$ and calculate the L^p norm of the function, whenever it is finite.

(iii) For $x\in\mathbb{R}^n$ and $\delta>0$, consider the function $f(x)=\frac{1}{|x|^\delta}\mathbf{1}_{\{|x|>1\}}$. Characterize the values of $\delta>0$, as a function of n,p so that the function f belongs to $L^p(\mathbb{R}^n)$. Consider all the range $0<p<+\infty$ and calculate the L^p norm of the function, whenever it is finite.

REMARK 1.22. An important notion that is implicit in the previous discussion is that of *local integrability* of a function. A function $f:\mathbb{R}^n\to\mathbb{C}$ is called *locally integrable* if for every compact set $K\subset\mathbb{R}^n$ we have that

$$\int_K|f(x)|dx<+\infty.$$

We then write $f\in L^1_{\text{loc}}(\mathbb{R}^n)$. Local integrability ignores the behavior of a function at infinity. We are thus left with only one obstruction: the possibility that f has local singularities. Observe that if $f\in L^p(\mathbb{R}^n)$ for any $p\geq 1$ then f will be locally integrable. Similarly we can define the space $L^p_{\text{loc}}(\mathbb{R}^n)$.

EXERCISE 1.23. Give a heuristic explanation of the fact that if $f\in L^p(\mathbb{R}^n)$ for any $p\geq 1$ then $f\in L^1_{\text{loc}}(\mathbb{R}^n)$ (**hint:** what is the only obstruction for a function to be locally integrable?). Give a rigorous proof by means of Hölder's inequality. Show also (which is the same) that on a finite measure space (X,μ), we have that $L^q(X,\mu)$ is continuously embedded in $L^p(X,\mu)$ whenever $0<p\leq q\leq\infty$, that is, show that

$$\|f\|_{L^p(X,\mu)}\lesssim_{p,q,\mu(X)}\|f\|_{L^q(X,\mu)}.$$

Determine the best value of the implied constant in the previous inequality and give an example showing that once cannot have any better constant.

EXERCISE 1.24. For $0<p\leq\infty$ consider the spaces $\ell^p(\mathbb{N})$ of all complex sequences $a=\{a_n\}_{n\in\mathbb{N}}$ such that

$$\|a\|_p=\left(\sum_{n=1}^{\infty}|a_n|^p\right)^{\frac{1}{p}}<\infty.$$

Show that if $0 < p_1 \leq p_2 \leq \infty$ we have that $\ell^{p_1} \subset \ell^{p_2}$ and the embedding is continuous

$$\|a\|_{p_2} \leq \|a\|_{p_1}.$$

A space (X, μ) is called *granular* if there is constant $c_o > 0$ such that $\mu(E) > c_o$ for all measurable sets E of positive measure. Show that the for a granular space $L^p(X, \mu)$ with constant $c_o > 0$ we have that $L^{p_1} \subset L^{p_2}$ whenever $0 < p_1 \leq p_2 \leq \infty$ with

$$\|f\|_{L^p(X,\mu)} \lesssim_{p,q,c_o} \|f\|_{L^q(X,\mu)},$$

whenever $0 < q \leq p \leq \infty$ and (X, μ). What is the best value of the implied constant?

REMARK 1.25. Note that the opposite embedding is true for $L^p(X, \mu)$ spaces with $\mu(X) < \infty$. The explanation for this is quite simple. Sequences on $\mathbb{N}$ (or $\mathbb{Z}$) cannot have local singularities so the only deciding factor for candidature to some ℓ^p space is decay at infinity. This also explains the embedding in this exercise. If a sequence belongs to some ℓ^p space, this means there is already sufficient decay at infinity for the series $\sum |a_n|^p$ to be summable. Raising the exponent p only improves the decay of $|a_n|^p$ as $n \to \infty$. A similar phenomenon occurs in general in granular spaces.

EXERCISE 1.26. Show the following statements:

(i) Let $0 < p_0 < \infty$ and suppose that $f \in L^{p_0} \cap L^\infty$. Show that $\|f\|_p \to \|f\|_\infty$ as $p \to \infty$.
(ii) If $f \notin L^\infty$ show that $\|f\|_p \to \infty$ as $p \to \infty$.

1.3. The dual space of L^p

Remember that for a Banach space Y over $\mathbb{C}$, its dual X^* is the space of all bounded linear functionals $x^* : X \to \mathbb{C}$. Let $1 \leq p < \infty$ and define p' be the duality relation $1/p + 1/p' = 1$. For any $g \in L^{p'}(X, \mu)$ we define the functional

$$g^* : L^p(X, \mu) \to \mathbb{C},$$

by means of the formula

$$g^*(f) = \int_X f(x)\overline{g(x)} d\mu(x).$$

It is obvious that g^* is linear and Hölder's inequality shows that g^* is continuous since

$$|g^*(f)| \leq \|g\|_{L^{p'}(X,\mu)} \|f\|_{L^p(X,\mu)},$$

for all $f \in L^p(X, \mu)$. Thus $g^* \in (L^p(X, \mu))^*$. Actually, in most cases the opposite is true, that is, *every* functional in $(L^p(X, \mu))^*$ is uniquely defined by a function in $L^{p'}$, whenever $1 \leq p < +\infty$ and the measure μ is σ-finite.

THEOREM 1.27. *Let $1 < p < \infty$ and $x^* \in (L^p(X, \mu))^*$. There exists a unique $g \in L^{p'}(X, \mu)$ such that $x^* = g^*$. The same is true when $p = 1$ and the measure μ is σ-finite.*

REMARK 1.28. Theorem 1.27 fails (in most cases) when $p = \infty$. In fact the dual of L^∞ can be characterized as a space of measures but we will not pursue that here. We have however the following substitute.

PROPOSITION 1.29. *Let (X,μ) be a σ-finite measure space and Σ denote the simple functions on (X,μ), of finite measure support. Let $1\le p\le\infty$ and f be a function such that $fg\in L^1(X,\mu)$ for all $g\in\Sigma$. If the quantity*

$$M_p(f) := \sup\left\{\left|\int_X f(x)\overline{g(x)}d\mu(x)\right| : g\in\Sigma,\ \|g\|_{L^{p'}(X,\mu)}\le 1\right\} \tag{1.3}$$

is finite then $f\in L^p(X,\mu)$ and $\|f\|_{L^p(X,\mu)} = M_p(f)$.

Observe however that for this we need to know *a priori* that $fg\in L^1(X,\mu)$ for all $g\in\Sigma$. A way to bypass this problem is to work with a dense subclass of functions. This is essentially a duality relation but the small point just mentioned doesn't allow one to show that the dual of L^∞ is L^1 (luckily since it's not true!). It is however a very useful device since it allows very often to "linearize" L^p norms. Furthermore this duality relationship shows that the norm of the functional $g^*\in(L^p)^*$ is $\|g\|_{L^{p'}}$. Thus $(L^p)^*$ is *isometrically isomorphic* to $L^{p'}$, p' being the dual exponent of p, for $1<p<\infty$ and also for $p=1$ whenever the measure μ is σ-finite.

EXERCISE 1.30. Show the duality relation (1.3) in the previous remark. This is essentially a consequence of Hölder's inequality. Using this duality relation give an alternative proof of the triangle inequality.

REMARK 1.31. Density arguments allow us to restrict g in the supremum in (1.3) to belong to any dense subclass of $L^{p'}(X,\mu)$.

1.4. Weak L^p-spaces

Going back to the example of the function $h(x)=1/x$, $x\in\mathbb{R}$, recall that this function does not belong to $L^1(\mathbb{R})$. For $\lambda>0$ the following estimate is obvious

$$|\{x\in\mathbb{R} : |h(x)|>\lambda\}| \le \frac{2}{\lambda}.$$

On the other hand observe that for every function $f\in L^1(\mathbb{R})$ we have that

$$\|f\|_{L^1(\mathbb{R})} = \int_{\mathbb{R}} |f(x)|dx \ge \lambda|\{x\in\mathbb{R} : |f(x)|>\lambda\}|.$$

That is, for all L^1-functions f the measure of the set $\{x\in\mathbb{R} : |f(x)|>\lambda\}$ behaves like $\sim\frac{1}{\lambda}$.

In general, for any measure space (X,μ) we define for $0<p<\infty$ the space weak-$L^p(X,\mu)$ or $L^{p,\infty}(X,\mu)$ to be the space of all functions f such that

$$\mu(\{x\in X : |f(x)|>\lambda\}) \le \frac{c^p}{\lambda^p},\quad \lambda>0, \tag{1.4}$$

for some constant $c>0$. We define the weak-$L^p(X,\mu)$ or the $L^{p,\infty}(X,\mu)$ norm of a function f to be the smaller constant $c>0$ such that (1.4) is true. Equivalently

$$\|f\|_{L^{p,\infty}(X,\mu)} := \sup_{\lambda>0} \lambda\mu(x\in X : |f(x)|>\lambda\})^{\frac{1}{p}}.$$

For $p=+\infty$ we have $L^{\infty,\infty}=L^\infty$. Note that $\|\cdot\|_{L^{p,\infty}}$ is not a norm since the triangle inequality fails. It is however a quasi-norm (the triangle inequality holds with a constant).

EXERCISE 1.32. Show that for $0 < p < \infty$ and $f, g \in L^{p,\infty}$ we have the quasi-triangle inequality

$$\|f + g\|_{L^{p,\infty}(X,\mu)} \lesssim_p \|f\|_{L^{p,\infty}(X,\mu)} + \|g\|_{L^{p,\infty}(X,\mu)}.$$

PROPOSITION 1.33. *Let $0 < p < \infty$. The space $L^{p,\infty}(X, \mu)$ is continuously embedded in $L^p(X, \mu)$:*

$$\|f\|_{L^{p,\infty}(X,\mu)} \leq \|f\|_{L^p(X,\mu)}.$$

PROOF. We just use Chebyshev's inequality to write

$$\begin{aligned}\|f\|_{L^p}^p &= \int_X |f(x)|^p d\mu(x) \geq \int_{\{x\in X:|f(x)|>\lambda\}} |f(x)|^p d\mu(x)\\ &\geq \lambda^p \mu(\{x \in X : |f(x)| > \lambda\}),\end{aligned}$$

for every $\lambda > 0$. □

Let us also recall how we can write the L^p-norm of a function in terms of the distribution function of f:

PROPOSITION 1.34. *For $0 < p < \infty$ we have that*

$$\|f\|_{L^p(X,\mu)}^p = p \int_0^\infty \lambda^{p-1} \mu(\{x \in X : |f(x)| > \lambda\}) d\lambda.$$

EXERCISE 1.35. Prove Proposition 1.34 above.
Hint: It is elementary to see that

$$|f(x)|^p = p \int_0^\infty \mathbf{1}_{\{x\in X:|f(x)|\geq\lambda\}} \lambda^p \frac{d\lambda}{\lambda}.$$

Use Fubini's theorem to complete the proof.

EXERCISE 1.36. Prove the following assertions:

(i) Let $0 < p < \infty$ and $f \in L^p(X, \mu)$. We define $g(x) := f(x)\mathbf{1}_{\{x:|f(x)|\leq 1\}}$ and $b := f - g$. It makes sense to call g the "good" part of f and b the "bad" part, although both are actually good for different reasons! Show that $g \in L^q(X, \mu)$ for any $q > p$ and $b \in L^r(X, \mu)$ for any $r < p$.

(ii) For $0 < p < \infty$ and $f \in L^p(X, \mu)$ show that

$$\int_X |f(x)|^p d\mu(x) \simeq_p \sum_{k\in\mathbb{Z}} 2^{kp} \mu(x \in X : |f(x)| \geq 2^k\}).$$

Hint: For (ii) observe first that

$$\int_X |f(x)|^p d\mu(x) \simeq_p \sum_{k\in\mathbb{Z}} 2^{kp} \mu(\{x \in X : 2^k < |f(x)| \leq 2^{k+1}\})$$

Now one direction (the $\lesssim_p$) is straightforward. For the opposite direction it will help to split the function f to a sum of a "good" part g and a "bad" part b, $f = g + b$, and use (i).

CHAPTER 2

Convolution, Dense subspaces and interpolation of operators

In this chapter we begin by recalling the notion of convolution of functions. This is a basic but extremely useful tool in analysis that will allow us for example to easily construct smooth (or smoother) approximations to given functions. It also formalizes the averaging operation which will appear many times in this course, in various different forms. As an application we will show that several classes of "nice" functions are dense in the L^p-spaces, at least for $p < +\infty$. The second part of the chapter deals with interpolation theorem for bounded linear (or sublinear) operators. We will give two examples of such theorems, namely the Marcinkiewicz and the Riesz-Thorin interpolation theorems. These should be thought of as basic examples of larger classes of interpolation theorem, the Marcinkiewicz one being the prototype for the "real method of interpolation" and the Riesz-Thorin theorem being a first representative of the "complex method of interpolation". Already these theorems are quite powerful. However, both these two interpolation theorems have more sophisticated variations which we will only briefly discuss in these notes.

2.1. Convolutions and approximations to the identity

We restrict our attention to the Euclidean case $(\mathbb{R}^n, \mathcal{L}, dx)$. As we have seen the space $L^1(\mathbb{R}^n)$ is a vector space: linear combinations of functions in $L^1(\mathbb{R}^n)$ remain in the space. There is however a "product" defined between elements of $L^1(\mathbb{R}^n)$ that turns L^1 into a Banach algebra. For $f, g \in L^1(\mathbb{R}^n)$ we define the *convolution* of $f * g$ to be the function

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy = \int_{\mathbb{R}^n} g(y)f(x-y)dy.$$

Furthermore, using Fubini's theorem to change the order of integration we can easily see that

$$\|f * g\|_{L^1(\mathbb{R}^n)} \leq \|f\|_{L^1(\mathbb{R}^n)}\|g\|_{L^1(\mathbb{R}^n)}.$$

Thus for $f, g \in L^1(\mathbb{R}^n)$ we have that their convolution $f * g$ is again an element of $L^1(\mathbb{R}^n)$. Note that the previous estimate is the main difficulty in showing that $(L^1(\mathbb{R}^n), *)$ is a Banach algebra.

More generally, the convolution of $f \in L^p(\mathbb{R}^n)$, $1 \leq p \leq +\infty$, and $g \in L^1(\mathbb{R}^n)$, is a well defined element of $L^p(\mathbb{R}^n)$ and we have that

$$\|f * g\|_{L^p(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)}\|g\|_{L^1(\mathbb{R}^n)}. \tag{2.1}$$

EXERCISE 2.1. Use the integral version of Minkowski's inequality to prove estimate (2.1) above.

Let us summarize some properties of convolution in the following proposition. We take the chance to give two definitions here that we will use throughout these notes.

DEFINITION 2.2. Let X be a topological space and $f \in C(X)$ be a continuous function. The *support* of $f : X \to \mathbb{C}$, denoted by $\operatorname{supp}(f)$, is the set

$$\operatorname{supp}(f) = \overline{\{x \in X : f(x) \neq 0\}} = \overline{f^{-1}(\mathbb{C} \setminus \{0\})}.$$

This is the smallest closed set in X outside which $f = 0$.

Observe that we gave the definition of the support of a function for *continuous* functions. This is mostly a technical issue. It is easily understood that, in general, the support of a measurable function can only be defined up to sets of measure zero. The precise definition is as follows.

DEFINITION 2.3. Let μ be a regular Borel measure on a topological space X and $f : X \to \mathbb{C}$ be a Borel measurable function. A point $x \in X$ is called a *support point* for f if

$$\mu(\{y \in U_x : f(y) \neq 0\}) > 0,$$

for every open neighborhood U_x of x. The set

$$\operatorname{supp}(f) := \{x \in X : x \text{ is a support point for } f\}$$

is called the *support* of f.

EXERCISE 2.4. Assume that the measure μ in the previous definition has the additional property that $\mu(U) > 0$ for every open set $U \subset X$. Use Exercise 1.1 to prove that for any continuous function $f : X \to \mathbb{C}$ the two definitions of $\operatorname{supp}(f)$, that is Definition 2.2 and Definition 2.3, coincide.

PROPOSITION 2.5. *Let $f, g, h : \mathbb{R}^n \to \mathbb{C}$ be such that the convolutions below are well defined.*

(i) *(commutative) $f * g = g * f$.*
(ii) *(associative) $(f * g) * h = f * (g * h)$.*
(iii) *(translations) For $x, y \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{C}$ we define the* translation operator

$$\tau_y(f)(x) = f(x - y).$$

For $y \in \mathbb{R}^n$ we have

$$\tau_y(f * g) = (\tau_y f) * g = f * (\tau_y g).$$

(iv) *(support) If $f, g \in C(\mathbb{R}^n)$ then*

$$\operatorname{supp}(f * g) \subset \overline{\{x + y : x \in \operatorname{supp}(f), y \in \operatorname{supp}(g)\}}.$$

PROOF. Statements (i), (ii) and (iii) are trivial consequences of changes of variables and Fubini's theorem. For (iv) observe that if

$$z \notin \overline{\{x + y : x \in \operatorname{supp}(f), y \in \operatorname{supp}(g)\}}$$

then for any $y \in \operatorname{supp}(g)$ we have $z - y \notin \operatorname{supp}(f)$. Thus $g(y)f(z - y) = 0$ for all $y \in \mathbb{R}^n$, so $(f * g)(z) = 0$. $\square$

A very useful property of the convolution of two functions is that it adopts the smoothness of the "nicest" function. Formally this is because any differentiation operator applied to $f * g$ can be transferred to either f or g:

$$\partial^\alpha (f * g) = (\partial^\alpha f) * g = f * (\partial^\alpha g).$$

Here we use the standard multi-index notation: for $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and $f : \mathbb{R}^n \to \mathbb{C}$ we write as usual $\partial^\alpha f = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f$. We also write $|\alpha| = \alpha_1 + \cdots + \alpha_n$. In practice we need one of the functions to have some regularity and some mild conditions on the second function to do this rigorously. For example we have the following:

PROPOSITION 2.6. *Let $f \in L^1(\mathbb{R}^n)$ and suppose that g has continuous partial derivatives up to k-th order, that is $g \in C^k(\mathbb{R}^n)$. Suppose also that $\partial^\alpha g$ is bounded for all $|a| \leq k$. Then $f * g$ has continuous derivatives up to k-th order, i.e. $f * g \in C^k(\mathbb{R}^n)$, and $\partial^\alpha (f * g) = f * (\partial^\alpha g)$.*

SKETCH OF PROOF. Let's just see the special case $n = 1$ and $k = 1$. The proof in the general case is identical. Call $d\mu(y) = f(y)dy$. Since $f \in L^1(\mathbb{R})$, $d\mu$ is a finite, absolutely continuous measure. We then need to show that

$$\frac{d}{dx} \int_{\mathbb{R}} g(x - y) d\mu(y) = \int_{\mathbb{R}} g'(x - y) d\mu(y).$$

Fix some sequence $x_n \to x$. Observe that $g'(x - y) = \lim_n \frac{g(x_n - y) - g(x - y)}{x_n - x} =: \lim_n h_n(x, y)$. By the mean value theorem we have that

$$|h_n(x, y)| \leq \|g'\|_\infty.$$

Using Lebesgue's dominated convergence theorem we get

$$(f * g)'(x) = \lim_n \int_{\mathbb{R}} \frac{g(x_n - y) - g(x - y)}{x - x_n} d\mu(y) = \int_{\mathbb{R}} \lim_n h_n(x, y) d\mu(y) = \int_{\mathbb{R}} g'(x - y) d\mu(y).$$

Observe that the hypothesis on the boundedness of the higher order derivatives will be used to show the uniform boundedness of (the analogues of) the functions $h_n(x, y)$ in the general case. □

2.1.1. The convolution as an averaging operator. It is instructive to fix one function g to be an indicator function, say $g_1(x) = \frac{1}{2} \mathbf{1}_{(-1,1)}(x)$ where the constant $1/2$ is there just in order to normalize the total L^1-mass of the function g_1 to 1. Usually we consider smooth versions of g_1 but let's just stick to case of the characteristic function for the sake of simplicity. Consider the reflection of g_1 give as $\tilde{g}_1(t) = g_1(-t)$. Since we have started with an even function this makes no difference so that $g_1 = \tilde{g}_1$. Observe that we can write

$$f * g(x) = \int f(y) g_1(x - y) dy = \int f(y) \tilde{g}_1(y - x) dy = \int f(y) (\tau_x \tilde{g}_1)(y) dy.$$

For some fixed $x \in \mathbb{R}$, the translations of $\tilde{g}_1$ by $x \in \mathbb{R}$, $\tau_x \tilde{g}_1$, centers the function $\tilde{g}_1$ at the point x. So $\tau_x \tilde{g}_1$ is (a multiple of) the indicator function of an interval of length 2, centered at x. Integrating against $f(y)$ essentially averages the function f around the point x with "weight", the function $\tilde{g}_1$. In this averaging process, our choice of g_1 implies that only the values of f at a scale 1 around x will be important. Thus the convolution of f and g_1, evaluated at $x \in \mathbb{R}$, replaces the value of f at the point x with the average of the values of f at a scale 1 around x. One can take this process one step further and consider sequences of functions that are more or and more concentrated around the origin, but have the same L^1 mass, say 1. For example the second function in this sequence would be $g_2 = \mathbf{1}_{(-\frac{1}{2},\frac{1}{2})}$, the third could be $g_3(x) = 2\mathbf{1}_{(-\frac{1}{4},\frac{1}{4})}$, and so on. Taking convolutions of the function f with the functions $g_1, g_2, g_3, \ldots$ amounts to averaging the function f around every point, in smaller and smaller scales around the point. Intuitively one thinks that, in the limit, one should recover the function itself, at least in some weak sense. This turns out to be indeed the case. But what is the gain in doing so? We just saw that taking convolutions of an integrable (say) function with a smooth bounded function gives us again a smooth function. Thus the previous process allows us to approximate (in some sense) any reasonable function by a sequence of very smooth functions. This has many technical advantages as one can think of any function as a limit, in the appropriate sense, of smooth approximations. This also gives a heuristic explanation of why the convolution of two functions behaves at least as good as the "nicest function" in the convolution; averaging is a smoothing operation.

We will now make the previous heuristic discussion precise. Let ϕ be a function on $\mathbb{R}^n$ and $t > 0$. We define the *dilations* of the function ϕ to be

$$\phi_t(x) = \frac{1}{t^n}\phi(\frac{x}{t}), \quad x \in \mathbb{R}^n.$$

Usually we will have a lot of freedom in choosing the function ϕ and we will require *at least* that $\phi \in L^1(R^n)$. Observe that dilating the function ϕ by $t > 0$ doesn't change the integral:

$$\int_{\mathbb{R}^n} \phi_t(x)dx = \frac{1}{t^n}\int_{\mathbb{R}^n} \phi\Big(\frac{x}{t}\Big)dx = \int_{\mathbb{R}^n} \phi(x)dx.$$

You should think of the function ϕ as a function concentrated around a point as was for example g_1 in the previous discussion or, even better, as smooth approximations of it (bump function). Thus for example ϕ could be a smooth function with compact support around the origin. Observe that as $t \to 0$, the mass of the function ϕ_t, which is constant, becomes more and more concentrated around the origin. We will refer to this construction as an "approximation to the identity". The reason is that, as was mentioned before, one can recover any reasonable function f by convolving with ϕ_t and taking the limit as $t \to 0$, at least in the L^p sense. An alternative motivation for the terminology "approximation to the identity" is that ϕ_t converges (in a weak sense) to a Dirac mass at 0.

THEOREM 2.7. *Let $\phi \in L^1(\mathbb{R}^n)$ with $\int \phi(x)dx = 1$. For $t > 0$ define the dilations of ϕ as before, $\phi_t(x) = t^{-n}\phi(t/x)$. Then, for any $1 \le p < \infty$ we have that $f * \phi_t \to f$ in L^p as $t \to 0$:*

$$\|f * \phi_t - f\|_{L^p(\mathbb{R}^n)} \to 0 \quad as \quad t \to 0.$$

PROOF. For $y \in \mathbb{R}^n$ we use the notation

$$(\tau_y f)(x) = f(x-y),$$

for the translation operator. Using the fact that ϕ_t has integral 1 we can write

$$\begin{aligned}(f*\phi_t)(x)-f(x)&=\int_{\mathbb{R}^n}[f(x-y)-f(x)]\phi_t(y)dy\\&=\int_{\mathbb{R}^n}[f(x-tu)-f(x)]\phi(u)du\\&=\int_{\mathbb{R}^n}[(\tau_{tu}f)(x)-f(x)]\phi(u)du.\end{aligned}$$

By Minkowski's integral inequality we get that

$$\begin{aligned}\|f*\phi_t-f\|_{L^p(\mathbb{R}^n)}&=\left\|\int_{\mathbb{R}^n}[(\tau_{tu}f)(x)-f(x)]\phi(u)du\right\|_{L^p(\mathbb{R}^n)}\\&\leq\int_{\mathbb{R}^n}\|(\tau_{tu}f)(x)-f(x)\|_{L^p(\mathbb{R}^n)}|\phi(u)|du.\end{aligned}$$

Now $\|\tau_{tu}f-f\|_{L^p(\mathbb{R}^n)}\to 0$ as $t\to 0$ (see remark below) and $\|\tau_{tu}f-f\|_{L^p(\mathbb{R}^n)}\leq 2\|f\|_{L^p(\mathbb{R}^n)}$ so by the dominated convergence theorem we get the result. □

REMARK 2.8. The translation operator is continuous in L^p for all $1\leq p<\infty$, that is

$$\|\tau_y f-f\|_{L^p(\mathbb{R}^n)}\to 0\quad\text{as}\quad y\to 0,$$

for all $f\in L^p(\mathbb{R}^n)$, $1\leq p<\infty$. Observe that for $p=\infty$, $\|\tau_y f-f\|_{L^\infty(\mathbb{R}^n)}\to 0$ as $y\to 0$ means that f is uniformly continuous. This explains why the previous theorem breaks down in L^∞.

EXERCISE 2.9. Show that the translation operator is continuous in $L^p(\mathbb{R}^n)$ for $1\leq p<\infty$. Use the fact that continuous functions with compact support are dense in L^p for $1\leq p<\infty$. See also §2.2.

EXERCISE 2.10. Let $\phi\in L^1$ with $\int\phi=1$. Show the following statements:

(i) If f is bounded and uniformly continuous then

$$\|f*\phi_t-f\|_{L^\infty}\to 0\quad\text{as}\quad t\to 0.$$

(ii) If f is bounded and continuous on an open set U show that

$$f*\phi_t\to f\quad\text{as}\quad t\to 0,$$

uniformly on compact subsets of U.

REMARK 2.11. There is a slight abuse of notation here. We use $\|\cdot\|_\infty$ for the norm in the space L^∞ defined in terms of the *essential supremum* of a function. However, the right norm in spaces of continuous functions should be defined in terms of the actual *supremum* of the function. Note however that for a continuous function, the two notions are identical so this should create no confusion.

EXERCISE 2.12. Let $1\leq p\leq\infty$ and p' be its dual exponent. Suppose that $f\in L^p(\mathbb{R}^n)$ and $g\in L^{p'}(\mathbb{R}^n)$. Show that $f*g$ exists for every $x\in\mathbb{R}^n$ and that it is bounded and uniformly continuous. Also show the estimate

$$\|f*g\|_{L^\infty(\mathbb{R}^n)}\leq\|f\|_{L^p(\mathbb{R}^n)}\|g\|_{L^{p'}(\mathbb{R}^n)}.$$

If $1<p<+\infty$ show that $f*g\in C_o(\mathbb{R}^n)$, that is that it is a continuous function that decays to 0 at infinity.

REMARK 2.13. If μ is a finite Borel measure on $\mathbb{R}^n$ and $f \in L^p(\mathbb{R}^n)$ it makes perfect sense to define the convolution of f with μ to be the function

$$(f * \mu)(x) = \int_{\mathbb{R}^n} f(x-y)d\mu(y).$$

We then have

$$\|f * \mu\|_{L^p(\mathbb{R}^n)} \leq \|\mu\| \, \|f\|_{L^p(\mathbb{R}^n)},$$

where μ is the total variation of the measure μ.

2.2. Some dense classes of functions

In this paragraph we will discuss some classes of functions that are dense in the L^p-spaces, at least for $1 \leq p < +\infty$. These will prove to be very useful as many estimates will be easier to establish for these special sub-classes. Also, many times, working with a dense class in L^p, help us avoid several technical difficulties or even define operators that are not obviously defined directly on some L^p space. We will state some of the results here in the generality of a Hausdorff (or locally Hausdorff) space noting that everything goes through for $\mathbb{R}^n$ equipped with the Lebesgue measure.

2.2.1. Simple functions: Let Σ be the class of all simple functions $s : X \to \mathbb{C}$ such that

$$\mu(\{x \in X : s(x) \neq 0\}) < \infty,$$

that is all simple complex valued functions that have support of finite measure. For $1 \leq p < \infty$ the space S is dense in $L^p(X,\mu)$. The space of all simple functions (not necessarily of finite measure support) is dense in L^p for $1 \leq p \leq \infty$.

2.2.2. Continuous functions with compact support: Let $(X, \mathcal{X}, \mu)$ be a measure space, where X is a locally Hausdorff space, $\mathcal{X}$ is a σ-algebra that contains all compact subsets of X and such that

(i) locally finite: $\mu(K) < +\infty$ for all compact sets $K \subset X$.
(ii) μ is inner regular, meaning $\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ is compact.}\}$
(iii) μ is outer regular, meaning $\mu(A) = \inf\{\mu(U) : A \subset U, U \in \mathcal{X} \text{ and } U \text{ is open.}\}$

We denote by $C_c(X)$ the space of continuous functions $f : X \to \mathbb{C}$ with compact support. Then, for every $1 \leq p < \infty$, $C_c(X)$ is dense in $L^p(X,\mu)$.

Remark here that whenever we embed $C_c(X)$ into $L^p(X,\mu)$, $C_c(X)$ automatically inherits the topology induced by the larger space, that is, the one defined by the norm $\|\cdot\|_{L^p(X,\mu)}$. Since L^p spaces are complete under our hypotheses, this says that $L^p(X,\mu)$ is the completion of $C_c(X)$ with respect to the norm of $L^p(X,\mu)$ for $p < \infty$. For $p = \infty$, the completion of $C_c(X)$ with respect to the $\|\cdot\|_{L^\infty(X,\mu)}$ is not $L^\infty(X,\mu)$ but the space of continuous functions on X that *vanish at infinity*.

2.2.3. Continuous functions that vanish at infinity: Let X be a locally compact Hausdorff space (a Hausdorff space where every point has a compact neighborhood). A function $f : X \to \mathbb{C}$ is said to *vanish at infinity* if for every $\epsilon > 0$ there exists a compact set $K \subset X$ such that $|f(x)| < \epsilon$ for all $x \notin K$. We denote by $C_o(X)$ the space of all complex valued continuous functions on X that vanish at infinity.

It is clear that $C_c(X) \subset C_o(X)$, and actually the two spaces coincide whenever X is compact. We can equip the space $C_o(X)$ with the norm

$$\|f\|_\infty = \sup_{x \in X} |f(x)|.$$

THEOREM 2.14. *If X is a locally compact Hausdorff space, then $C_o(X)$ is the completion of $C_c(X)$ with respect to the supremum norm defined above.*

For the proofs of the previous classical results see for example [F] or [R].

All the previous results apply to the Euclidean setup $(\mathbb{R}^n, \mathcal{L}, dx)$. Of course simple functions with support of finite measure are dense in $L^p(\mathbb{R}^n)$ whenever $1 \leq p < +\infty$. A bit more can be said as we can choose our simple functions to be linear combinations of (n-dimensional) bounded intervals, and these are still dense in $L^p(\mathbb{R}^n)$. Continuous functions with compact support are also dense in $L^p(\mathbb{R}^n)$ for all $1 \leq p < \infty$. We can also restrict to a smaller class of more regular functions:

2.2.4. Infinitely differentiable functions with compact support: Let us consider the space of functions $f : \mathbb{R}^n \to C$ which are infinitely differentiable and have compact support. We denote this space by $\mathcal{D}(\mathbb{R}^n) = C_c^\infty(\mathbb{R}^n)$. First of all it is not totally trivial that this space is non-empty.

LEMMA 2.15. *There exists a function $\phi_1 \in \mathcal{D}(\mathbb{R})$. From this we easily conclude that there is a $\phi \in \mathcal{D}(\mathbb{R}^n)$.*

EXERCISE 2.16. Consider the function

$$g(t) = \begin{cases} e^{-\frac{1}{t}} & t > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Show that g, together with its derivatives of any order, is infinitely differentiable and bounded.
(ii) Consider the function $\phi_1(t) = g(1+t)g(1-t)$. Show that $\phi_1(t) = e^{-2/(1-t^2)}$ if $|t| < 1$ and $\phi_1(t) = 0$ otherwise. It is obvious then that $\phi_1 \in \mathcal{D}(\mathbb{R})$.
(iii) For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ consider the function $\phi(x) = \phi_1(x_1) \cdots \phi_1(x_n)$ belongs to $\mathcal{D}(R^n)$.
(iv) For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ consider the function

$$\psi(x) = \begin{cases} e^{-2/(1-|x|^2)}, & |x| < 1, \\ 0, & \text{otherwise.} \end{cases} .$$

Show that $\psi \in \mathcal{D}(\mathbb{R}^n)$.

Obviously $\mathcal{D}(\mathbb{R}^n) \subset C_c(\mathbb{R}^n)$. However, it is not hard to see the space $\mathcal{D}(\mathbb{R}^n)$ is still dense in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$. It will however be easier to show that once we've introduced some more tools from real analysis and, in particular, convolution.

2.2.5. Schwartz functions: Here we introduce the space of *Schwartz* functions $\mathcal{S}(\mathbb{R}^n)$, which will turn out to be extremely useful in what follows. So let $\mathcal{S}(\mathbb{R}^n)$ be the space of all infinitely differentiable (C^∞) functions $f : \mathbb{R}^n \to \mathbb{C}$ such that

$$\sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta f(x)| < \infty,$$

for all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n), \beta = (\beta_1, \ldots, \beta_n)$, of nonnegative integers. In other words, Schwartz functions are smooth functions whose partial derivatives of every order decay faster than any polynomial power at infinity. Of course every function in the class $\mathcal{D}(\mathbb{R}^n)$ is trivially a Schwartz function since it vanishes identically at infinity together with its derivatives of every order. A more interesting example of a Schwartz function is the Gaussian function $\phi : \mathbb{R}^n \to \mathbb{R}$:

$$\phi(x) = e^{-\delta|x|^2}, \quad \delta > 0.$$

The space $\mathcal{S}(\mathbb{R}^n)$ is also dense in all $L^p(\mathbb{R}^n)$ spaces for $1 \leq p < \infty$. Of course this is immediate once one shows that $\mathcal{D}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$.

Schematically we have the following inclusions

$$\mathcal{D}(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n),$$
$$\mathcal{D}(\mathbb{R}^n) \subset C_c(\mathbb{R}^n) \subset L^p(\mathbb{R}^n),$$

and each space in this chain is dense in $L^p(\mathbb{R}^n)$ with the topology induced by $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$. We will discuss the space of Schwartz functions in much more detail in what follows. For now you can think of it as another nice class of functions that is dense in all the spaces $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$.

In the following proposition we use convolutions to show the previous denseness properties:

PROPOSITION 2.17. *The space $\mathcal{D}(\mathbb{R}^n)$, and thus also the space $\mathcal{S}(\mathbb{R}^n)$, is dense in $L^p(\mathbb{R}^n)$ for all $1 \leq p < \infty$. Also the space $\mathcal{D}(\mathbb{R}^n)$ is dense in $C_o(\mathbb{R}^n)$ in the supremum norm.*

PROOF. Let $f \in L^p(\mathbb{R}^n)$ and $\epsilon > 0$. Since the space $C_c(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$, there is a $g \in C_c(\mathbb{R}^n)$ such that

$$\|f - g\|_{L^p(\mathbb{R}^n)} < \frac{\epsilon}{2}.$$

Let $\phi \in \mathcal{D}(\mathbb{R}^n)$ with $\int \phi = 1$. By 2.7 we have that there is $\phi_t * g \to g$ in $L^p(\mathbb{R}^n)$ as $t \to 0$. Thus for t small enough we have that

$$\|g * \phi_t - g\|_{L^p(\mathbb{R}^n)} < \frac{\epsilon}{2}.$$

We conclude that

$$\|g * \phi_t - f\|_{L^p(\mathbb{R}^n)} < \epsilon.$$

It remains to verify that $g * \phi_t$ is in $\mathcal{D}(\mathbb{R}^n)$ for every $t > 0$. Note however that $g * \phi_t$ is smooth by Proposition 2.6. Also, since both g and ϕ_t have compact support, Proposition 2.5 shows that $g * \phi_t$ also has compact support and we are done. Observe that the same argument applies if we start with a $f \in C_o(\mathbb{R}^n)$. Using the fact $C_c(\mathbb{R}^n)$ is dense in $C_o(\mathbb{R}^n)$ it suffices to approximate a function $g \in C_o(\mathbb{R}^n)$. However, functions in $C_o(\mathbb{R}^n)$ are obviously bounded, so Exercise 2.10 completes the proof in this case as well. □

Let us go back to approximations of the identity and justify their name.

EXERCISE 2.18 (convergence of approximations to the identity in the sense of distributions). For $a \in \mathbb{R}^n$ we denote by δ_a the Dirac measure at a:

$$\int_E d\delta_a = \begin{cases} 1, & a \in E, \\ 0, & a \notin E. \end{cases}$$

Let $\phi \in L^1(\mathbb{R}^n)$ with $\int_{\mathbb{R}^n} \phi = 1$ and consider the approximation to the identity $\phi_t(x) = t^{-n}\phi(x/t)$, $t > 0$. Show that

$$\lim_{t\to 0} \int_{\mathbb{R}^n} \phi_t(x)\psi(x)dx = \int_{\mathbb{R}^n} \psi(x)d\delta_0(x),$$

for every $\psi \in \mathcal{D}(\mathbb{R}^n)$. We say that $\phi_t(x)$ (considered as a sequence of finite measures) *converges in the sense of distributions* to the measure $d\delta_0$. We will come back to that point later on in the course.

2.3. Operators on L^p spaces; boundedness and interpolation

Having set up our main environment, the spaces $L^p(X, \mathcal{X}, \mu)$, we come to the core of this introduction: operators acting on these spaces and their properties. In general, we will consider operators T taking functions on some measure space $(X, \mathcal{X}, \mu)$ to function on some other measure space $(Y, \mathcal{Y}, \nu)$. Many times our operators will be initially defined on 'nice functions' such as smooth functions with compact support of Schwartz functions. The goal would then be to extend the operator to a standard normed vector space such as $L^p(X, \mu)$.

Suppose that $(Z, \|\cdot\|_Z)$ and $(W, \|\cdot\|_W)$ are two normed vector spaces (usually Banach spaces of functions) and $T : Z \to W$ be a linear operator, that is, we have

$$T(ax + by) = aTx + bTy,$$

for all $x, y, \in Z$ and complex numbers a, b. We will say that T is bounded if there is a constant $c > 0$ such that $\|Tz\|_W \leq c\|z\|_Z$ for every $z \in Z$. The norm of the operator T, denoted by $\|T\|_{Z\to W}$ or just $\|T\|$, is the smallest constant $c > 0$ so that such an inequality is true. We thus have

$$\|T\| = \sup_{z\in Z} \frac{\|Tz\|_W}{\|z\|_Z} = \sup_{\|z\|_Z=1} \|Tz\|_W.$$

Continuity is equivalent to boundedness for linear operators:

LEMMA 2.19. *Let $T : Z \to W$ be a linear operator. The following are equivalent:*

(i) *The operator T is continuous.*
(ii) *The operator T is continuous at 0.*
(iii) *The operator T is bounded.*

Suppose that we want to show that a linear operator $T : Z \to W$ is a well defined bounded linear operator, where Z, W are Banach spaces. Many times however we can only define the operator on some dense linear subspace $Z_o \subset Z$. Suppose we have then that $T : Z_o \to W$. When can we extend T to the whole space Z? Given $z \in Z$, the obvious thing to do is to consider some sequence $\{z_n\} \subset Z_o$ such that $z_n \to z$. We then need to examine whether the limit Tz_n exists. Suppose that T is bounded on the dense subspace, that is,

$$\|Tz\|_W \leq \|T\|\,\|z\|_Z,$$

for all $z \in Z_o$. Using the boundedness of T on the dense subspace and linearity (essential) we can conclude that

$$\|Tz_m - Tz_n\|_W \leq \|T\|\,\|z_m - z_n\|_Z,$$

so the sequence $\{Tz_n\}$ is a Cauchy sequence. The completeness of W then implies that the limit of $\{Tz_n\}$ does indeed exist, so we can define

$$Tz := \lim_n Tz_n.$$

Observe also that for any other sequence $y_n \to z$ we must have

$$\|Tz_n - Ty_n\|_W \leq \|T(z_n - y_n)\|_W \leq \|T\| \, \|z_n - y_n\|_Z \to 0$$

as $n \to +\infty$. Since both sequences $\{Tz_n\}$ and $\{Ty_n\}$ converge we conclude that they must have the same limit thus the extension is unique. Many times we will only define the operator T on the dense subspace and show its continuity there. We will then say that T is *densely defined*.

We will use this device many times in trying to show that some linear operator $T : L^p \to L^q$ is well defined and bounded, by examining the continuity of T on one of the dense classes that we have considered before (depending on what is more convenient).

A more general class of operators we will come across quite often is that of sublinear operators. Suppose that T is an operator acting on a vector space of measurable functions. Then T is called *sublinear* if $|T(af)| = |a||Tf|$ for all complex constants a and

$$|T(f+g)(x)| \leq |T(f)(x)| + |T(g)(x)|,$$

for all f, g in the vector space. Of course all linear operators are sublinear. However, the most typical example of a sublinear operators we will come across is a maximal type operator. Such an operator has the form

$$Tf = \sup_{t \in \Lambda} |T_t f|,$$

where T_t is a family of linear operators acting on some vector space of measurable functions, Λ is an infinite countable or uncountable index set, and the function $t \to T_t f$ is a measurable function of t. Such operators are called *maximal operators* and the linearity of each T_t guarantees that T is sublinear.

DEFINITION 2.20. Let $0 < p, q \leq \infty$ and T be a sublinear operator mapping functions in $L^q(X, \mu)$ to measurable functions on (Y, ν).

(i) We will say that T is of *strong type* (p, q) if

$$\|Tf\|_{L^q(Y)} \lesssim_{p,q,T,X,Y} \|f\|_{L^p(X)},$$

for all $f \in L^p(X)$, where the implied constant depends only on p, q, X, Y and T. In this case we write $\|T\|_{L^p \to L^q}$ for the norm of the operator $T : L^p(X, \mu) \to L^q(Y, \nu)$.

(ii) We will say that T is of *weak type* (p, q) if

$$\|Tf\|_{L^{q,\infty}(Y,\nu)} \lesssim_{p,q,T,X,Y} \|f\|_{L^p(X,\mu)},$$

for all $f \in L^p(X, \mu)$. We will write $\|T\|_{L^p \to L^{q,\infty}}$ for the norm of the sublinear operator $T : L^p(X, \mu) \to L^{q,\infty}(Y, \nu)$.

Observe that for fixed (p, q), the strong type (p, q) property of T trivially implies that T is of weak type (p, q). The opposite, of course, is not true. However, we will see that in many cases the strong type bound can be deduced by interpolating between suitable endpoint weak type bounds. The first such result is the Marcinkiewicz interpolation theorem.

THEOREM 2.21 (Marcinkiewicz interpolation theorem). *Let (X,μ) and (Y,ν) be measure spaces, $1 \leq p_1 < p_2 \leq \infty$, and let T be a sublinear operator defined on $L^{p_1}(X,\mu) + L^{p_2}(X,\mu)$ and taking values in the space of measurable functions on (Y,ν). Suppose that T is of weak type (p_1,p_1) with constant A_1*

$$\|Tf\|_{L^{p_1}(Y,\nu)} \leq A_1 \|f\|_{L^p_1(X,\mu)},$$

and of weak type (p_2,p_2) with constant A_2

$$\|Tf\|_{L^{p_2}(Y,\nu)} \leq A_2 \|f\|_{L^p_2(X,\mu)}.$$

Then T is of strong type (p,p) for any $p_1 < p < p_2$.

REMARK 2.22. Before going into the proof of this theorem let us discuss a bit its hypothesis. Given a function $f \in L^p(X,\mu)$ we first need to show that $T(f)$ is well defined. Having the information that T is well defined on $L^{p_1} + L^{p_2}$ we essentially need to see that $L^p \subset L^{p_1} + L^{p_2}$ whenever $p_1 < p < p_2$. To see this, fix a positive constant $\beta > 0$, to be defined later, and consider the functions

$$f_1(x) = f(x)\mathbf{1}_{\{x\in X:|f(x)|>\beta\}},$$
$$f_2(x) = f(x)\mathbf{1}_{\{x\in X:|f(x)|\leq\beta\}}.$$

Obviously we have $f(x) = f_1(x) + f_2(x)$. Moreover,

$$\int_X |f_1(x)|^{p_1} d\mu(x) = \int_X |f_1(x)|^p |f_1(x)|^{p_1-p} dx \leq \beta^{p_1-p} \int_X |f(x)|^p dx.$$

Similarly we can estimate

$$\int_X |f_2(x)|^{p_2} d\mu(x) = \int_X |f_2(x)|^p |f_2(x)|^{p_2-p} dx \leq \beta^{p_2-p} \int_X |f(x)|^p dx.$$

This shows that we can decompose any function $f \in L^p(X,\mu)$ to a sum of two functions $f_1 \in L^{p_1}(X,\mu)$ and $f_2 \in L^{p_2}(X,\mu)$, whenever $p_1 < p < p_2$, thus $L^p \subset L^{p_1} + L^{p_2}$. In particular, $T(f)$ is well defined for any $f \in L^p(X,\mu)$.

PROOF. We first prove the theorem when $p_2 < \infty$. Since our hypothesis involves the distribution sets of of Tf it is convenient to recall the representation of the L^p norm of a function in terms of its distribution set. Indeed, from Proposition 1.34 we have

$$\|Tf\|^p_{L^p(Y,\nu)} = \int_Y |Tf(y)|^p d\nu(y) = p\int_0^\infty \lambda^{p-1}\nu(\{y \in X : |Tf(y)| > \lambda\})d\lambda.$$

The measure of the set $\{x \in X : |Tf(y)| > \lambda\}$ will appear many times in the proof so it is convenient to give it a shorter notation:

$$\rho(\lambda) = \nu(\{y \in Y : |Tf(y)| > \lambda\}), \quad \lambda > 0.$$

With this notation

$$\|T(f)\|^p_{L^p(Y,\nu)} = p\int_0^\infty \lambda^{p-1}\rho(\lambda)d\lambda. \tag{2.2}$$

Fix $\lambda > 0$ for a moment and consider the decomposition of the function $f = f_1 + f_2$ at level λ as in the remark before:

$$f_1(x) = f(x)\mathbf{1}_{\{x\in X:|f(x)|>\lambda\}},$$
$$f_2(x) = f(x)\mathbf{1}_{\{x\in X:|f(x)|\leq\lambda\}}.$$

The sublinearity of T allows us to write

$$|Tf(y)| \leq |Tf_1(y)| + |Tf_2(y)|,$$

for any $y \in Y$. Thus,

$$\{|Tf| > \lambda\} \subset \{|Tf_1| > \lambda/2\} \cup \{|Tf_2| > \lambda/2\},$$

so that

$$\rho(\lambda) \leq \nu(\{y \in Y : |Tf_1(y)| > \lambda/2\}) + \nu(\{y \in Y : |Tf_2(y)| > \lambda/2\}).$$

Since $f_1 \in L^{p_1}(X,\mu)$ and T is of weak type (p_1, p_1) we can estimate the first summand as

$$\nu(\{y \in Y : |Tf_1(y)| > \lambda/2\}) \leq (2A_1)^{p_1} \frac{\|f_1\|^{p_1}_{L^{p_1}(X,\mu)}}{\lambda^{p_1}}.$$

Similarly, since $f_2 \in L^{p_2}(X,\mu)$ and T is of weak type (p_2, p_2) we have

$$\nu(\{y \in Y : |Tf_2(y)| > \lambda/2\}) \leq (2A_2)^{p_2} \frac{\|f_2\|^{p_2}_{L^{p_2}(X,\mu)}}{\lambda^{p_2}}.$$

Combining the previous estimates we can write

$$\rho(\lambda) \leq \Big(\frac{2A_1\|f_1\|_{L^{p_1}(X,\mu)}}{\lambda}\Big)^{p_1} + \Big(\frac{2A_2\|f_2\|_{L^{p_2}(X,\mu)}}{\lambda}\Big)^{p_2}$$

Recalling the definitions of f_1, f_2 the previous estimate yields

$$\rho(\lambda) \leq \Big(\frac{2A_1}{\lambda}\Big)^{p_1} \int_{\{x\in X:|f(x)|>\lambda\}} |f(x)|^{p_1} d\mu(x) + \Big(\frac{2A_2}{\lambda}\Big)^{p_2} \int_{\{x\in X:|f(x)|\leq\lambda\}} |f(x)|^{p_2} d\mu(x). \tag{2.3}$$

In order to recover the L^p norm of $T(f)$ observe by (2.2) that it's enough to multiply $\rho(\lambda)$ by $p\lambda^{p-1}$ and integrate in $\lambda \in (0,\infty)$.

Multiplying the first summand on the right hand side of (2.3) by $p\lambda^{p-1}$ and integrating we get

$$\begin{aligned}&(2A_1)^{p_1} p \int_0^\infty \lambda^{p-p_1-1} \int_{\{x\in X:|f(x)|>\lambda\}} |f(x)|^{p_1} d\mu(x)\, d\lambda \\ &= (2A_1)^{p_1} p \int_X |f(x)| \int_0^{|f(x)|} \lambda^{p-p_1-1} d\lambda\, d\mu(x) \; = p\frac{(2A_1)^{p_1}}{p-p_1}\|f\|^p_{L^p(X,\mu)}.\end{aligned}$$

Similarly, multiplying the second summand in (2.3) by $p\lambda^{p-1}$ and integrating we have

$$\begin{aligned}(2A_2)^{p_2} p \int_0^\infty \lambda^{p-p_2-1} \int_{\{x\in X:|f(x)|\leq\lambda\}} |f(x)|^{p_2} d\mu(x)\, d\lambda &= (2A_2)^{p_2} p \int_X |f(x)| \int_{|f(x)|}^\infty \lambda^{p-p_2-1} d\lambda\, d\mu(x) \\ &= p\frac{(2A_2)^{p_2}}{p_2-p}\|f\|^p_{L^p(X,\mu)}.\end{aligned}$$

Summing up the previous two estimates we conclude that

$$\|Tf\|^p_{L^p(Y,\nu)} \leq p\Big(\frac{(2A_1)^{p_1}}{p-p_1} + \frac{(2A_2)^{p_2}}{p_2-p}\Big)\|f\|^p_{L^p(X,\mu)},$$

which shows that T is of strong type (p,p) with

$$\|T\|_{L^p\to L^p} \leq p^{\frac{1}{p}}\Big(\frac{(2A_1)^{p_1}}{p-p_1} + \frac{(2A_2)^{p_2}}{p_2-p}\Big)^{\frac{1}{p}}.$$

Observe that there is no claim here that this quantitative estimate on the norm of T is optimal in general.

The proof in the case $p_2 = \infty$ is very similar. Now the hypothesis that T is of weak type (p_2, p_2) is replaced by the hypothesis that T maps L^∞ to L^∞. That is, there exists some constant $A_2 > 0$, depending only on T and X, such that

$$\|Tg\|_{L^\infty(Y,\nu)} \leq A_2\|g\|_{L^\infty(X,\mu)},$$

for all $g \in L^\infty(X,\mu)$. We fix some level $\lambda > 0$ and we split the function f as $f = f_1 + f_2$ where $f_2(x) = f(x)\mathbf{1}_{\{x\in X:|f(x)|<\lambda/2A_2\}}$. Obviously $f_2 \in L^\infty(X,\mu)$ so by the hypothesis we have that $\|Tf_2\|_{L^\infty(Y,\nu)} \leq A_2\|f_2\|_{L^\infty(X,\mu)} \leq \lambda/2$. Arguing as in the case $p_2 < \infty$ we can write

$$\rho(\lambda) \leq \nu(\{y \in Y : |Tf_1(y)| > \lambda/2\}) + \mu(\{x \in X : |Tf_2(y)| > \lambda/2\}).$$

Since $\|Tf_2\|_{L^\infty(Y,\nu)} \leq \lambda/2$, the second summand in the previous estimate vanishes identically. We conclude that

$$\begin{aligned}
\|Tf\|^p_{L^p(Y,\nu)} &= p\int_0^\infty \lambda^{p-1}\rho(\lambda)d\lambda \leq (2A_1)^{p_1}p\int_0^\infty \lambda^{p-1-p_1}\int_X |f_1(x)|^{p_1}d\mu(x)\,d\lambda\\
&= (2A_1)^{p_1}p\int_0^\infty \lambda^{p-p_1-1}\int_{\{x\in X:|f(x)|>\lambda/(2A_2)\}} |f(x)|^{p_1}d\mu(x)d\lambda\\
&= (2A_1)^{p_1}p\int_X |f(x)|^{p_1}\int_0^{2A_2|f(x)|} \lambda^{p-p_1-1}d\lambda\,d\mu(x)\\
&= \frac{(2A_1)^{p_1}(2A_2)^{p-p_1}}{p-p_1}\|f\|^p_{L^p(X,\mu)}.
\end{aligned}$$

This concludes the proof in the case $p_2 = \infty$ as well as providing the quantitative estimate $\|T\|_{L^p\to L^p} \leq 2\left(\frac{A_1^{p_1}A_2^{p-p_1}}{p-p_1}\right)^{\frac{1}{p}}$. □

EXERCISE 2.23. Modify the proof above to show that under they hypotheses of the Marcinkiewicz interpolation theorem we can conclude that

$$\|T\|_{L^p\to L^p} \leq 2p^{\frac{1}{p}}\left(\frac{1}{p-p_1} + \frac{1}{p-p_2}\right)^{\frac{1}{p}} A_1^{1-\theta}A_2^{\theta},$$

where $\frac{1}{p} := \frac{1-\theta}{p_1} + \frac{\theta}{p_2}$ for some $0 < \theta < 1$.

Hint: This is already the constant appearing in the case $p_2 = \infty$. For the case $p_2 < \infty$ split the function f at the level $c\lambda$ (instead of λ), for some $c > 0$, and optimize in the parameter $c > 0$ at the end of the proof. For this, use the heuristic that a sum is optimized when the terms in the sum are roughly equal in size.

EXERCISE 2.24. Let $0 < p_1 < p_2 \leq \infty$ and suppose that $f \in L^{p_1,\infty}(X,\mu) \cap L^{p_2,\infty}(X,\mu)$. Show that $f \in L^p(X,\mu)$ for all $p_1 < p < p_2$.

Hint: The proof is very similar to the proof of the Marcinkiewicz interpolation theorem, only simpler. Use again the fact that

$$\|f\|^p_{L^p(X,\mu)} = p\int_0^\infty \lambda^{p-1}\mu(\{x \in X : |f(x)| > \lambda\})d\lambda,$$

and split the range of $\lambda \in (0,\infty)$ as $(0,\infty) = (0,\beta)\cup(\beta,\infty)$, at an appropriate level $\beta > 0$. Use the weak integrability conditions for f in the appropriate intervals of λ.

EXERCISE 2.25. Let X be a finite set equipped with counting measure and let $f : X \to \mathbb{C}$ be a function. Show that for any $0 < p < \infty$ we have that

$$\|f\|_{L^{p,\infty}(X)} \leq \|f\|_{L^p(X)} \lesssim_p \log(1 + |X|)\|f\|_{L^{p,\infty}(X)}.$$

Thus on finite sets, the spaces L^p and $L^{p,\infty}$ are equivalent. Here $|X|$ denotes the cardinality of X.

Hint: Observe that $|\{x \in X : |f(x)| > \lambda\}| \leq \min(\|f\|^p_{L^{p,\infty}}/\lambda^p, |X|)$ and use the representation of the L^p norm in terms of the measure of the level sets.

EXERCISE 2.26 (Dual formulation of $L^{p,\infty}$). Let $1 < p \leq \infty$. Show that for every $f \in L^{p,\infty}(X, \mu)$, we have

$$\|f\|_{L^{p,\infty}(X,\mu)} \simeq_p \sup\Big\{\mu(E)^{-\frac{1}{p'}} \int_E |f(x)| dx : 0 < \mu(E) < \infty\Big\},$$

where $\frac{1}{p} + \frac{1}{p'} = 1$.

Hint: As in the previous exercise, write

$$\int_E |f(x)| d\mu(x) = \int_0^\infty \mu(\{x \in E : |f(x)| > \lambda\}) d\lambda.$$

Since the set E has finite measure one can estimate further the measure of the level set by

$$\mu(\{x \in E : |f(x)| > \lambda\}) \leq \min\Big(|E|, \|f\|^p_{L^{p,\infty}}/\lambda^p\Big).$$

Now split the integral we want to estimate accordingly in order to take advantage of this estimate. See also the hint in the previous exercise. This will give you one direction of the estimate, the other direction being trivial.

While the Marcinkiewicz interpolation theorem is the prototype of real interpolation, complex methods can be used to derive similar conclusions. An example of such a method has already been used via the three lines lemma applied to exhibit the log convexity of the L^p norms (which is also a form of interpolation). We will now describe the prototype of complex interpolation.

The following theorem has some differences compared to the Marcinkiewicz interpolation theorem. First of all we assume that T is linear rather than sublinear. Note as well that our hypotheses concern *strong* type bounds for the operator T rather than weak endpoint bounds. On the other hand, the conclusion gives a good estimate for the norm of the operator when interpolating between the endpoints and allows more freedom in the choice of the exponents at the endpoints.

THEOREM 2.27 (Riesz-Thorin interpolation theorem). *Let* $1 \leq p_0, p_1 \leq \infty$ *and* $1 \leq q_0, q_1 \leq \infty$. *Let*

$$T : L^{p_0}(X, \mu) + L^{p_1}(X, \mu) \to L^{q_0}(Y, \nu) + L^{q_1}(Y, \nu),$$

be a linear operator that is of strong type (p_0, q_0) *with norm* k_0 *and of strong type* (p_1, q_1) *with norm* k_1. *That is we have that*

$$\|Tf\|_{L^{q_0}(Y,\nu)} \leq k_0\|f\|_{L^{p_0}(X,\mu)},$$

for all $f \in L^{p_0}(X, \mu)$ *and*

$$\|Tf\|_{L^{q_1}(Y,\nu)} \leq k_1\|f\|_{L^{p_1}(X,\mu)},$$

for all $f \in L^{p_1}(X,\mu)$. *Then* T *is of strong type* (p_θ, q_θ) *with norm at most* $k_\theta = k_0^{1-\theta} k_1^\theta$*:*

$$\|Tf\|_{L^{q_\theta}(Y,\nu)} \leq k_\theta \|f\|_{L^{p_\theta}(X,\mu)},$$

for all $f \in L^{p_\theta}(X,\mu)$*, where* $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$ *and* $\frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}$*, with* $0 \leq \theta \leq 1$.

PROOF. We divide the proof in several steps:
step 1: It is enough to prove the theorem for $k_0 = k_1 = k_\theta = 1$. To see this just observe that we can always replace the measures μ, ν by $c_\mu \mu$, $c_\nu \nu$ respectively, for appropriate constants $c_\mu, c_\nu > 0$. We can choose these constants so that $k_0 = k_1 = 1$ and then we also have $k_\theta = 1$. Doing the calculations you will see that we need to define the constants c_μ, c_ν by means of the equations

$$c_\nu^{\frac{1}{q_0}} c_\mu^{-\frac{1}{p_0}} k_0 = 1 \quad \text{and} \quad c_\nu^{\frac{1}{q_1}} c_\mu^{-\frac{1}{p_1}} k_1 = 1.$$

In what follows we will therefore assume that $k_0 = k_1 = k_\theta = 1$ in the statement of the theorem.
step 2: Let us now get rid of the easy case. If $p_0 = p_1 = p_\theta$ then by the log-convexity of the L^p norm we get directly that

$$\|Tf\|_{L^{q_\theta}} \leq \|Tf\|_{L^{q_0}}^{1-\theta} \|Tf\|_{L^{q_1}}^{\theta} \leq \|f\|_{L^{p_\theta}},$$

as desired. Thus, for the rest of the proof we can and will assume that $p_0 < p_1$ and $1 < p_\theta < +\infty$.
step 3: We have that

$$\Big| \int_Y (Tf) g d\nu \Big| \leq \|f\|_{L^{p_\theta}} \|g\|_{L^{q'_\theta}}, \tag{2.4}$$

for all simple functions of finite measure support f, g. Here q'_θ is the dual exponent of q_θ.

First of all, since T is of strong type (p_0, q_0), Hölder's inequality shows that

$$\Big| \int_Y (Tf) g d\nu \Big| \leq \|f\|_{L^{p_0}} \|g\|_{L^{q'_0}}, \tag{2.5}$$

and, similarly, by the (p_1, q_1) type of T we get that

$$\Big| \int_Y (Tf) g d\nu \Big| \leq \|f\|_{L^{p_1}} \|g\|_{L^{q'_1}}. \tag{2.6}$$

Thus, estimate (2.4) is true for $\theta = 0, 1$. It is obvious that we need to interpolate between the two endpoint estimates above. We will do that by means of the three lines convexity lemma. First we define the map

$$\mathbb{C} \ni z \mapsto F(z) = \int_Y \Big(T\Big[|f|^{(1-z)p_\theta/p_0 + z p_\theta/p_1} \operatorname{sgn}(f) \Big] \Big) |g|^{(1-z)q'_\theta/q'_0 + z q'_\theta/q'_1} \operatorname{sgn}(g) d\nu,$$

where $\operatorname{sgn}(h) = h/|h|$. In the case $q_0 = q_1 = q_\theta = 1$ then we understand that $q'_\theta/q'_0 = q'_\theta/q'_1 = 1$. The function F is a holomorphic function of z. Furthermore, since f, g are simple functions of finite measure support, it is not hard to see that F is actually bounded on the strip $S = \{z = x + iy : y \in \mathbb{R}, 0 \leq x \leq 1\}$. Furthermore, for $z = \theta + 0i$ we see that $F(\theta) = \int_Y (Tf) g$. Now, on the boundary of the strip we have that

$$|F(0 + iy)| \leq \|f\|_{L^{p_\theta}}^{\frac{p_\theta}{p_0}} \|g\|_{L^{q'_\theta}}^{\frac{q_\theta{}'}{q_0{}'}}.$$

from (2.5) and similarly

$$|F(1+iy)| \leq \|f\|_{L^{p_\theta}}^{\frac{p_\theta}{p_1}} \|g\|_{L^{q'_\theta}}^{\frac{q_\theta{}'}{q_1{}'}}.$$

from (2.6). Using the three lines lemma we get that

$$|F(x+iy)| \leq \|f\|_{L^{p_\theta}}^{\frac{(1-x)p_\theta}{p_0}} \|g\|_{L^{q'_\theta}}^{\frac{(1-x)q_\theta{}'}{q_0{}'}} \|f\|_{L^{p_\theta}}^{\frac{xp_\theta}{p_1}} \|g\|_{L^{q'_\theta}}^{\frac{xq_\theta{}'}{q_1{}'}}.$$

The right hand side however is equal to $\|f\|_{L^{p_\theta}}\|g\|_{L^{q'_\theta}}$. Applying the result for $x=\theta$ and $y=0$ we get the claim of step 2. Observe that nothing really changes in the case $q_0=q_1=q_\theta=1$.

step 4: Let f,g be any simple functions of finite measure support. Then $Tf \in L^{q_0}+L^{q_1}$ by the hypothesis thus $gTf \in L^1$ for all simple functions g with finite measure support. Now Step 3 together with Proposition 1.29 show that

$$\|Tf\|_{L^{q_\theta}} \leq \|f\|_{L^{p_\theta}}$$

for all functions f of finite measure support. Since the set of simple functions of finite measure support is dense in L^{p_θ} (remember that $p_\theta < +\infty$), the discussion in the beginning of §2.3 shows that the operator T has a unique extension to a bounded linear operator $\tilde{T}$ on L^{p_θ} and satisfies the bounds

$$\|\tilde{T}f\|_{L^{q_\theta}} \leq \|f\|_{L^{p_\theta}}$$

for all $f \in L^{p_\theta}$. It remains to check that this extension coincides with the original operator T.

step 5: Let $f \in L^{p_\theta}$ and $\{f_n\}$ be a sequence of simple functions of finite measure support such that $|f_1| \leq \cdots \leq |f_n| \leq \cdots \leq |f|$ and $|f_n| \nearrow |f|$ almost everywhere. Let $g := f\mathbf{1}_{\{|f|\leq 1\}}$, $b := f\mathbf{1}_{\{|f|>1\}}$, $g_n := f_n\mathbf{1}_{\{|f|\leq 1\}}$, $b_n := f_n\mathbf{1}_{\{|f|>1\}}$. We have that $g_n, g \in L^{p_\theta}\cap L^{p_1}$ and $b_n, b \in L^{p_0}\cap L^{p_\theta}$ and

$$\|b_n-b\|_{p_0}, \|b_n-b\|_{p_\theta} \to 0,$$

and

$$\|g_n-g\|_{L^{p_1}}, \|g_n-g\|_{p_\theta} \to 0,$$

as $n\to\infty$. By the hypotheses we have that

$$\left\|T(b_n-b)\right\|_{L^{p_0}} \leq \|b-b_n\|_{L^{p_0}} \to 0 \quad \text{and} \quad \left\|T(g_n-g)\right\|_{L^{p_1}} \leq \|g-g_n\|_{L^{p_1}} \to 0.$$

Passing to subsequences, if necessary, we conclude that $Tf_n = \tilde{T}(b_n+g_n) \to Tf$ almost everywhere. Since $T,\tilde{T}$ coincide on the set of simple functions with finite measure support we conclude that

$$\|Tf\|_{L^{q_\theta}} = \|\lim_n Tf_n\|_{L^{q_\theta}} = \|\lim_n Tf_n\|_{L^{q_\theta}} \leq \liminf_n \|f_n\|_{L^{q_\theta}} = \|f\|_{L^{q_\theta}}$$

by Fatou's lemma. Thus T satisfies the same estimate as $\tilde{T}$ on L^{q_θ} which means that the two operators coincide. Indeed, if f_n is any sequence of simple functions of finite measure support with $f_n \to f$ in L^{p_θ} we get

$$\|Tf-\tilde{T}f\|_{L^{q_\theta}} = \|T(f-f_n)+\tilde{T}(f-f_n)+\tilde{T}f_n - Tf_n\|_{L^{q_\theta}} \leq 2\|f-f_n\|_{L^{p_\theta}} \to 0,$$

so that $T\equiv\tilde{T}$. □

As a first application of the Riesz-Thorin interpolation theorem we will now prove Young's inequality on convolutions of functions.

PROPOSITION 2.28 (Young's inequality). *Let $f, g : \mathbb{R}^n \to \mathbb{C}$. Let $1 \le p, q, r \le \infty$ be such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$. If $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, then $f * g$ is a well defined function in $L^r(\mathbb{R}^n)$ and we have the estimate*

$$\|f * g\|_{L^r(\mathbb{R}^n)} \le \|f\|_{L^p(\mathbb{R}^n)} \|g\|_{L^q(\mathbb{R}^n)}.$$

PROOF. For $1 \le q \le \infty$ and $g \in L^q(\mathbb{R}^n)$ fixed we define the operator

$$T(f) = g * f.$$

As we have already seen (see Exercise 2.1) we have the bound $\|T(f)\|_{L^q} \le \|g\|_{L^q} \|f\|_{L^1}$, that is, T is of strong type $(1, q)$. It is also very easy to see that if q' is the conjugate exponent of q then we have

$$|(f * g)(x)| = \Big| \int f(x - y) g(y) dy \Big| \le \|f\|_{L^{q'}} \|g\|_{L^q},$$

that is $\|T(f)\|_{L^\infty} \le \|g\|_{L^q} \| \|f\|_{L^{q'}}$ and T is of strong type (q', ∞). Letting $\frac{1}{q_\theta} = \frac{1}{r} = \frac{1-\theta}{q} + \frac{\theta}{\infty}$ and $\frac{1}{p_\theta} = \frac{1-\theta}{1} + \frac{\theta}{q'}$, the Riesz-Thorin interpolation theorem shows that T is of strong type (p_θ, q_θ). Replacing $1 - \theta = q/r$ and using the hypothesis $1/p + 1/q = 1/r + 1$ we get that $p_\theta = p$. Thus we conclude that T is of strong type (p, r) with norm at most $\|g\|_{L^q}^{1-\theta} \|g\|_{L^q}^{\theta} = \|g\|_{L^q}$. That is we have $\|f * g\|_{L^r} \le \|g\|_{L^q} \|f\|_{L^p}$ as we wanted to show. □

EXERCISE 2.29 (Schur's test). Let $1 \le p_1, q_0 \le \infty$ and $B_0, B_1 > 0$. Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces and $K : X \times Y \to \mathbb{C}$ be a $\mathcal{X} \otimes \mathcal{Y}$-measurable function such that

(i) For almost every $x \in X$ we have that

$$\|K(x, \cdot)\|_{L^{q_0}(Y,\nu)} \le B_0.$$

(ii) For almost every $y \in Y$ we have that

$$\|K(\cdot, y)\|_{L^{p'_1}(X,\mu)} \le B_1.$$

Define $p_0 = 1$ and $q_1 = \infty$. Show that the operator T, defined as

$$Tf(x) := \int K(x, y) f(x) d\mu(x)$$

is of strong type (p_θ, q_θ) with norm at most $B_0^\theta B_1^{1-\theta}$ and

$$\frac{1}{q_\theta} := \frac{1-\theta}{q_0} + \frac{\theta}{q_1}, \quad \frac{1}{p_\theta} := \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$$

as in the Riesz-Thorin interpolation theorem.

CHAPTER 3

The Fourier transform and the space of tempered distributions

3.1. The Fourier transform on $L^1(\mathbb{R}^n)$.

For $f \in L^1(\mathbb{R}^n)$, the *Fourier transform* of f is the function

$$\mathcal{F}(f)(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}dx, \quad \xi \in \mathbb{R}^n.$$

Here $x \cdot y = \langle x, y\rangle$ denotes the *inner product* of $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$:

$$x \cdot y = \langle x, y\rangle = x_1y_1 + \cdots x_ny_n.$$

Observe that this inner product in $\mathbb{R}^n$ is compatible with the Euclidean norm since $x \cdot x = |x|^2$. It is easy to see that the integral above converges for every $\xi \in \mathbb{R}^n$ and that the Fourier transform of an L^1 function is a uniformly continuous function.

THEOREM 3.1. *Let $f, g \in L^1(\mathbb{R}^n)$. We have the following properties.*

(i) *The Fourier transform is linear $\widehat{f+g} = \hat{f} + \hat{g}$ and $\widehat{cf} = c\hat{f}$ for any $c \in \mathbb{C}$.*
(ii) *The function $\hat{f}(\xi)$ is uniformly continuous.*
(iii) *The operator $\mathcal{F}$ is bounded from $L^1(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$ and*

$$\|\hat{f}\|_{L^\infty(\mathbb{R}^n)} \leq \|f\|_{L^1(\mathbb{R}^n)}.$$

(iv) *(Riemann-Lebesgue) We have that*

$$\lim_{|\xi|\to+\infty} \hat{f}(\xi) = 0.$$

PROOF. Properties (i), (ii) and (iii) are easy to establish and are left as an exercise. There are several ways to see (iv) based on the idea that it is enough to establish this property for a dense subspace of $L^1(\mathbb{R}^n)$. For example, observe that if f is the indicator function of an interval of the real line, $f = \mathbf{1}_{[a,b]}$, then we can calculate explicitly to show that

$$|\hat{f}(\xi)| = \left|\int_a^b e^{-2\pi i x\xi}dx\right| = \left|\frac{e^{-2\pi i\xi a} - e^{-2\pi i\xi b}}{2\pi i\xi}\right| \lesssim \frac{1}{|\xi|} \to 0 \quad \text{as} \quad |\xi| \to +\infty.$$

Tensoring this one dimensional result one easily shows that $\lim_{|\xi|\to+\infty} f(\xi) = 0$ whenever f is the indicator function of an n-dimensional interval of the form $[a_1, b_1] \times \cdots \times [a_n, b_n]$. Obviously the same is true for finite linear combinations of n-dimensional intervals since the Fourier transform is linear.

Now let f be any function in $L^1(\mathbb{R}^n)$ and $\epsilon > 0$ and consider a finite linear combination g of indicators of n-dimensional intervals with $\|f - g\|_1 < \epsilon/2$. Let also $M > 0$ be large enough so that $|\hat{g}(\xi)| < \epsilon/2$ whenever $|\xi| > M$. Using (iii) and the linearity of the Fourier transform we have that

$$|\hat{f}(\xi)| \leq |(\widehat{f-g})(\xi)| + |\hat{g}(\xi)| \leq \|f-g\|_{L^1} + |\hat{g}(\xi)| < \epsilon,$$

whenever $|\xi| > M$, which finishes the proof. □

In view of (ii) and (iv) we immediately get the following.

COROLLARY 3.2. *If $f \in L^1(\mathbb{R}^n)$ then $\hat{f} \in C_o(\mathbb{R}^n)$.*

EXERCISE 3.3. Show the properties (ii) and (iii) in the previous Theorem.

The discussion above and especially Corollary 3.2 shows that a necessary condition for a function g to be a Fourier transform of some function in $L^1(\mathbb{R}^n)$ is $g \in C_o(\mathbb{R}^n)$. However, this condition is not sufficient as there are functions $g \in C_o(\mathbb{R}^n)$ which are not Fourier transforms of L^1 functions. See Exercise 3.20.

Let us now see two important examples of Fourier transforms that will be very useful in what follows.

EXAMPLE 3.4. For $a > 0$ let $f(x) = e^{-\pi a|x|^2}$. Then

$$\hat{f}(\xi) = a^{-\frac{n}{2}} e^{-\frac{\pi|\xi|^2}{a}}.$$

PROOF. Observe that in one dimension we have

$$\begin{aligned}\hat{f}(\xi) &= \int_{\mathbb{R}} e^{-\pi a x^2} e^{-2\pi i x\xi} dx = \int_{\mathbb{R}} e^{-\pi a(x+i\frac{\xi}{a})^2} dx\, e^{-\frac{\pi\xi^2}{a}} \\ &= \int_{\mathbb{R}} e^{-\pi a x^2} dx\, e^{-\frac{\pi\xi^2}{a}} = a^{-\frac{1}{2}} e^{-\frac{\pi^2\xi^2}{a}},\end{aligned}$$

where the third equality is a consequence of Cauchy's theorem from complex analysis. The n-dimensional case is now immediate by tensoring the one dimensional result. □

REMARK 3.5. Replacing $a = 1$ in the previous example we see that $e^{-\pi|x|^2}$ is its own Fourier transform.

EXAMPLE 3.6. For $a > 0$ let $g(x) = e^{-2\pi a|x|}$. Then

$$\hat{g}(\xi) = c_n \frac{a}{(a^2 + |\xi|^2)^{\frac{n+1}{2}}},$$

where $c_n = \Gamma((n+1)/2)/\pi^{\frac{n+1}{2}}$.

PROOF. The first step here is to show the subordination identity

$$e^{-\beta} = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-\beta^2/4u} du, \quad \beta > 0, \tag{3.1}$$

which is a simple consequence of the identities

$$\begin{aligned} e^{-\beta} &= \frac{2}{\pi} \int_0^\infty \frac{\cos\beta x}{1+x^2} dx, \\ \frac{1}{1+x^2} &= \int_0^\infty e^{-(1+x^2)u} du. \end{aligned}$$

Using (3.1) we can write

$$\begin{aligned}\hat{g}(\xi) &= \int_{\mathbb{R}^n} e^{-2\pi a|x|} e^{-2\pi i x\cdot\xi} dx = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}^n} \Big(\int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-4\pi^2 a^2 |x|^2/4u} du \Big) e^{-2\pi i x\cdot\xi} dx \\ &= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} \frac{1}{a^n} \Big(\sqrt{\frac{u}{\pi}} \Big)^{\frac{n}{2}} e^{-\frac{u|\xi|^2}{a^2}} du = \frac{1}{\pi^{\frac{n+1}{2}} a^n} \int_0^\infty u^{\frac{n-1}{2}} e^{-u\frac{|\xi|^2}{a^2}} e^{-u} du \\ &= \frac{1}{\pi^{\frac{n+1}{2}} a^n} \frac{1}{\Big(1+\frac{|\xi|^2}{a^2}\Big)^{\frac{n+1}{2}}} \int_0^\infty u^{\frac{n-1}{2}} e^{-u} du = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}} \frac{a}{\big(a^2+|\xi|^2\big)^{\frac{n+1}{2}}},\end{aligned}$$

by the definition of the Γ-function. □

EXERCISE 3.7. This exercise gives a first (qualitative) instance of the *uncertainty principle*. Prove that there does not exist a non-zero integrable function on $\mathbb{R}$ such that both f and $\hat{f}$ have compact support.

Hint: Observe that the function

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i x\xi} dx,$$

extends to an entire function (why ?).

The definition of the Fourier transform extends without difficulty to finite Borel measures on $\mathbb{R}^n$. Let us denote by $\mathcal{M}(\mathbb{R}^n)$ this class of finite Borel measures and let $\mu \in \mathcal{M}(\mathbb{R}^n)$. We define the Fourier transform of μ to be the function

$$\mathcal{F}(\mu)(\xi) = \hat{\mu}(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i x\cdot\xi} d\mu(x), \quad \xi \in \mathbb{R}^n.$$

We have the analogues of (i), (ii) and (iii) of Theorem 3.1 if we replace the L^1 norm by the total variation of the measure. However property (iv) fails as can be seen by considering the Fourier transform of a Dirac mass at the point 0. Indeed, observe that

$$\widehat{\delta_0}(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i x\cdot\xi} d\delta_0(x) = 1,$$

which is a constant function. This remark can be used as a first instance of the heuristic that "regularity of the function implies decay of the Fourier transform". Observe that $f \in L^1(\mathbb{R}^n)$ implies that $\hat{f}$ decays to 0 at infinity, while $\mu \in \mathcal{M}(\mathbb{R}^n)$ does not imply any decay. Here the regularity hypothesis is quite weak, $f \in L^1$; functions in L^1 are however more regular than general measures $\mu \in \mathcal{M}(\mathbb{R}^n)$.

The Fourier transform interacts very nicely with convolutions of functions, turning them to products. This turns out to be quite important when considering translation invariant operators as we shall see later on in the course.

PROPOSITION 3.8. *Let $f, g \in L^1(\mathbb{R}^n)$. Then $\widehat{f * g} = \hat{f}\hat{g}$.*

EXERCISE 3.9. Prove Proposition 3.8.

Another elementary but important property of Fourier transforms is the *multiplication formula*.

PROPOSITION 3.10 (Multiplication formula). *Let* $f, g \in L^1(\mathbb{R}^n)$. *Then*

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

We will now describe some easily verified symmetries of the Fourier transform. We introduce the following basic operations on functions:

- Translation operator: $(\tau_{x_o} f)(x) := f(x - x_o), \quad x, x_o \in \mathbb{R}^n$
- Modulation operator: $\mathrm{Mod}_{x_o}(f)(x) := e^{2\pi i x \cdot x_o} f(x), \quad x, x_o \in \mathbb{R}^n$
- Dilation operator: $\mathrm{Dil}^p_\lambda(f)(x) := \lambda^{-\frac{n}{p}} f(x/\lambda), \quad x \in \mathbb{R}^n, \lambda > 0, 1 \le p \le \infty.$

PROPOSITION 3.11. *Let* $f \in L^1(\mathbb{R}^n)$ *We have the following symmetries:*

(i) $\mathcal{F}\tau_{x_o} = \mathrm{Mod}_{-x_o}\mathcal{F}$,
(ii) $\mathcal{F}\mathrm{Mod}_{\xi_o} = \tau_{\xi_o}\mathcal{F}$,
(iii) $\mathcal{F}\mathrm{Dil}^p_\lambda = \mathrm{Dil}^{p'}_{\lambda^{-1}}\mathcal{F}$, *where* $\frac{1}{p} + \frac{1}{p'} = 1$.

EXERCISE 3.12. Prove the symmetries in Proposition 3.11 above. Also, let $U : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation, that is, $U \in GL(\mathbb{R}^n)$. Define the general dilation operator

$$(\mathrm{Dil}^p_U f)(x) = |\det U|^{-\frac{1}{p}} f(U^{-1}x), \quad x \in \mathbb{R}^n, 1 \le p \le \infty.$$

Prove that

$$\mathcal{F}\mathrm{Dil}^p_U = \mathrm{Dil}^{p'}_{(U^*)^{-1}}\mathcal{F},$$

where U^* is the (real) adjoint of U, that is the matrix for which we have $\langle Ux, y\rangle = \langle x, U^*y\rangle$ for all $x, y \in \mathbb{R}^n$.

We now come to one of the most interesting properties of the Fourier transform, the way it commutes with derivatives.

PROPOSITION 3.13. *We have the following statements:*

(a) *Suppose that* $f \in L^1(\mathbb{R}^n)$ *and that* $x_k f(x) \in L^1(\mathbb{R}^n)$ *for some* $1 \le k \le n$. *Then* $\hat{f}$ *is differentiable with respect to* ξ_k *and*

$$\frac{\partial}{\partial \xi_k}\mathcal{F}(f)(\xi) = \mathcal{F}(-2\pi i x_k f)(\xi).$$

(b) *We will say that a function* f *has a partial derivative in the* L^p *norm with respect to* x_k *if there exists a function* $g \in L^p(\mathbb{R}^n)$ *such that*

$$\lim_{h_k \to 0} \left(\int_{\mathbb{R}^n} \left| \frac{f(x+h) - f(x)}{h_k} - g \right|^p dx \right)^{\frac{1}{p}} = 0,$$

where $h = (0, \ldots, 0, h_k, 0, \ldots, 0)$ *is a non-zero vector along the k-th coordinate axis. If* f *has a partial derivative* g *with respect to* x_k *in the* L^1*-norm, then*

$$\hat{g}(\xi) = 2\pi i \xi_j \hat{f}(\xi).$$

EXERCISE 3.14. Prove Proposition 3.13.

A similar result that involves the classical derivatives of a function is the following:

PROPOSITION 3.15. *For k a non-negative integer, suppose that $f \in C^k(\mathbb{R}^n)$ and that $\partial^\alpha f \in L^1(\mathbb{R}^n)$ for all $|\alpha| \leq k$, and $\partial^\alpha f \in C_o(\mathbb{R}^n)$ for $|\alpha| \leq k-1$. Show that*

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

EXERCISE 3.16. Prove Proposition 3.15.

Several remarks are in order. First of all observe that Proposition 3.13 and Proposition 3.15 assert that the following commutation relations are true

(i) $\mathcal{F}(-2\pi i x_k) = \partial_{\xi_k}\mathcal{F}$,
(ii) $\mathcal{F}\partial x_k = (2\pi i \xi_k)\mathcal{F}$,

where here we abuse notation and denote by $2\pi i x_k$ the operator of *multiplication* by $2\pi i x_k$. Thus the Fourier transform turns derivatives to multiplication by the corresponding variable, and vice versa, it turns multiplication by the coordinate variable to a partial derivative, whenever this is technically justified. This is a manifestation of the heuristic principle that smoothness of a function translates to decay of the Fourier transform and on the other hand, decay of a function at infinity translates to smoothness of the Fourier transform.

A second remark is that these commutation relations generalize in an obvious way to higher derivatives. To make this more precise let P be a polynomial on $\mathbb{R}^n$:

$$P(x) = \sum_{|\alpha|\leq d} c_\alpha x^\alpha.$$

Slightly abusing notation again we write $P(\partial_x^\alpha)$ for the differential operator

$$P(\partial_x) = \sum_{|\alpha|\leq d} c_\alpha \partial_x^\alpha = \sum_{|\alpha|\leq d} c_\alpha \partial_{x_1}^{\alpha_1}\cdots\partial_{x_n}^{\alpha_n}.$$

We then have that the following commutation relations are true

(i′) $\mathcal{F}P(-2\pi i x) = P(\partial_\xi^\alpha)\mathcal{F}$,
(ii′) $\mathcal{F}P(\partial_x^\alpha) = P(2\pi i \xi)\mathcal{F}$.

Observe that for “nice” functions, for example $f \in C_c^\infty(\mathbb{R}^n)$ or $f \in \mathcal{S}(R^n)$, Propositions 3.13 and 3.15 are automatically satisfied.

3.1.1. Inverting the Fourier transform. On of the most important problems in the theory of Fourier transforms is that of the *inversion* of the Fourier transform. That is, given the Fourier transform $\hat{f}$ of an L^1 function, when can we recover the original function f from $\hat{f}$? We begin with a simple case where the recovery is quite easy.

PROPOSITION 3.17. *Let $f \in L^1(\mathbb{R}^n)$ be such that $\hat{f} \in L^1(\mathbb{R}^n)$. Then the* inversion formula *holds true. In particular we have that*

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

for almost every $x \in \mathbb{R}^n$.

PROOF. The proof is based on the following calculation. For $a > 0$ we have that

$$\begin{aligned}\int_{\mathbb{R}^n} \hat{f}(\xi)e^{-a|\xi|^2}e^{2\pi i x\cdot\xi}d\xi &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(y)e^{-2\pi i y\cdot\xi}dye^{-a|\xi|^2}e^{2\pi i x\cdot\xi}d\xi \\ &= \int_{\mathbb{R}^n} f(x+y)\int_{\mathbb{R}^n} e^{-2\pi i y}e^{-a|\xi|^2}d\xi dy \\ &= (\frac{\pi}{a})^{\frac{n}{2}}\int_{\mathbb{R}^n} f(x+y)e^{-\frac{\pi^2|y|^2}{a}}dy \\ &= \int_{\mathbb{R}^n} f(x+\sqrt{a}y)e^{-\pi|y|^2}dy,\end{aligned}$$

where in the last equality we have used Example 3.4. We can thus write

$$\begin{aligned}\int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^n}\hat{f}(\xi)e^{-a|\xi|^2e^{2\pi i x\cdot\xi}}d\xi - f(x)\right|dx &= \int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^n} f(x+\sqrt{a}y)e^{-\pi|y|^2}dy - f(x)\right|dx \\ &= \int_{\mathbb{R}^n}\left|\int_{\mathbb{R}^n}\{f(x+\sqrt{a}y)-f(x)\}e^{-\pi|y|^2}dy\right|dx \\ &\leq \int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|f(x+\sqrt{a}y)-f(x)|dxe^{-\pi|y|^2}dy \\ &= \int_{\mathbb{R}^n}\|f-\tau_{-\sqrt{a}y}f\|_{L^1(\mathbb{R}^n)}e^{-\pi|y|^2}dy.\end{aligned}$$

Since $\|f-\tau_{-\sqrt{a}y}f\|_{L^1(\mathbb{R}^n)} \to 0$ as $a \to 0$ and $\|f-\tau_{-\sqrt{a}y}f\|_{L^1(\mathbb{R}^n)} \leq 2\|f\|_{L^1(\mathbb{R}^n)}$, Lebesgue's dominated convergence theorem shows that f is almost everywhere equal to the L^1-limit of the sequence of functions

$$g_a(x) = \int_{\mathbb{R}^n}\hat{f}(\xi)e^{-a|\xi|^2}e^{2\pi i x\cdot\xi}d\xi,$$

as $a \to 0$ (technically speaking we need to consider a sequence $a_k \to 0$). On the other hand since $\hat{f} \in L^1(\mathbb{R}^n)$, another application of Lebesgue's dominated theorem shows that the L^1-limit of the functions g_a is also equal to $\int_{\mathbb{R}^n}\hat{f}(\xi)e^{2\pi i x\cdot\xi}d\xi$. This completes the proof of the proposition. □

An immediate corollary is that the Fourier transform is a one-to-one operator:

COROLLARY 3.18. *Let $f_1, f_2 \in L^1(\mathbb{R}^n)$ and suppose that $\hat{f_1}(\xi) = \hat{f_2}(\xi)$ for all $\xi \in \mathbb{R}^n$. The we have that $f_1(x) = f_2(x)$ for almost every $x \in \mathbb{R}^n$.*

The proof is an obvious application of Proposition 3.17.

EXERCISE 3.19. Show the following

(i) If $f \in C_c^{n+1}(\mathbb{R}^n)$ then

$$|\hat{f}(\xi)| \lesssim (1+|\xi|^2)^{-(n+1)/2}.$$

Conclude that whenever $f \in C_c^{n+1}(\mathbb{R}^n)$, we have that

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

(ii) Show that $\mathcal{F}$ maps the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ onto $\mathcal{S}(\mathbb{R}^n)$.

EXERCISE 3.20. The purpose of this exercise is to show that $\mathcal{F}(L^1(\mathbb{R}^n))$ is a *proper* subset of $C_o(\mathbb{R}^n)$ but also that it is a *dense* subset of $C_o(\mathbb{R}^n)$.

(i) Show that $\mathcal{F}(L^1(\mathbb{R}))$ is a proper subset of $C_o(\mathbb{R})$.

Hint: While there are different ways to do that, a possible approach is the following. For simplicity we just consider the case $n = 1$:

(a) Show that $\left|\int_a^b \frac{\sin x}{x} dx\right| \leq B$ for all $0 \leq |a| < |b| < \infty$ where $B > 0$ is a numerical constant that does not depend on a, b.

(b) Suppose that $f \in L^1(\mathbb{R})$ is such that $\hat{f}$ is an odd function. Use (a) to show that for every $b > 0$ we have that

$$\left|\int_1^b \frac{\hat{f}(\xi)}{\xi} d\xi\right| < A,$$

for some numerical constant $A > 0$ which *does not depend on* b.

(c) Construct a function $g \in C_o(\mathbb{R})$ which is not the Fourier transform of an L^1 function. To do this note that it is enough to find a function $g \in C_o(\mathbb{R})$ which does not satisfy the condition in (b).

(ii) Show that $\overline{\mathcal{F}(L^1(\mathbb{R}^n))} = C_o(\mathbb{R}^n)$ where the closure is taken in the C_o topology.

Hint: Observe that $C_c^\infty(\mathbb{R}^n)$ is dense in $C_o(\mathbb{R}^n)$, in the topology of the supremum norm.

It is convenient to define the formal inverse of the Fourier transform in the following way. For $f \in L^1(\mathbb{R}^n)$ we set

$$\mathcal{F}^{-1}(f)(\xi) = \mathcal{F}^*(f)(\xi) = \check{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{2\pi i x\cdot\xi}d\xi = \hat{f}(-\xi) = \tilde{\hat{f}}(\xi) = \hat{\tilde{f}}(\xi).$$

Here we denote by $\tilde{g}$ the *reflection* of a function g, that is, $\tilde{g}(x) = g(-x)$. Observe that $\mathcal{F}^*$ is the conjugate of the Fourier transform. Thus the operator $\mathcal{F}^*$ is very closely connected to the operator $\mathcal{F}$ and enjoys essentially the same symmetries and properties.

As we shall see later on, it is also the adjoint of the Fourier transform with respect to the L^2 inner product

$$\langle f, g\rangle = \int_{\mathbb{R}^n} f\bar{g}.$$

Although we haven't yet defined the Fourier transform on L^2 we can calculate for $f, g \in L^1 \cap L^2(\mathbb{R}^n)$ that

$$\begin{aligned}
\int_{\mathbb{R}^n}(\mathcal{F} f)\bar{g} &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}dx\bar{g}(\xi)d\xi \\
&= \int_{\mathbb{R}^n} f(x)\overline{\int_{\mathbb{R}^n} g(\xi)e^{2\pi i x\cdot\xi}d\xi}\, dx \\
&= \int_{\mathbb{R}^n} f\overline{(\mathcal{F}^*(g))}.
\end{aligned}$$

Proposition 3.17 claims that $\mathcal{F}^*$ is also the inverse of the Fourier transform in the sense that

$$\mathcal{F}^*\mathcal{F} f = f,$$

whenever $f, \mathcal{F} f \in L^1(\mathbb{R}^n)$.

The proof of Proposition 3.17 is quite interesting in the following ways. First of all observe that we have actually showed that whenever $f \in L^1(\mathbb{R}^n)$, f is equal (a.e.) to the L^1 limit of the functions

$$\int_{\mathbb{R}^n} \hat{f}(\xi) e^{-a|\xi|^2} e^{2\pi i x\cdot\xi} d\xi,$$

as $a \to 0$. This does not require any additional hypothesis and actually provides us with a method of inverting the Fourier transform of *any* L^1 function, at least in the L^1 sense. The second remark is that the proof of Proposition 3.17 can be generalized to different *methods of summability*. Indeed, let $\Phi \in L^1(\mathbb{R}^n)$ be such that $\phi = \hat{\Phi} \in L^1(\mathbb{R}^n)$ and $\Phi(0) = \int \phi(x)dx = 1$. For $\epsilon > 0$ we consider the integrals

$$\int_{\mathbb{R}^n} \hat{f}(\xi) \Phi(\epsilon\xi) e^{2\pi i x\cdot\xi} d\xi, \tag{3.2}$$

which we will call the Φ-means of the integral $\int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x\cdot\xi}$, or just the Φ-means of $\check{f}$. Using the multiplication formula in Proposition 3.10 we can rewrite the means (3.2) as

$$\int_{\mathbb{R}^n} \hat{f}(\xi) \Phi(\epsilon\xi) e^{2\pi i x\cdot\xi} d\xi = (f * \tilde{\phi}_\epsilon)(x), \quad x \in \mathbb{R}^n. \tag{3.3}$$

The following more general version of Proposition 3.17 is true.

PROPOSITION 3.21. *Let $\Phi \in L^1(\mathbb{R}^n)$ be such that $\phi = \hat{\Phi} \in L^1(\mathbb{R}^n)$ with $\int \phi = 1$. We then have that the* Φ-means of $\int \hat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi$,

$$\int_{\mathbb{R}^n} \hat{f}(\xi) \Phi(\epsilon\xi) e^{2\pi i x\cdot\xi} d\xi,$$

converge to f in L^1, as $\epsilon \to 0$.

PROOF. The proof is just a consequence of formula (3.3). Indeed, $\tilde{\phi}_\epsilon$ is an approximation to the identity since $\tilde{\phi} \in L^1$ and $\int \tilde{\phi}(x)dx = 1$ and thus $f * \tilde{\phi}_\epsilon$ converges to f in the L^1 norm as $\epsilon \to 0$. □

Proposition 3.17 says that the inversion formula is true whenever $f, \hat{f} \in L^1(\mathbb{R}^n)$. This however is not the most natural assumption since the Fourier transform of an L^1 function need not be integrable. The idea behind Proposition 3.21 is to "force" $\hat{f}$ in L^1 by multiplying it by the L^1 function $\Phi(\epsilon\xi)$. Thus, we artificially impose some decay on $\hat{f}$. This is equivalent to smoothing out the function f itself by convolving it with a smooth function $\tilde{\phi}_\epsilon$. Although no smoothness is explicitly assumed in Proposition 3.21, there is a hidden smoothness hypothesis in the requirement $\Phi, \phi \in L^1$. Indeed, we could have replaced this assumption by directly assuming that ϕ is (say) a smooth function with compact support and taking $\Phi = \hat{\phi}$; then the conclusion $\hat{\phi} \in L^1(\mathbb{R}^n)$ would follow automatically. The trick of multiplying the Fourier transform of a general L^1 function with an appropriate function in L^1 or, equivalently, smoothing out the function f itself allows us then to invert the Fourier transform, at least in the L^1-sense. This process is usually referred to as a *summability method*.

As we shall see now, the inversion of a Fourier transform by means of a summability method is also valid in a pointwise sense. Because of formula (3.3), in order to understand the pointwise convergence of the Φ-means of $\check{\hat{f}}$ we have to examine the pointwise convergence of the convolution $f * \phi_\epsilon$ to f, whenever ϕ is an approximation to the identity.

DEFINITION 3.22. Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. The *Lebesgue set of f* is the set of points $x \in \mathbb{R}^n$ such that

$$\lim_{r\to 0} \frac{1}{r^n} \int_{|y|<r} |f(x-y) - f(x)| dy = 0.$$

The Lebesgue set of a locally integrable function f is closely related to the set where the integral of f is differentiable:

DEFINITION 3.23. Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. The *set of points where the integral of f is differentiable* is the set of points $x \in \mathbb{R}^n$ such that

$$\lim_{r\to 0} \frac{1}{\Omega_n r^n} \int_{|y|<r} f(x-y) = f(y),$$

where Ω_n is the volume of the unit ball $B(0,1)$ in $\mathbb{R}^n$. In other words, we say that the integral of f is differentiable at some point $x \in \mathbb{R}^n$ if the *average* of f with respect to Euclidean balls centered at x the value of f at the point x.

We shall come back to these notions a bit later in the course when we will introduce the maximal function of f which is just the maximal average of f around every point. For now we will use as a black box the following theorem:

THEOREM 3.24. *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Then the integral of f is differentiable at almost every point $x \in \mathbb{R}^n$.*

While postponing the proof of this theorem for later on, we can already see the following simple proposition connecting the Lebesgue set of f to the set of points where the integral of f is differentiable. In particular we see that almost every point in $\mathbb{R}^n$ is Lebesgue point of f.

COROLLARY 3.25. *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Then almost every $x \in \mathbb{R}^n$ is a Lebesgue point of f.*

PROOF. First assume that f is real valued. For any rational number q we have that the function $f(x) - q$ is locally integrable. Theorem 3.24 then implies that

$$\lim_{r\to 0} \frac{1}{r^n} \int_{|y|\leq r} \Big\{ |f(x-y) - q| - |f(x) - q| \Big\} dy = 0,$$

for almost every $x \in \mathbb{R}^n$. Thus the set F_q where the previous statement is not true has measure zero and so does the set $F := \cup_{q\in\mathbb{Q}} F_q$. Now let $x \in \mathbb{R}^n \setminus F$. Indeed, let $\epsilon > 0$ and $q \in \mathbb{Q}$ be such that $|f(x) - q| < \epsilon/2$. We then have

$$\frac{1}{\Omega_n r^n} \int_{|y|<r} |f(x-y) - f(x)| dy \leq \frac{1}{\Omega_n r^n} \int_{|y|<r} |f(x-y) - q| dy + \frac{1}{\Omega_n r^n} \int_{|y|<r} |f(x) - q| dy.$$

The first summand converges to $|f(x) - q| < \epsilon/2$ as $r \to 0$ since $x \notin F$ while the second summand is smaller than $\epsilon/2$. This shows that the Lebesgue set of f is contained in $\mathbb{R}^n \setminus F$ and thus that almost every point in $\mathbb{R}^n$ is a Lebesgue point of f.

For general complex valued f we just apply the result just proved to the real part and the imaginary part of f and use the triangle inequality. □

We can now give the following pointwise convergence result for approximations to the identity.

THEOREM 3.26. *Let $\phi \in L^1(\mathbb{R}^n)$ with $\int \phi = 1$. We define $\psi(x) := \operatorname{ess\,sup}_{|y|\geq|x|} |\phi(y)|$. If $\psi \in L^1(\mathbb{R}^n)$ and $f \in L^p(\mathbb{R}^n)$ for $1 \leq p \leq \infty$ then*

$$\lim_{\epsilon\to 0}(f * \phi_\epsilon)(x) = f(x),$$

whenever x is a Lebesgue point for f.

PROOF. Let x be a Lebesgue point of f and fix $\delta > 0$. By Corollary 3.25 there exists $\eta > 0$ such that

$$\frac{1}{r^n}\int_{|y|<r} |f(x-y) - f(x)|dy < \delta, \tag{3.4}$$

whenever $|r| < \eta$.

We can estimate as usual

$$\begin{aligned}
|(f*\phi_\epsilon)(x) - f(x)| &= \left|\int_{\mathbb{R}^n} [f(x-y) - f(x)]\phi_\epsilon(y)dy\right| \\
&\leq \left|\int_{|y|<\eta} [f(x-y) - f(x)]\phi_\epsilon(y)dy\right| + \left|\int_{|y|\geq\eta} [f(x-y) - f(x)]\phi_\epsilon(y)dy\right| \\
&=: I_1 + I_2.
\end{aligned}$$

We claim that

$$\lim_{|x|\to+\infty} |x|^n \psi(x) = 0. \tag{3.5}$$

First of all observe that ψ is radially decreasing. We will abuse notation and write $\psi(x) = \psi(|x|)$. For every $r > 0$ we have that

$$\int_{r/2\leq|x|<r} \psi(x)dx \geq \psi(r)(r^n - (r/2)^n)\Omega_n \simeq_n r^n\psi(r).$$

Since $\psi \in L^1$ the left hand side tends to 0 as $r \to +\infty$ which proves the claim.

We write (3.4) in polar coordinates to get

$$\frac{1}{r^n}\int_{S^{n-1}}\int_0^r |f(x-sy') - f(x)|s^{n-1}ds d\sigma_{n-1}(y') < \delta.$$

Setting $g(s) = \int_{S^{n-1}} |f(x-sy') - f(x)|d\sigma_{n-1}(y')$ we can rewrite the previous estimate in the form

$$G(r) := \int_0^r g(s)s^{n-1}ds \leq \delta r^n,$$

whenever $|r| < \eta$ and, by continuity, for $|r| = \eta$ as well. We now estimate I_1 as follows

$$
\begin{aligned}
I_1 &\leq \int_{S^{n-1}} \int_0^\eta |f(x-ry') - f(x)| \psi_\epsilon(r) d\sigma_{n-1}(y') r^{n-1} dr \\
&= \int_0^\eta g(r) r^{n-1} \frac{1}{\epsilon^n} \psi(r/\epsilon) dr = \int_0^\eta G'(r) \psi_\epsilon(r) dr.
\end{aligned}
$$

At this point the proof simplifies a bit if we assume that ψ is differentiable. In this case we have that $\psi' \leq 0$ and we can estimate the last integral by

$$
\begin{aligned}
\int_0^\eta G'(r)\psi_\epsilon(r)dr &= G(\eta)\psi_\epsilon(\eta) - \int_0^\eta G(r)\Big(\frac{d}{dr}\psi_\epsilon\Big)(r)dr \\
&\lesssim_{n,\phi} \delta\eta^n \psi_\epsilon(\eta) - \delta \int_0^\eta r^n \Big(\frac{d}{dr}\psi_\epsilon\Big)(r)dr \\
&= \delta n \int_0^\eta r^{n-1}\psi_\epsilon(r)dr = \frac{\delta}{\omega_n}\int_{\mathbb{R}^n} \psi(x)dx,
\end{aligned}
$$

where ω_n is the surface measure of the unit sphere S^{n-1} in $\mathbb{R}^n$.

REMARK 3.27. A monotone non-decreasing function F is almost everywhere differentiable and $\int_a^b F' \leq F(b) - F(a)$. By considering the function $-F$ we see that a monotone, non-increasing function G is almost everywhere differentiable and $\int_a^b G \geq G(b) - G(a)$.

For I_2 we estimate as follows

$$
I_2 \leq \|f\|_p \|\psi_\epsilon \mathbf{1}_{\{|x|\geq \eta\}}\|_{p'} + |f(x)| \|\mathbf{1}_{\{|x|\geq\eta\}}\psi_\epsilon\|_1.
$$

For the second summand we have that

$$
\|\mathbf{1}_{\{|x|\geq\eta\}}\psi_\epsilon\|_1 = \frac{1}{\epsilon^n}\int_{|x|\geq\eta} \psi_\epsilon(x/\epsilon)dx = \int_{|x|\geq\eta/\epsilon}\psi(x)dx \to 0,
$$

as $\epsilon \to 0$, since $\psi \in L^1$.

On the other hand, we have

$$
\begin{aligned}
\|\psi_\epsilon \mathbf{1}_{\{|x|\geq\eta\}}\|_{p'} &= \Big(\int_{|x|\geq\eta} [\psi_\epsilon(x)]^{p'} dx\Big)^{\frac{1}{p'}} = \Big(\int_{|x|\geq\eta/\epsilon} [\psi(x)]^{\frac{p'}{p}} \psi(x)dx\Big)^{\frac{1}{p'}} \\
&\leq \|\psi\mathbf{1}_{\{|x|\geq\eta/\epsilon\}}\|_\infty^{\frac{1}{p}} \|\psi\mathbf{1}_{\{|x|\geq\eta/\epsilon\}}\|_1.
\end{aligned}
$$

Now since ψ_ϵ is decreasing we have that

$$
\|\psi\mathbf{1}_{\{|x|\geq\eta/\epsilon\}}\|_\infty \leq \psi(\eta/\epsilon) = \eta^{-n}(\eta/\epsilon)^n\psi(\eta/\epsilon) \to 0
$$

as $\epsilon \to 0$ by (3.5). We get that the first summand also vanishes as $\epsilon \to 0$ and thus $I_2 \to 0$ as $\epsilon \to 0$.

We have showed that

$$
\limsup_{\epsilon\to 0} |(f * \phi_\epsilon)(x) - f(x)| \lesssim_{n,\phi} \delta,
$$

whenever x is a Lebesgue point of f. Since $\delta > 0$ was arbitrary this completes the proof of the theorem. ∎

REMARK 3.28. The previous theorem is true, for example, in the case that ϕ itself is a radially decreasing function in L^1 or if it is a function that satisfies a bound of the form $|\phi(x)| \lesssim_{n,\delta} (1+|x|)^{-(n+\delta)}$ for some $\delta > 0$. In particular it holds for our favorite functions $\phi(x) = e^{\pi|x|^2}$ and $\phi(x) = c_n(1+|x|^2)^{-\frac{n+1}{2}}$.

We conclude the discussion on the inversion of the Fourier transform with a useful corollary.

COROLLARY 3.29. *Let $f \in L^1(\mathbb{R}^n)$ and assume that f is continuous at 0 and that $\hat{f} \geq 0$. Then $\hat{f} \in L^1(\mathbb{R}^n)$ and*

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

for almost every $x \in \mathbb{R}^n$. In particular,

$$f(0) = \int_{\mathbb{R}^n} \hat{f}(\xi) d\xi.$$

PROOF. By identity (3.3) we have that

$$\int_{\mathbb{R}^n} \hat{f}(\xi)\Phi(\epsilon\xi) e^{2\pi i x\cdot\xi} d\xi = (f * \tilde{\phi}_\epsilon)(x),$$

for all $x \in \mathbb{R}^n$. Observe that the functions on both sides of this identity are continuous functions of x. Now let ϕ, Φ satisfy the conditions of Theorem 3.26. Assume furthermore that Φ is non-negative and continuous at 0. For example we can consider the function $\Phi(\xi) = \phi(\xi) = e^{-\pi|\xi|^2}$. Now since the point 0 is a point of continuity of f, it certainly belongs to the Lebesgue set of f. Thus we have that $\lim_{\epsilon\to 0}(f * \tilde{\phi}_\epsilon)(0) = f(0)$ which gives

$$\lim_{\epsilon\to 0} \int_{\mathbb{R}^n} \hat{f}(\xi)\Phi(\epsilon\xi) d\xi = f(0).$$

Since $\hat{f}\Phi$ is positive, we can use Fatou's lemma to write

$$\int_{\mathbb{R}^n} \hat{f}(\xi) d\xi = \int_{\mathbb{R}^n} \liminf_{\epsilon_k\to 0} \hat{f}(\xi)\Phi(\epsilon_k\xi) d\xi \leq f(0),$$

so $\hat{f} \in L^1(\mathbb{R}^n)$. Thus the inversion formula holds true for f and we get

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

for almost every $x \in \mathbb{R}^n$. However

$$f(0) = \lim_{\epsilon\to 0} \int_{\mathbb{R}^n} \hat{f}(\xi)\Phi(\epsilon\xi) d\xi = \int_{\mathbb{R}^n} \lim_{\epsilon\to 0} \hat{f}(\xi)\Phi(\epsilon\xi) d\xi = \int_{\mathbb{R}^n} \hat{f}(\xi) d\xi,$$

since $\hat{f} \in L^1$. □

3.1.2. Two special summability methods. We describe in detail two summability methods that are of special interest. These are based on the Examples 3.4 and 3.6 in the beginning of this set of notes and on applications of Theorem 3.26 and Proposition 3.21.

3.1.2.1. *The Gauss-Weierstrass summability method.* By dilating the function $W(x) = e^{-\pi|x|^2}$ we get

$$W(x,t) := W_{\sqrt{4\pi t}}(x) = (4\pi t)^{-\frac{n}{2}} e^{-\frac{|x|^2}{4t}}.$$

The function $W(x,t)$, $t > 0$, is called the *Gauss kernel* and it gives rise to the *Gauss-Weierstrass method of summability*. The Fourier transform of W is

$$\widehat{W_{\sqrt{4\pi t}}}(\xi) = \widehat{W}(\sqrt{2\pi t}\xi) = e^{-4\pi^2 t|\xi|^2}.$$

It is also clear that

$$\int_{\mathbb{R}^n} W(x,t)dx = 1,$$

for all $t > 0$. The discussion in the previous sections applies to the Gauss-Weierstrass summability method and we have that the means

$$w(x,t) := \int_{\mathbb{R}^n} f(y)W(y-x,t)dy = \int_{\mathbb{R}^n} \hat{f}(\xi)e^{-4\pi^2 t|\xi|^2} e^{2\pi i x\cdot\xi} d\xi$$

converge to f in $L^1(\mathbb{R}^n)$ as $t \to 0$, and also in the pointwise sense, for every x in the Lebesgue set of f. One of the aspects of Gauss-Weierstrass summability is that the function $w(x,t)$ defined above satisfies the *heat equation*:

$$\begin{aligned} \frac{\partial w}{\partial t} - \Delta_x w = 0, &\quad \text{on } \mathbb{R}^{n+1}_+, \\ w(x,0) = f(x), &\quad x \in \mathbb{R}^n. \end{aligned}$$

Here $\Delta_x := \sum_{j=1}^n \partial^2_{x_j}$ denotes the Laplacian in the space variable x only. To see that the Gauss-Weierstrass means of $\check{f}$ satisfy the Heat equation with initial data f, one can use the formula for $w(x,t)$ and calculate everything explicitly. However it is easier to consider the Fourier transform of the solution $u(x,t)$ of the Heat equation in the x variable and show that it must agree with the Fourier transform of $w(x,t)$, again in the x variable. Observe that under suitable assumptions on the initial data f we get that the solution $w(x,t)$ converges to the initial data f as "time" $t \to 0$.

EXERCISE 3.30. Let $f(x) = e^{-\pi x^2}$, $x \in \mathbb{R}$. Using the properties of the Fourier transform show that the function $\hat{f}$ satisfies the initial value problem

$$\begin{aligned} &u' + 2\pi x u = 0, \\ &u(0) = 1. \end{aligned}$$

Solve the initial value problem to give an alternative proof of the fact that $\hat{f}(\xi) = e^{-\pi\xi^2}$. Observe that the differential equation above is invariant under the Fourier transform.

The Abel summability method. We consider the function $P(x) = c_n \frac{1}{(1+|x|^2)^{\frac{n+1}{2}}}$ where $c_n = \frac{\Gamma((n+1)/2}{\pi^{\frac{n+1}{2}}}$. By dilating the function P we have

$$P(x,t) := P_t(x) = c_n \frac{t}{(t^2+|x|^2)^{\frac{n+1}{2}}}.$$

The function $P(x,t),\ t>0$, is called the *Poisson kernel* (for the upper half plane) and it gives rise to the *Abel method of summability*. The Fourier transform of P_t is

$$\widehat{P_t}(\xi) = \hat{P}(t\xi) = e^{-2\pi t|\xi|}.$$

This is just a consequence of the calculation in Example 3.6, the inversion formula and the easily verified fact that $P \in L^1(\mathbb{R}^n)$. It is also clear by a direct calculation or through the previous Fourier transform relation that

$$\int_{\mathbb{R}^n} P(x,t)dx = 1,$$

for all $t>0$. Everything we have discussed in these notes applies to the Abel summability method. In particular we have that whenever $f \in L^1(\mathbb{R}^n)$, the means

$$u(x,t) := \int_{\mathbb{R}^n} f(y)P(y-x,t)dy = \int_{\mathbb{R}^n} \hat{f}(\xi)e^{-2\pi t|\xi|}e^{-2\pi i x\cdot\xi}d\xi,$$

converge to f in L^1 as $t \to 0$ and also in the pointwise sense for all x in the Lebesgue set of f. The function $u(x,t)$ is also called *the Poisson integral or extension of* f. It is not difficult to see that it satisfies the *Dirichlet* problem

$$\Delta u = 0, \quad \text{on } \mathbb{R}^{n+1}_+,$$
$$u(x,0) = f(x), \quad x \in \mathbb{R}^n.$$

Here we denote by $\mathbb{R}^{n+1}_+$ the *upper half plane* $\mathbb{R}^{n+1}_+ = \{(x,y) : x \in \mathbb{R}^n, y > 0\}$. Thus, if we are given an L^1 function on the "boundary" $\mathbb{R}^n$, the Poisson integral of f provides us with a harmonic function $u(x,t)$ in the upper half plane which has boundary value f in the sense that $u(x,t)$ converges to f as $t \to 0$ both in the L^1 sense as well as almost everywhere.

REMARK 3.31. It is not hard to see that *the Poisson extension* of $f \in L^p(\mathbb{R}^n)$, $1 \le p \le \infty$

$$u(x,t) = \int_{\mathbb{R}^n} f(y)P(x-y,t)dy,$$

is *harmonic* in $\mathbb{R}^{n+1}_+$, that is, that it satisfies the Laplace equation:

$$\Delta_{x,t}u(x,t) = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_k^2}u(x,t) + \frac{\partial^2}{\partial t^2}u(x,t) = 0.$$

This is a consequence of the fact that $\Delta_{x,t}P(x,t) = 0$ for $(x,t) \in \mathbb{R}^{n+1}_+$.

In general, if $f \in L^p(\mathbb{R}^n)$, $1 < p < +\infty$, then the Poisson extension of f, $u(x,t)$, is harmonic in $\mathbb{R}^{n+1}_+$ and $\lim_{t\to 0} u(\cdot,t) = f \in L^p(\mathbb{R}^n)$ where the limit is taken L^p-sense. The function $u(x,t)$ also converges to $f(x)$ for almost every $x \in \mathbb{R}^n$, as $t \to 0^+$. The same is true if $p = \infty$ and $f \in C_o(\mathbb{R}^n) \subset L^\infty(\mathbb{R}^n)$, in which case $u(\cdot,t)$ converges to f uniformly as $t \to 0^+$.

If one considers the Dirichlet problem in the upper half plane, a natural question is that of uniqueness. That is, given a boundary value f on $\mathbb{R}^n$ we want to find a function u which is harmonic in $\mathbb{R}^{n+1}_+$, continuous up to the boundary, and equals f on $\mathbb{R}^n$. If we further ask that u is bounded in $\mathbb{R}^{n+1}_+$, then the solution is unique (and given by the Poisson extension of f). However, unless some additional assumption is place u (such as boundedness), the previous result fails. For example, both the functions $u_1(x,t) = t$ and $u_2(x,t) = 0$ are harmonic in $\mathbb{R}^{n+1}_+$ and continuous through the boundary, where they both vanish. See [SW] for more details on Poisson extensions and properties of harmonic functions.

3.2. The Fourier transform of the Schwartz class

In this section we go back to the space of Schwartz functions $\mathcal{S}(\mathbb{R}^n)$ and we define the Fourier transform on this space. This will turn out to be extremely useful and flexible. The reason for this is the fact that Schwartz functions are much "nicer" than functions that are just integrable. On the other hand, Schwartz functions are dense in all L^p spaces, $p < \infty$, so many statements established initially for Schwartz functions go through in the more general setup of L^p spaces. A third reason is the dual of the space $\mathcal{S}(\mathbb{R}^n)$, the space of tempered distributions, is rich enough to allow us to define the Fourier transform of much rougher objects than integrable functions

3.2.1. The space of Schwartz functions as a Fréchet space. We recall that the space of Schwartz functions $\mathcal{S}(\mathbb{R}^n)$ consists of all smooth (i.e. infinitely differentiable) functions $f : \mathbb{R}^n \to \mathbb{C}$ such that the function itself together with all its derivatives decay faster than any polynomial at infinity. To make this more precise it is useful to introduce the *seminorms* p_N defined for any non-negative integer N as

$$p_N(f) := \sup_{|\alpha|\leq N, |\beta|\leq N} \sup_{x\in\mathbb{R}^n} |x^\alpha \partial^\beta f(x)|,$$

where $\alpha, \beta \in \mathbb{N}^n_o$ are multi-indices and as usual we write $|\alpha| = \alpha_1 + \cdots + \alpha_n$. Thus $f \in \mathcal{S}(\mathbb{R}^n)$ if and only if $f \in C^\infty(\mathbb{R}^n)$ and $p_N(f) < +\infty$ for $N \in \mathbb{N}_o$.

It is clear that $\mathcal{S}(\mathbb{R}^n)$ is a vector space. We have already seen that a basic example of a function in $\mathcal{S}(\mathbb{R}^n)$ is the Gaussian $f(x) = e^{-\pi|x|^2}$ and it is not hard to check that the more general Gaussian function $f(x) = e^{-\langle Ax, x\rangle}$, where A is a positive definite real matrix, is also in $\mathcal{S}(\mathbb{R}^n)$. Furthermore, the product of two Schwartz functions is again a Schwartz function and the space $\mathcal{S}(\mathbb{R}^n)$ is closed under taking partial derivatives or multiplying by complex valued polynomials of any degree. As we have already seen (and it's obvious by the definitions) the space of infinitely differentiable functions with compact support is contained in $\mathcal{S}(\mathbb{R}^n)$, $\mathcal{D}(\mathbb{R}^n) = C^\infty_c(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$, and each one of these spaces is a dense subspace of $L^p(\mathbb{R}^n)$ for any $1 \leq p < \infty$ and also in $C_o(\mathbb{R}^n)$, in the corresponding topologies.

The seminorms defined above define a topology in $\mathcal{S}(\mathbb{R}^n)$. In order to study this topology we need the following definition:

DEFINITION 3.32. A Fréchet space is a locally convex topological vector space which is induced by a complete invariant metric.

A translation invariant metric on $\mathcal{S}(\mathbb{R}^n)$. It is not hard to actually define a metric on $\mathcal{S}(\mathbb{R}^n)$ which induces the topology. Indeed for two functions $f, g \in \mathcal{S}(\mathbb{R}^n)$ we set

$$\rho(f,g) = \sum_{N=0}^{\infty} \frac{1}{2^N} \frac{p_N(f-g)}{1 + p_N(f-g)}.$$

The function $\rho : \mathcal{S}(\mathbb{R}^n) \times \mathcal{S}(\mathbb{R}^n) \to [0, +\infty)$ is translation invariant, symmetric and it separates the elements of $\mathcal{S}(\mathbb{R}^n)$. The metric ρ induces a topology in $\mathcal{S}(\mathbb{R}^n)$; a set $U \subset \mathcal{S}(\mathbb{R}^n)$ is open if and only if there exists exists $f \in U$ and $\epsilon > 0$ such that

$$B_\rho(f, \epsilon) := \{g \in \mathcal{S}(\mathbb{R}^n) : \rho(f, g) < \epsilon\} \subset U.$$

Convergence in $\mathcal{S}(\mathbb{R}^n)$. By definition, a sequence $\{\phi_k\}_{k\in\mathbb{N}} \subset \mathcal{S}(\mathbb{R}^n)$ converges to 0 if $\rho(\phi_k, 0) \to 0$ as $k \to \infty$. A more handy description of converging sequences in $\mathcal{S}(\mathbb{R}^n)$ is given by the following lemma.

LEMMA 3.33. *A sequence $\{\phi_k\}_{k\in\mathbb{N}} \subset \mathcal{S}(\mathbb{R}^n)$ converges to 0 if and only if*

$$p_N(\phi_k) \to 0 \quad \text{as} \quad k \to \infty,$$

for all $N \in \mathbb{N}_o$.

PROOF. First assume that $\rho(\phi_k, 0) \to 0$ as $k \to \infty$. Then, since

$$\sum_{N=1}^{\infty} \frac{1}{2^N} \frac{p_N(\phi_k)}{1 + p_N(\phi_k)}$$

converges to zero as $k \to \infty$ and all summands are positive, we conclude that for every N we have that

$$\frac{p_N(\phi_k)}{1 + p_N(\phi_k)} \to 0,$$

as $k \to \infty$. However, this easily implies that $p_N(\phi_k) \to 0$ as $k \to \infty$, for every $N \in \mathbb{N}_o$.

Assume now that $p_N(\phi_k) \to 0$ as $k \to \infty$ for every $N \in \mathbb{N}_o$ and let $\epsilon > 0$. We choose a positive integer M such that $2^{-M} < \frac{\epsilon}{2}$.

Thus,

$$\begin{aligned}
\rho(\phi_k, 0) &= \sum_{N=1}^{M} \frac{1}{2^N} \frac{p_N(\phi_k)}{1 + p_N(\phi_k)} + \sum_{N=M+1}^{\infty} \frac{1}{2^N} \frac{p_N(\phi_k)}{1 + p_N(\phi_k)} \\
&\leq \sum_{N=1}^{M} \frac{1}{2^N} \frac{p_N(\phi_k)}{1 + p_N(\phi_k)} + \frac{\epsilon}{2}.
\end{aligned}$$

Now, every term in the finite sum of the first summand converges to 0 as $k \to \infty$ and we get that $\rho(\phi_k, 0) \to 0$ as $k \to \infty$. □

$\mathcal{S}(\mathbb{R}^n)$ is a topological vector space. The topology induced by ρ turns $\mathcal{S}(\mathbb{R}^n)$ into a topological vector space. To see this we need to check that addition of elements in $\mathcal{S}(\mathbb{R}^n)$ and multiplication by complex constants are continuous with respect to ρ. This is very easy to check and is left as an exercise.

Local convexity. For $\epsilon > 0$ and $N \in \mathbb{N}_o$ consider the family of sets

$$U_{\epsilon,N} := \{f \in \mathcal{S}(\mathbb{R}^n) : p_N(f) < \epsilon\}.$$

We claim that $\{U_{\epsilon,N}\}_{\epsilon>0, N\in\mathbb{N}_o}$ is a neighborhood basis of the point 0 for the topology induced by ρ. Indeed, the system $B_\rho(0, \epsilon)$ defines a neighborhood basis of 0. On the other hand it is implicit in the proof of Lemma 3.33 that for every $\epsilon > 0$ there is some $\epsilon' > 0$ and some $N > 0$ such that $U_{\epsilon',N} \subset B_\rho(0, \epsilon)$. This proves the claim.

Now, in order to show that $\mathcal{S}(\mathbb{R}^n)$ endowed with the topology induced by ρ is locally convex it suffices (by translation invariance) to show that the point

0 has a neighborhood basis which consists of convex sets. This is clear for the neighborhood basis $U_{\epsilon,N}$ defined above since the seminorms p_N are positive homogeneous. Observe however that the balls $B_\rho(0,\epsilon)$ are not convex.

EXERCISE 3.34. Show that the balls $B_\rho(0,\epsilon)$, $\epsilon > 0$, are *not* convex sets.

Completeness. The space $\mathcal{S}(\mathbb{R}^n)$ is a complete topological vector space with the topology induced by ρ. If $\{\phi_k\}_k$ is a Cauchy sequence in $\mathcal{S}(\mathbb{R}^n)$ then for every $\alpha,\beta \in \mathbb{N}_o^n$, the sequence

$$x^\alpha \partial^\beta \phi_k$$

is a Cauchy sequence in the space $C_o(\mathbb{R}^n)$, with the topology induced by the supremum norm. Since this space is complete we conclude that ϕ_k converges uniformly to some $\phi \in C_o(\mathbb{R}^n)$. A standard uniform convergence argument shows now that $\phi \in \mathcal{S}(\mathbb{R}^n)$.

REMARK 3.35. In general, a sequence $\{\phi_k\}_k$ in a topological vector space is called a *Cauchy sequence* if for every open neighborhood of zero U, there exists some positive integer N so that $\phi_k - \phi_{k'} \in U$ for all $k,k' > N$. If the topology is induced by a translation invariant metric, this definition coincides with the more familiar one, that is: for every $\epsilon > 0$ there exists $N > 0$ such that $\rho(\phi_k,\phi_{k'}) < \epsilon$ whenever $k,k' > N$.

The discussion above gives the following:

THEOREM 3.36. *The space $\mathcal{S}(\mathbb{R}^n)$, endowed with the metric ρ and the topology induced by ρ, is a Fréchet space.*

We now give a general lemma that describes continuity of linear operators acting on $\mathcal{S}(\mathbb{R}^n)$.

LEMMA 3.37. *Let $(X, \|\cdot\|_X)$ be a Banach space.*

(i) *A linear operator $T : \mathcal{S}(\mathbb{R}^n) \to X$ is continuous if and only if there exists $N \geq 0$ and $C > 0$ such that*

$$\|T(\phi)\|_X \leq C p_N(\phi), \tag{3.6}$$

for all $\phi \in \mathcal{S}(\mathbb{R}^n)$.

(ii) *Let $T : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ be a linear operator. Then T is continuous if and only if for each $N > 0$ there exists $N' > 0$ and $C > 0$ such that*

$$p_N(T(\phi)) \leq C p_{N'}(\phi), \tag{3.7}$$

for all $\phi \in \mathcal{S}(\mathbb{R}^n)$.

PROOF. For (i) it is clear that T is continuous if (3.6) holds. On the other hand, suppose that $T : \mathcal{S}(\mathbb{R}^n) \to X$ is continuous and let $B_X(0,1)$ be the open ball of center 0 and radius 1 in X. Then $T^{-1}(B_X(0,1))$ is an open neighborhood of 0 in $\mathcal{S}(\mathbb{R}^n)$ and hence it contains some $U_{\epsilon,N}$. Thus $p_N(\phi) < \epsilon$ implies that $\|T(\phi)\|_X < 1$. Now we have that

$$\|T(\phi)\|_X = \frac{2}{\epsilon}|p_N(\phi)| \left\| T\Big(\frac{\epsilon}{2p_N(\phi)}\phi\Big) \right\| \lesssim p_N(\phi).$$

Similarly, if $T : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is continuous then for every N, ϵ there is N', ϵ' so that

$$T^{-1}(U_{N,\epsilon}) \supset U_{N',\epsilon'}.$$

This implies (3.7) using the same trick we used to deduce (3.6). $\square$

It is obvious that for every $0 < p \leq \infty$, $\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$. Let us show however that this embedding is also continuous:

PROPOSITION 3.38. *Let* $0 < p \leq \infty$. *Then the identity map* $\mathrm{Id} : \mathcal{S}(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ *is continuous, that is, there exists* N *so that*

$$\|f\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} p_N(f),$$

for all $f \in \mathcal{S}(\mathbb{R}^n)$.

PROOF. Let $f \in \mathcal{S}(\mathbb{R}^n)$. For $p < \infty$ and $N > n/p$ we have that

$$\begin{aligned}
\|f\|_{L^p(\mathbb{R}^n)} &\leq \Big(\int_{|x|\leq 1} |f(x)|^p dx\Big)^{\frac{1}{p}} + \Big(\int_{|x|>1} |f(x)|^p dx\Big)^{\frac{1}{p}} \\
&\leq \|f\|_{L^\infty} |B(0,1)|^{\frac{1}{p}} + \sup_{x\in\mathbb{R}^n} |x|^N |f(x)| \Big(\int_{|x|>1} |x|^{-Np} dx\Big)^{\frac{1}{p}} \\
&\lesssim_{n,p} p_N(f).
\end{aligned}$$

If $p = \infty$ observe that $\|f\|_\infty = p_0(f)$ so there is nothing to prove. □

3.2.2. The Fourier transform on the Schwartz class. Since $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ there is no difficulty in defining the Fourier transform on $\mathcal{S}(\mathbb{R}^n)$ by means of the formula

$$\mathcal{F}(f)(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i x\cdot\xi} dx, \quad f \in \mathcal{S}(\mathbb{R}^n),\ \xi \in \mathbb{R}^n.$$

All the properties of $\mathcal{F}$ that we have seen in the previous sections are of course valid for the Fourier transform on $\mathcal{S}(\mathbb{R}^n)$. As we shall now see, there is much more we can say for the Fourier transform on $\mathcal{S}(\mathbb{R}^n)$.

For $f \in \mathcal{S}(\mathbb{R}^n)$ and every polynomial P we have that $P(-2\pi i x)f, P(\partial^\alpha)f \in \mathcal{S}(\mathbb{R}^n)$. Using the commutation relations

$$\begin{aligned}
\mathcal{F}(P(-2\pi i x)f)(\xi) &= P(\partial_\xi^\alpha)\hat{f}(\xi), \\
\mathcal{F}(P(\partial_x^\alpha)f)(\xi) &= P(2\pi i\xi)\hat{f}(\xi),
\end{aligned}$$

we see that $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$. Furthermore, since $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ we can use the inversion formula to write

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi = \mathcal{F}^{-1}(\hat{f}) = \mathcal{F}^{-1}\mathcal{F} f.$$

This shows that $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is onto and of course it is a one to one operator as we have already seen. Finally let us see that it is also a continuous map. To see this observe that

$$\begin{aligned}
p_N(\hat{f}) &= \sup_{|\alpha|,|\beta|\leq N} \|\xi^\alpha \partial^\beta \hat{f}\|_{L^\infty(\mathbb{R}^n)} = \sup_{|\alpha|,|\beta|\leq N} |2\pi|^{-|\alpha|} \|\mathcal{F}(\partial_x^\alpha (-2\pi i x)^\beta f)\|_{L^\infty(\mathbb{R}^n)} \\
&\leq \sup_{|\alpha|,|\beta|\leq N} |2\pi|^{|\beta|-|\alpha|} \|\partial_x^\alpha x^\beta f\|_{L^1(\mathbb{R}^n)} \lesssim_N \sup_{|\alpha|,|\beta|\leq N} \|x^\beta \partial_x^\alpha f\|_{L^1(\mathbb{R}^n)} \\
&\leq \sup_{|\alpha|,|\beta|\leq N} p_M(x^\beta \partial^\alpha f),
\end{aligned}$$

for every $M > n$, by Proposition 3.38. However, $\sup_{|\alpha|,|\beta|\leq N} p_M(x^\beta \partial^\alpha f) \leq p_{M+N}(f)$ so we get that

$$p_N(\hat{f}) \lesssim_N p_{M+N}(f),$$

for every $M > N$ which shows that $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is continuous.

We have thus proved the following:

THEOREM 3.39. *The Fourier transform is a* homeomorphism *of* $\mathcal{S}(\mathbb{R}^n)$ *onto itself. The operator*

$$\mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n), \quad g \mapsto \mathcal{F}^{-1}(g)(x) = \int_{\mathbb{R}^n} f(\xi) e^{2\pi i x\cdot\xi} d\xi,$$

is the continuous inverse of $\mathcal{F}$ *on* $\mathcal{S}(\mathbb{R}^n)$*:*

$$\mathcal{F}\mathcal{F}^{-1} = \mathcal{F}^{-1}\mathcal{F} = \mathrm{Id},$$

on $\mathcal{S}(\mathbb{R}^n)$.

We immediately get Plancherel's identities:

COROLLARY 3.40. *Let* $f, g \in \mathcal{S}(\mathbb{R}^n)$*. We have that*

$$\int_{\mathbb{R}^n} f(x)\overline{g(x)} dx = \int_{\mathbb{R}^n} \hat{f}(\xi)\overline{\hat{g}(\xi)} d\xi.$$

In particular, for every $f \in \mathcal{S}(\mathbb{R}^n)$ *we have that*

$$\|\hat{f}\|_{L^2(\mathbb{R}^n)} = \|f\|_{L^2(\mathbb{R}^n)}.$$

PROOF. Using the multiplication formula for the Fourier transform we can write

$$\int f\hat{g} = \int \hat{f} g,$$

for $f, g \in L^1(\mathbb{R}^n)$ and thus for $f, g \in \mathcal{S}(\mathbb{R}^n)$. Now let $f, g \in \mathcal{S}(\mathbb{R}^n)$ and apply this formula to the functions $f, h \in \mathcal{S}(\mathbb{R}^n)$ where $h = \bar{\hat{g}}$. Observing that $\hat{\bar{\hat{g}}} = \bar{g}$ we get the first of the identities in the corollary. Applying this identity to the functions f and $g = f$ we also get the second. □

We also get a nice proof of the fact that convolution of Schwartz functions is again a Schwartz function.

COROLLARY 3.41. *Let* $f, g \in \mathcal{S}(\mathbb{R}^n)$*. Then* $f * g \in S$.

PROOF. For $f, g \in \mathcal{S}(\mathbb{R}^n)$ we have that $\widehat{f * g} = \hat{f}\hat{g}$. Since $\hat{f}, \hat{g} \in S$ we conclude that $\widehat{f * g} \in \mathcal{S}(\mathbb{R}^n)$ and thus that $f * g \in \mathcal{S}(\mathbb{R}^n)$. □

3.3. The Fourier transform on $L^2(\mathbb{R}^n)$

We have already seen that the Fourier transform is defined for functions $f \in L^1(\mathbb{R}^n)$ by means of the formula

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

While this integral converges absolutely for $f \in L^1(\mathbb{R}^n)$, this is not the case in general for $f \in L^2(\mathbb{R}^n)$. However, Corollary 3.40 says that the Fourier transform is a bounded linear operator on $\mathcal{S}(\mathbb{R}^n)$ which is a dense subset of $L^2(\mathbb{R}^n)$ and in fact we have that

$$\|f\|_{L^2(\mathbb{R}^n)} = \|\hat{f}\|_{L^2(\mathbb{R}^n)} \tag{3.8}$$

for every $f \in \mathcal{S}(\mathbb{R}^n)$. As we have seen several times already, this means that the Fourier transform has a unique bounded extension, which we will still denote by $\mathcal{F}$, throughout $L^2(\mathbb{R}^n)$. In fact the Fourier transform $\mathcal{F}$ is an *isometry* on $L^2(\mathbb{R}^n)$ as identity (3.8) shows.

DEFINITION 3.42. A linear operator $S : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ which is an isometry and maps onto $L^2(\mathbb{R}^n)$ is called a *unitary operator*.

COROLLARY 3.43. *The Fourier transform is a unitary operator on $L^2(\mathbb{R}^n)$.*

The definition of the Fourier transform on L^2 given above suggests that given $f \in L^2(\mathbb{R}^n)$, one should find a sequence $\{h_k\}_k \subset \mathcal{S}(\mathbb{R}^n)$ such that $h_k \to f$ in L^2 and define

$$(\mathcal{F} f)S(\xi) = \hat{f}(\xi) = L^2 - \lim_{k\to\infty} \int_{\mathbb{R}^n} h_k(x) e^{-2\pi i x\cdot\xi} dx.$$

This, however, is a bit too abstract. The following lemma gives us an alternative way to calculate the Fourier transform on $L^2(\mathbb{R}^n)$.

LEMMA 3.44. *Let $f \in L^2(\mathbb{R}^n)$. The following formulas are valid*

$$\hat{f}(\xi) = L^2 - \lim_{R\to+\infty} \int_{|x|\le R} f(x) e^{-2\pi i x\cdot\xi} dx,$$
$$f(x) = L^2 - \lim_{R\to+\infty} \int_{|\xi|\le R} \hat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi,$$

where the notation above means that the limits are considered in the L^2 norm.

PROOF. Given $f \in L^2(\mathbb{R}^n)$ let us define the functions

$$f_R(x) = \begin{cases} f(x), & \text{if } |x| \le R, \\ 0, & \text{if } |x| > R. \end{cases}$$

Then on the one hand we have that $\lim_{R\to+\infty} f_R = f$ in $L^2(\mathbb{R}^n)$. On the other hand the functions f_R belong to $L^1(\mathbb{R}^n)$ for all $R > 0$ so we can write

$$\widehat{f_R}(\xi) = \int_{|x|\le R} f(x) e^{-2\pi i x\cdot\xi} dx, \quad \xi \in \mathbb{R}^n.$$

Since the Fourier transform is an isometry on $L^2(\mathbb{R}^n)$ we also have that $\widehat{f_R} \to \hat{f}$ as $R \to +\infty$ in $L^2(\mathbb{R}^n)$. The proof of the second formula is similar. □

3.4. The Fourier transform on $L^p(\mathbb{R}^n)$ and Hausdorff-Young

Having defined the Fourier transform on $L^1(\mathbb{R}^n)$ and on $L^2(\mathbb{R}^n)$ we can now interpolate between these two spaces. Indeed, we have established that

$$\mathcal{F} : L^1(\mathbb{R}^n) + L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n) + L^\infty(\mathbb{R}^n),$$

and that $\mathcal{F}$ is of strong type $(1,\infty)$ and of strong type $(2,2)$, with norm 1 in both cases. We have also seen that it is well defined on the simple functions with finite measure support and on the Schwartz class, both dense subsets of all L^p spaces for $p < \infty$. Setting $\frac{1}{p} = \frac{1-\theta}{2} + \frac{\theta}{2}$ we get $\theta = \frac{2}{p'}$ where p' is the dual exponent of p. This shows that $\frac{1}{q} = \frac{1-\theta}{2} + \frac{\theta}{\infty} = \frac{1}{p'}$. The Riesz-Thorin interpolation theorem now applies to show the following:

THEOREM 3.45 (Hausdorff-Young Theorem). *For* $1 \leq p \leq 2$ *the Fourier transform extends to a bounded linear operator*

$$\mathcal{F} : L^p(\mathbb{R}^n) \to L^{p'}(\mathbb{R}^n),$$

of norm at most 1*, that is we have*

$$\|\mathcal{F} f\|_{L^{p'}(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)}, \quad f \in L^p(\mathbb{R}^n), \quad 1 \leq p \leq 2.$$

REMARK 3.46. This is one instance where the Riesz-Thorin interpolation theorem fails to give the sharp norm, although the endpoint norms are sharp. Indeed, the actual norm of the Fourier transform is

$$\|\mathcal{F}\|_{L^p \to L^{p'}} = \frac{p^{\frac{1}{2p}}}{p'^{\frac{1}{2p'}}} < 1, \quad 1 \leq p \leq 2.$$

This is a deep theorem that has been proved firstly by K.I. Babenko in the special case that p is an even integer and then by W. Beckner in the general case.

EXERCISE 3.47. Let f be a general Gaussian function of the form

$$f(x) = ce^{2\pi i x\cdot\xi_o} e^{-\langle A(x-x_o), x-x_o\rangle},$$

for some positive definite real matrix $A : \mathbb{R}^n \to \mathbb{R}^n$. Show that

$$\|\mathcal{F} f\|_{L^{p'}(\mathbb{R}^n)} = \frac{p^{\frac{1}{2p}}}{p'^{\frac{1}{2p'}}} \|f\|_{L^p(\mathbb{R}^n)}.$$

Observe that this gives a lower bound on the norm $\|\mathcal{F}\|_{L^p \to L^{p'}}$.

Hint: Write f as a composition of translations, modulations and generalized dilations of the basic Gaussian function $e^{-\pi|x|^2}$.

REMARK 3.48. The inversion problem for L^p, $1 < p < 2$ has a similar solution as the L^1 case. One can easily see that the Φ means of $\check{f}$ converge to f in L^p as well as for every Lebesgue point of f if Φ is appropriately chose. In particular this is the case for the Abel or Gauss means of $\check{f}$.

We also have the following extension on the action of the Fourier transform on convolutions.

PROPOSITION 3.49. *Let* $f \in L^1(\mathbb{R}^n)$ *and* $g \in L^p(\mathbb{R}^n)$ *for some* $1 \leq p \leq 2$*. Then, as we know, the function* $f * g$ *belongs to* $L^p(\mathbb{R}^n)$*. We have that*

$$(\widehat{f * g})(x) = \hat{f}(x)\hat{g}(x),$$

for almost every $x \in \mathbb{R}^n$*.*

We close this section by discussing the possibility of other mapping properties of the Fourier transform, besides the ones given by the Hausdorff-Young theorem. In particular we have seen that the Fourier transform is of strong type (p, p') for all $1 \leq p \leq 2$. But are there any other pairs (p, q) for which the Fourier transform is of strong, or even weak type (p, q)?

The easiest thing to see is that whenever $\mathcal{F}$ is of type (p, q) we must have that $q = p'$.

EXERCISE 3.50. Suppose that $\mathcal{F}$ is of weak type (p,q). Show that we must necessarily have $q = p'$.

Hint: Exploit the scale invariance of the Fourier transform; in particular remember the symmetry $\mathcal{F}\mathrm{Dil}^p_\lambda = \mathrm{Dil}^{p'}_{\lambda^{-1}}\mathcal{F}$.

The previous exercise thus shows that the only possible type for $\mathcal{F}$ is of the form (p,p'). The Hausdorff-Young theorem shows that this is actually true whenever $1 \le p \le 2$. It turns out however that the bound (p,p') fails for $p > 2$. The following exercise describes one way to prove this.

EXERCISE 3.51. Show that $\mathcal{F}$ is not of strong type (p,p') when $p > 2$.

(i) Let N be a large positive integer and $g(x) = e^{-\pi|x|^2}$. For $y \in \mathbb{R}^n$ consider the function

$$f(x) = \sum_{j=1}^N e^{2\pi i x\cdot jy} g(x - jy), \quad x \in \mathbb{R}^n.$$

Show that

$$\hat{f}(\xi) = \sum_{j=1}^N e^{-2\pi i \xi\cdot jy}\hat{g}(\xi - jy).$$

(ii) For any $1 \le p \le \infty$ show that

$$\|f\|_{L^p(\mathbb{R}^n)} \simeq_p N^{\frac{1}{p}},$$

if N and $|y|$ are large enough. For this show first the endpoint bounds for $p = 1$ and $p = \infty$. This will also give you the intermediate upper bounds by log-convexity. For the lower bounds, consider the values of f close to integer multiples of y.

(iii) The previous steps show that

$$\|\hat{f}\|_{L^{p'}(\mathbb{R}^n)} \simeq_p N^{\frac{1}{p'}-\frac{1}{p}}\|f\|_{L^p(\mathbb{R}^n)},$$

which allows you to conclude the proof.

3.5. The space of tempered distributions $\mathcal{S}'(\mathbb{R}^n)$

The purpose of this paragraph is to introduce a space of "generalized functions" that is much larger than all the spaces we have seen so far, namely *the space of tempered distributions*. Let us begin with an informal discussion, drawing some analogies with some more classical function spaces.

We have seen already that whenever $1 \le p < \infty$ and the underlying measure is σ-finite, then the space $L^{p'}(\mathbb{R}^n)$ can be identified with the dual $(L^p(\mathbb{R}^n))^*$, by means of the pairing:

$$g \in L^{p'} \mapsto g^* : L^p \to \mathbb{C}, \quad g^*(f) = \int_{\mathbb{R}^n} f(x)\overline{g(x)}dx.$$

This is already quite interesting. A function in L^p is already a generalized object in the sense that it is only defined up to sets of measure zero; so, in fact, it represents an equivalence class. Furthermore, it can be identified with a linear functional acting on another function space.

We have seen that the space $\mathcal{S}(\mathbb{R}^n)$ is contained in every L^p space and furthermore that it is dense in $L^p(\mathbb{R}^n)$ for all $p<\infty$. Restricting our attention to a smaller class of functions, the space $\mathcal{S}(\mathbb{R}^n)$, we get a larger dual space:

$$\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n) \implies L^{p'}(\mathbb{R}^n) = (L^p(\mathbb{R}^n))^* \subset (\mathcal{S}(\mathbb{R}^n))^*.$$

We thus obtain a space of generalized functions that contains the classical L^p spaces. As we shall see, this space is much bigger and in particular it allows us to differentiate (in the appropriate sense) and remain in this class of generalized functions and, most notably, consider the Fourier transform of these objects and still remain in the class. These operation many times are not even available on L^p spaces; for example we cannot even define the Fourier transform on $L^p(\mathbb{R}^n)$ for $p>2$. Furthermore, even when there is a way to define these operations on L^p functions we don't necessarily stay in the given class of functions. For example, while it is perfectly legitimate to define the Fourier transform of an L^1 function, the resulting function $\hat{f}$ is not in general an integrable function. We shall see that the fact that $\mathcal{S}(\mathbb{R}^n)$ is closed under taking partial derivatives, multiplying by polynomials and by taking the Fourier transform of its elements, its dual space is also closed under the corresponding operations.

In what follows we will many times write $\mathcal{S}'$ for the dual $(\mathcal{S}(\mathbb{R}^n))^*$ and $\langle f,g\rangle$ for the pairing $\int f\bar{g}$.

DEFINITION 3.52. A linear functional $\lambda:\mathcal{S}(\mathbb{R}^n)\to\mathbb{C}$ will be called a *tempered distribution* if it is continuous on $\mathcal{S}(\mathbb{R}^n)$ with respect to the topology on $\mathcal{S}(\mathbb{R}^n)$ described in the previous sections.

That is, the linear functional $\lambda:\mathcal{S}(\mathbb{R}^n)\to\mathbb{C}$ is a tempered distribution if and only if there exists some $N\in\mathbb{N}_o$ and $C>0$ such that

$$|\lambda(\phi)| \le Cp_N(\phi),$$

for all $\phi\in\mathcal{S}(\mathbb{R}^n)$.

We equip the space $(\mathcal{S}(\mathbb{R}^n))^*$ with the weak-* topology; a sequence of tempered distributions λ_k converges to a limit λ if one has $\lambda_k(\phi)\to\lambda(\phi)$ for all $\phi\in\mathcal{S}(\mathbb{R}^n)$. This is the weakest topology such that for each $f\in\mathcal{S}(\mathbb{R}^n)$ the functional

$$f^*:(\mathcal{S}(\mathbb{R}^n))^*\to\mathbb{C},\quad f^*(\lambda)=\lambda(f)$$

is continuous. The space $(\mathcal{S}(\mathbb{R}^n))^*$ equipped with this topology will also be denoted by $\mathcal{S}'(\mathbb{R}^n)$.

In what follows we will also use the notation $(f,\lambda)=(\lambda,f)$ for $\lambda(f)$ whenever $\lambda\in\mathcal{S}'(\mathbb{R}^n)$ and $f\in\mathcal{S}(\mathbb{R}^n)$. Be careful not to confuse this pairing with $\langle f,g\rangle = \int f\bar{g}$.

3.5.1. Examples of tempered distributions. We now describe several examples of classes of tempered distributions. We begin by showing how we can identify some known function classes with tempered distributions.

(i) Any element $f\in L^p(\mathbb{R}^n)$, $1\le p\le\infty$ can be identified with an element $\lambda_f\in\mathcal{S}'(\mathbb{R}^n)$ by means of the formula

$$\lambda_f(\phi)=\int_{\mathbb{R}^n} f(x)\phi(x)dx,\quad \phi\in\mathcal{S}(\mathbb{R}^n),$$

and the map $L^p \ni f \mapsto \lambda_f$ is continuous. We will say in this case that the tempered distribution λ_f is an L^p function.

It is clear that λ_f is linear. Furthermore we have that

$$|\lambda_f(\phi)| \le \|f\|_{L^p(\mathbb{R}^n)}\|\phi\|_{L^q(\mathbb{R}^n)} \lesssim_{p,n} \|f\|_{L^p(\mathbb{R}^n)} p_N(\phi),$$

for some non-negative integer N, by Proposition 3.38. By Lemma 3.37 this shows that $\lambda_f \in \mathcal{S}'(\mathbb{R}^n)$. Furthermore, the mapping $f \mapsto \lambda_f$ is continuous. Indeed, if $f_k \to f$ in $L^p(\mathbb{R}^n)$ we set $\lambda_k = \lambda_{f_k}$. We need to show that $\lambda_k \to \lambda_f$ in the weak-* topology, that is, that $\lambda_k(\phi) - \lambda_f(\phi) \to 0$ for every $\phi \in \mathcal{S}(\mathbb{R}^n)$. However this is a consequence of the previous estimate.

(ii) Any element $\psi \in \mathcal{S}(\mathbb{R}^n)$ can be identified with an element $\lambda_\psi \in \mathcal{S}'(\mathbb{R}^n)$ by means of the formula

$$\lambda_\psi(\phi) = \int_{\mathbb{R}^n} \psi(x)\phi(x)dx, \quad \phi \in \mathcal{S}(\mathbb{R}^n),$$

and the map $\mathcal{S}(\mathbb{R}^n) \ni \psi \mapsto \lambda_\psi$ is continuous. We will say in this case that the tempered distribution λ_ϕ is an Schwartz function. The proof is very similar to that of (i).

(iii) If $\mu \in \mathcal{M}(\mathbb{R}^n)$ be a finite Borel measure. Then μ can be identified with a tempered distribution $\lambda_\mu \in \mathcal{S}'(\mathbb{R}^n)$ by means of the formula

$$\lambda_\mu(\phi) = \int_{\mathbb{R}^n} \phi(x)d\mu(x),$$

and the map $\mathcal{M}(\mathbb{R}^n) \mapsto \lambda_\mu$ is continuous. We will say in this case that the tempered distribution λ_μ is a (finite Borel) measure. The proof is the same as that of the preceding cases.

(iv) Let $1 \le p \le \infty$. A measurable function f such that $(1+|x|^2)^{-k} f(x) \in L^p(\mathbb{R}^n)$ for some non-negative integer k is called *a tempered L^p-function*. Again the functional λ_f is an element of $\mathcal{S}'(\mathbb{R}^n)$. For $p = \infty$ such a function is often called *a slowly increasing function*. Similarly a Borel measure μ such that

$$\int_{\mathbb{R}^n} (1+|x|^2)^{-k} d|\mu|(x) < +\infty,$$

is called *a tempered Borel measure* and it defines an element of $\mathcal{S}'(\mathbb{R}^n)$ by setting

$$\lambda_\mu(\phi) = \int_{\mathbb{R}^n} \phi(x)d\mu(x).$$

We will say that the tempered distribution λ_μ is a tempered Borel measure.

EXERCISE 3.53. Show that if μ is a tempered Borel measure then $\lambda_\mu \in \mathcal{S}'(\mathbb{R}^n)$ and the map $\mu \mapsto \lambda_\mu$ is continuous. Conclude the corresponding statement if f is a tempered L^p function. Observe that $f(x)dx$ defines a tempered measure.

EXERCISE 3.54. Show that a Borel measure μ is a tempered measure if and only if it is of *polynomial growth*: for every $R > 0$ we have that

$$\mu(B(0,R)) \lesssim R^k,$$

for some positive integer k and all $R \geq 1$. In particular, μ is locally finite.

REMARK 3.55. From the previous definitions one gets the impression that the term "tempered" is closely connected to "of at most polynomial growth". This is in some sense correct since all functions or measures of at most polynomial growth define tempered distributions. On the other hand, the opposite claim is not true. Indeed, observe that the function $\sin(e^x)$ is a slowly increasing function (actually it is bounded) and thus defines a tempered distribution. Thus, the derivative of this function, $e^x\cos(e^x)$ is also a tempered distribution although it grows exponentially fast.

All the previous examples identify functions and measures (of moderate growth) with tempered distributions and the embeddings are continuous. However the space $\mathcal{S}'(\mathbb{R}^n)$ also contains "rougher" objects which are neither functions nor measures.

EXERCISE 3.56. Show that the functional $\delta_0' : \phi \mapsto -\phi'(0)$ for all $\phi \in \mathcal{S}(\mathbb{R}^n)$ is a tempered distribution which does not arise from a tempered measure (and thus it does not arise from a tempered function either).

EXAMPLE 3.57 (The principal value distribution). We define the functional $\mathrm{p.v.}\,\frac{1}{x}$ as

$$(\mathrm{p.v.}\,\frac{1}{x}\phi) := \lim_{\epsilon\to 0}\int_{|x|>\epsilon}\frac{\phi(x)}{x}dx.$$

Then $\mathrm{p.v.}\,\frac{1}{x} \in \mathcal{S}(\mathbb{R})$.

To see that $\mathrm{p.v.}\,\frac{1}{x} \in \mathcal{S}'(\mathbb{R}^n)$ let us fix some $0<\epsilon<1$ and $\phi \in \mathcal{S}(\mathbb{R}^n)$ and write

$$\int_{|x|>\epsilon}\frac{\phi(x)}{x}dx = \int_{\epsilon<|x|<1}\frac{\phi(x)-\phi(0)}{x}dx + \int_{|x|>1}\frac{\phi(x)}{x}dx.$$

Now observe that $\left|\frac{\phi(x)-\phi(0)}{x}\right| \leq \|\phi'\|_\infty$ thus the limit of the first summand as $\epsilon\to 0$ exists and

$$(\mathrm{p.v.}\,\frac{1}{x}\phi) = \int_{|x|<1}\frac{\phi(x)-\phi(0)}{x}dx + \int_{|x|>1}\frac{\phi(x)}{x}dx.$$

Moreover we have that

$$|(\mathrm{p.v.}\,\frac{1}{x}\phi)| \lesssim \|\phi'\|_\infty + \|x\phi\|_\infty \leq p_1(\phi).$$

Furthermore this distribution does not arise from any locally finite Borel measure. For this consider a Schwartz function ϕ adopted to an interval of the form $(\delta,1)$ for $\delta\to 0$.

EXERCISE 3.58 (The principal value distribution in many dimensions). Let $K : \mathbb{R}^n \to \mathbb{C}$ be a homogeneous function of degree $-n$. This means that

$$K(\lambda x) = \lambda^{-n}K(x), \quad \lambda > 0.$$

(i) Show that there exists a function $\Omega : S^{n-1} \to \mathbb{C}$ such that $K(x) = \Omega(x')/|x|^n$ where $x' = x/|x| \in S^{n-1}$.

(ii) Assume that $\Omega \in L^1(S^{n-1})$ and $\int_{S^{n-1}}\Omega(x')d\sigma_{n-1}(x') = 0$. For $\phi \in \mathcal{S}(\mathbb{R}^n)$ we define

$$\mathrm{p.v.}\,K(\phi) = \lim_{\epsilon\to 0}\int_{|x|>\epsilon}K(x)\phi(x)dx.$$

Show that the limit in the previous definition the limit exists and that p. v. K defines a tempered distribution.

3.5.2. Basic operations on the space of tempered distributions. We have already seen that the space $\mathcal{S}(\mathbb{R}^n)$ is closed under several basic operations: differentiation, multiplying by polynomials, multiplying by Schwartz functions and, most notably, the Fourier transform. The space of tempered distributions has very similar properties:

3.5.2.1. *Derivatives in $\mathcal{S}'(\mathbb{R}^n)$:* We begin the discussion by considering $\phi, \psi \in \mathcal{S}(\mathbb{R}^n)$ and writing down the integration by parts formula

$$\int_{\mathbb{R}^n} (\partial^\beta \psi)(x)\phi(x)dx = (-1)^{|\beta|} \int_{\mathbb{R}^n} \psi(x)(\partial^\beta \phi)(x)dx.$$

According to the previous definitions we can rewrite the previous formula as

$$(\partial^\beta \psi, \phi) = (-1)^{|\beta|}(\psi, \partial^\beta \phi),$$

or

$$\lambda_{\partial^\beta \psi}(\phi) = (-1)^{|\beta|}\lambda_\psi(\partial^\beta \phi).$$

The right hand side of the previous identity though makes sense for any $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ in the place of λ_ψ. Also, for $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ the mapping $\phi \mapsto \lambda(\partial^\beta \phi)$ is continuous since λ is continuous and the map $\phi \mapsto \partial^\beta \phi$ is continuous. We thus define the partial derivative $\partial^\beta \lambda$ of any $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ by means of

$$(\partial^\beta \lambda)(\phi) := (-1)^{|\beta|}\lambda(\partial^\beta \psi).$$

The previous discussion implies that $\partial^\beta \lambda \in \mathcal{S}'(\mathbb{R}^n)$.

EXAMPLE 3.59. Let f be the tempered L^∞ function defined as

$$f(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

The function f is many times called the *Heaviside step function*. Clearly f defines a tempered distribution λ_f in the usual way

$$\lambda_f(\phi) = \int_{\mathbb{R}} f(x)\phi(x)dx, \quad \phi \in \mathcal{S}(\mathbb{R}).$$

For every $\phi \in \mathcal{S}(\mathbb{R})$ we then have

$$\lambda_f'(\phi) = -\lambda_f(\phi') = -\int_{\mathbb{R}} f(x)\phi'(x)dx = -\int_0^\infty \phi'(x)dx = \phi(0) = \int_{\mathbb{R}} \phi(x)d\delta_0(x).$$

That is $\lambda_f' = d\delta_0$.

REMARK 3.60. The fact that the distributional derivative of the Heaviside step function is the Dirac mass at 0 is intuitively obvious. The function f is differentiable everywhere except at 0 and $f'(x) = 0$ whenever $x \neq 0$. On the other hand there is a jump discontinuity of weight equal to 1 at 0 which roughly speaking requires an infinite derivative to be realized. In general, a jump discontinuity of weight a at a point x_o has a distributional derivative which coincides with Dirac mass of weight a at the point x_o .

EXAMPLE 3.61. Let δ_0 be a Dirac mass at 0. We then have

$$(\partial^\beta\delta_0)(\phi) = (-1)^{|\beta|}\delta_0(\partial^\beta\phi) = (-1)^{|\beta|}\partial^\beta\phi(0).$$

This also explains the minus sign in Exercise 3.56.

EXERCISE 3.62. In dimension $n = 1$ show that:

(i) The distributional derivative of the signum function $\operatorname{sgn}(x) := x/|x|$ is $2\delta_0$.
(ii) The distributional derivative of the locally integrable function $\log|x|$ is equal to $\mathrm{p.v.}\,\frac{1}{x}$.
(iii) The distributional derivative of the locally integrable function $|x|$ is equal to $\operatorname{sgn}(x)$.

3.5.2.2. *Translations, Modulations, Dilations and reflections in $\mathcal{S}'(\mathbb{R}^n)$:* We have see that the translation operator τ_h maps a measurable function f to the function $f(\cdot - h)$, where $h \in \mathbb{R}^n$. A trivial change of variables shows that whenever $f\phi \in L^1(\mathbb{R}^n)$ we have that

$$\int_{\mathbb{R}^n}(\tau_h f)(x)\phi(x)dx = \int_{\mathbb{R}^n} f(x)(\tau_{-h}\phi)(x)dx.$$

Now assume that f is a tempered L^p function (say). In the language of distributions we can rewrite the previous identity as

$$\lambda_{\tau_h f}(\phi) = \lambda_f(\tau_{-h}\phi),$$

for all $\phi \in \mathcal{S}(\mathbb{R}^n)$. Again, the right hand side of this identity is well defined for any $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ and we define the translation of any distribution $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ as

$$(\tau_h\lambda)(\phi) := \lambda(\tau_{-h}\phi), \quad \phi \in \mathbb{R}^n.$$

It is easy to see that $\tau_h\lambda \in \mathcal{S}'(\mathbb{R}^n)$.

Similarly we define for $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ and $\phi \in \mathcal{S}(\mathbb{R}^n)$ the tempered distributions

$$\begin{aligned}\tilde{\lambda}(\phi) &:= \lambda(\tilde{\phi}),\\ (\mathrm{Mod}_y)\lambda(\phi) &:= \lambda(\mathrm{Mod}_y\phi),\\ (\mathrm{Dil}_t^p)\lambda(\phi) &:= \lambda(\mathrm{Dil}_{t^{-1}}^{p'}\phi), \quad t > 0.\end{aligned}$$

3.5.2.3. *Convolution in $\mathcal{S}'(\mathbb{R}^n)$:* Let $f, g, h \in \mathcal{S}(\mathbb{R}^n)$. Then it is an easy application of Fubini's theorem that

$$\int_{\mathbb{R}^n}(f * g)(x)h(x)dx = \int_{\mathbb{R}^n} f(x)(\tilde{g} * h)(x)dx,$$

where $\tilde{g}(x) = g(-x)$ is the reflection of g. In the language of distributions the previous identity reads

$$\lambda_{f*g}(h) = \lambda_f(\tilde{g} * h), \quad h \in \mathcal{S}(\mathbb{R}^n).$$

Now the right hand side of the previous identity is well defined whenever $g*h \in \mathcal{S}(\mathbb{R}^n)$ while in order to define the distribution $f * g$ we need to have that $h \in \mathcal{S}(\mathbb{R}^n)$. Now assume that g is a function such that $g*\phi \in \mathcal{S}(\mathbb{R}^n)$ for all $\phi \in \mathcal{S}(\mathbb{R}^n)$. This is obviously the case if $g \in \mathcal{S}(\mathbb{R}^n)$. Thus we can define the convolution of any $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ with a function $g \in \mathcal{S}(\mathbb{R}^n)$ by means of the formula

$$(\lambda * g)(\phi) := \lambda(\tilde{g} * \phi), \quad \phi \in \mathcal{S}(\mathbb{R}^n).$$

It is easy to see that the function $\lambda * g$ is continuous as a composition of the continuous maps $\phi \mapsto \tilde{g} * \phi$ and $\psi \mapsto \lambda(\psi)$ thus $\lambda * g \in \mathcal{S}'(\mathbb{R}^n)$ for every $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ and $g \in \mathcal{S}(\mathbb{R}^n)$.

EXERCISE 3.63. Actually, the condition $g \in \mathcal{S}(\mathbb{R}^n)$ is a bit too much to ask if one just wants to define the convolution $\lambda * g$. As we have observed, the only requirement is that $g * \phi \in \mathcal{S}(\mathbb{R}^n)$ whenever $\phi \in \mathcal{S}(\mathbb{R}^n)$. Suppose that g is a rapidly decreasing function, that is $|x|^k f(x) \in L^\infty(\mathbb{R}^n)$ for all $k = 0, 1, 2, \ldots$. Show the convolution of $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ and g can be defined and that it is again an element of $\mathcal{S}'(\mathbb{R}^n)$.

It turns out that the convolution of a tempered distribution with a Schwartz function is a function:

THEOREM 3.64. *Let $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ and $h \in \mathcal{S}(\mathbb{R}^n)$. Then the convolution $\lambda * h$ is the* function *f given by the formula*

$$(\lambda * h)(x) = \lambda(\tau_x \tilde{h}), \quad x \in \mathbb{R}^n.$$

Moreover, $f \in C^\infty(\mathbb{R}^n)$ and for all multi-indices α the function $\partial^\alpha f$ is slowly increasing.

For the proof of this theorem see [SW].

3.5.2.4. *The Fourier transform on $\mathcal{S}'(\mathbb{R}^n)$:* We now come to the definition and action of the Fourier transform on tempered distribution. As in all the other definitions, first we investigate what happens in the case the tempered distribution is a Schwartz function. So, letting $\phi, f \in \mathcal{S}(\mathbb{R}^n)$ the multiplication formula implies that

$$\int_{\mathbb{R}^n} \phi(x)\hat{f}(x)dx = \int_{\mathbb{R}^n} \hat{\phi}(x) f(x)dx.$$

In the language of tempered distributions we have that

$$\lambda_{\hat{f}}(\phi) = \lambda_f(\hat{\phi}).$$

Observing once more that the right hand side is well defined for all $\phi \in \mathcal{S}(\mathbb{R}^n)$ and that the map $\mathcal{S}(\mathbb{R}^n) \ni \phi \mapsto \lambda(\hat{\phi})$ is well defined and continuous we define the Fourier transform of any tempered distribution $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ as

$$\mathcal{F}(\lambda)(\phi) = \hat{\lambda}(\phi) := \lambda(\hat{\phi}), \quad \phi \in \mathcal{S}(\mathbb{R}^n).$$

We have that $\hat{\lambda} \in \mathcal{S}'(\mathbb{R}^n)$ whenever $\lambda \in \mathcal{S}'(\mathbb{R}^n)$. It is also trivial to define the inverse Fourier transform of a tempered distribution as

$$\mathcal{F}^{-1}(\lambda)(\phi) := \check{\lambda}(\phi) = \lambda(\check{\phi}),$$

and to show that $\mathcal{F}$ is a homeomorphism of $\mathcal{S}'(\mathbb{R}^n)$ onto itself. Also the operator $\mathcal{F} : \mathcal{S}'(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$ satisfies all the symmetry properties that the classical Fourier transform satisfies and commutes with derivatives in the same way.

EXAMPLE 3.65 (The Fourier transform of $|x|^{-2}$ in $\mathbb{R}^3$). We consider the function

$$f(x) = \frac{1}{|x|^2}, \quad x \in \mathbb{R}^3.$$

Note that f is locally integrable in $\mathbb{R}^3$ and it decays at infinity thus it can be identified with a tempered distribution which we will still call f. On the other hand f is not in any L^p space so we can't consider its Fourier transform in the classical sense. We claim that the Fourier transform of f in the sense of distributions is given as

$$\hat{f}(\xi) = \frac{\pi}{|\xi|}.$$

First of all observe that it suffices to show that

$$\int_{\mathbb{R}^3} \frac{1}{|x|^2}\hat{\phi}(x)dx = \int_{\mathbb{R}^3} \frac{\pi}{|\xi|}\phi(\xi)d\xi, \tag{3.9}$$

for all $\phi \in \mathcal{S}(\mathbb{R}^3)$. Here it is convenient to express the function $1/|x|^2$ as an average of functions with known Fourier transforms. Indeed, this can be done by means of the identity

$$\frac{1}{2\pi|x|^2} = \int_0^\infty te^{-\pi t^2|x|^2}dt,$$

which can be proved by simple integration by parts. Now fix a function $\phi \in \mathcal{S}(\mathbb{R}^3)$. We have that

$$\begin{aligned}\int_{\mathbb{R}^3} \frac{1}{|x|^2}\hat{\phi}(x)dx &= 2\pi \int_{\mathbb{R}^3}\int_0^\infty te^{-\pi t^2|x|^2}dt\ \hat{\phi}(x)dx\\ &= 2\pi\int_0^\infty t\Big(\int_{\mathbb{R}^3} e^{-\pi t^2|x|^2}\hat{\phi}(x)dx\Big)dt,\end{aligned}$$

by an application of Fubini's theorem since the function $te^{-\pi t^2|x|^2}\hat{\phi}(x)$ is an integrable function on $(0,\infty)\times\mathbb{R}^3$. The inner integral can be calculated now by using the multiplication formula and the (known) Fourier transform of a Gaussian. Indeed we have

$$\int_{\mathbb{R}^3} e^{-\pi t^2|x|^2}\hat{\phi}(x)dx = \int_{\mathbb{R}^3}\frac{1}{t^3}e^{-\pi\frac{|x|^2}{t^2}}\phi(x)dx.$$

Putting the last two identities together we get

$$\int_{\mathbb{R}^3} \frac{1}{|x|^2}\hat{\phi}(x)dx = 2\pi\int_0^\infty \Big(\int_{\mathbb{R}^3}\frac{1}{t^2}\phi(x)e^{-\pi|x|^2/t^2}dx\Big)dt.$$

Now observe that by changing variables $s = |x|/t$ we have

$$\int_0^\infty \frac{1}{t^2}e^{-\pi|x|^2/t^2}dt = \frac{1}{|x|}\int_0^\infty e^{-\pi s^2}ds = \frac{1}{2|x|},$$

and thus

$$\int_0^\infty\int_{\mathbb{R}^3}|\phi(x)|\frac{1}{t^2}e^{-\pi|x|^2/t^2}dt = \int_{\mathbb{R}^3}|\phi(x)|\frac{1}{|x|}dx < \infty,$$

since $|x|^{-1}$ is locally integrable in $\mathbb{R}^3$ and $\phi \in \mathcal{S}(\mathbb{R}^3)$. A second application of Fubini's theorem then gives (3.9) and proves the claim.

EXAMPLE 3.66 (The Fourier transform of the principal value distribution). It will be quite useful and instructive to calculate the Fourier transform of the tempered distribution $\mathrm{p.v.}\,\frac{1}{x}$. We will show in this example that

$$\mathcal{F}(\mathrm{p.v.}\,\frac{1}{x})(\xi) = -\pi i\,\mathrm{sgn}(\xi).$$

Since this tempered distribution does not arise from a function this is not completely straightforward. Instead we argue as follows. Let $\epsilon > 0$ and set $\psi_\epsilon(x) := \frac{1}{x}\mathbf{1}_{\{|y|>\epsilon\}}(x)$ and $Q_\epsilon(x) := \frac{x}{\epsilon^2+x^2}$. We first show that

$$\lim_{\epsilon\to 0}(\psi_\epsilon - Q_\epsilon) = 0 \quad \text{in} \quad \mathcal{S}'(\mathbb{R}^n). \tag{3.10}$$

Indeed for $\phi \in \mathcal{S}(\mathbb{R}^n)$ we have

$$\begin{aligned}(Q_\epsilon - \psi_\epsilon)(\phi) &= \int_{|x|>\epsilon} \frac{\phi(x)}{x}dx - \int_{\mathbb{R}} \frac{x\phi(x)}{\epsilon^2 + x^2}dx \\ &= \int_{|x|<\epsilon} \frac{x\phi(x)}{\epsilon^2 + x^2} + \int_{|x|>\epsilon}\Big(\frac{x}{\epsilon^2 + x^2} - \frac{1}{x}\Big)\phi(x)dx \\ &= \int_{|x|<1} \frac{x\phi(\epsilon x)}{1 + x^2}dx - \int_{|x|>1} \frac{\phi(\epsilon x)}{x(1 + x^2)}dx.\end{aligned}$$

Taking limits as $\epsilon \to 0$ and applying dominated convergence we get the difference of the integrals

$$\int_{|x|<1} \frac{x\phi(0)}{1 + x^2}dx, \qquad \int_{|x|>1} \frac{\phi(0)}{x(1 + x^2)}dx$$

which are both identically 0 since the functions under the integral sign are odd, and the domain of integration is symmetric around 0. This shows (3.10).

Now, by considering Fourier transforms in $\mathcal{S}'(\mathbb{R}^n)$ and using the asymptotic (3.10) we get that

$$\mathcal{F}(\mathrm{p.\,v.}\,\frac{1}{x})(\xi) = \lim_{\epsilon\to 0}\widehat{Q_\epsilon}(\xi).$$

Observe that the Fourier transform of Q_ϵ has to be understood in the sense of distributions since Q_ϵ is not integrable.

Now one can show that $Q_\epsilon(x) = \mathcal{F}^{-1}(-\pi i\,\mathrm{sgn}(\xi)e^{-2\pi\epsilon|\xi|})(x)$. Indeed we have

$$\begin{aligned}\mathcal{F}^{-1}(-\pi i\,\mathrm{sgn}(\xi)e^{-2\pi\epsilon|\xi|})(x) &= -i\pi\int_{\mathbb{R}} \mathrm{sgn}(\xi)e^{-2\pi\epsilon|\xi|}e^{2\pi i x\xi}d\xi \\ &= \pi i\int_{-\infty}^{0} e^{2\pi(\epsilon+ix)\xi}d\xi - \pi i\int_{0}^{+\infty} e^{2\pi(-\epsilon+ix)\xi}d\xi \\ &= \pi i\Big(\frac{1}{2\pi(\epsilon + ix)} + \frac{1}{2\pi(-\epsilon + ix)}\Big) \\ &= \pi i\frac{2ix}{2\pi(-x^2 - \epsilon^2)} = \frac{x}{x^2 + \epsilon^2}.\end{aligned}$$

The fact that the inverse Fourier transform of the function $-\pi i\,\mathrm{sgn}(\xi)e^{2\pi i\epsilon|\xi|}$ is Q_ϵ and the injectivity of the Fourier transform on $\mathcal{S}'(\mathbb{R}^n)$ imply that $\widehat{Q_\epsilon}(\xi) = -\pi i\,\mathrm{sgn}(\xi)e^{2\pi\epsilon|\xi|}$ in the sense of distributions. Now clearly $\lim_{\epsilon\to 0}\widehat{Q_\epsilon}(\xi) = -i\,\mathrm{sgn}(\xi)$ so we are done.

EXERCISE 3.67. The purpose of this exercise is to show that the Fourier transform in the sense of distributions acts on convolutions as the usual Fourier transform.

(i) Let f be a smooth function such that for all multi-indices α the partial derivatives $\partial^\alpha f$ have at most polynomial growth: $|\partial^\alpha f(x)| \lesssim (1 + |x|^2)^k$, for some $k \geq 0$. Then the *product* of a tempered distribution $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ with f is well defined by means of the formula

$$(\lambda f)(\phi) = \lambda(f\phi), \quad \phi \in \mathcal{S}(\mathbb{R}^n),$$

and $\lambda f \in \mathcal{S}'(\mathbb{R}^n)$.

(ii) If $\lambda \in \mathcal{S}'(\mathbb{R}^n)$ and $f \in \mathcal{S}(\mathbb{R}^n)$ then show that

$$\widehat{\lambda * f} = \hat{\lambda}\hat{f} \quad \text{in} \quad \mathcal{S}'(\mathbb{R}^n).$$

REMARK 3.68. The definition of the Fourier transform on $\mathcal{S}'(\mathbb{R}^n)$ implies that whenever $f \in L^p(\mathbb{R}^n)$, $1 \leq p \leq 2$ we have that

$$\hat{\lambda}_f = \lambda_{\hat{f}}.$$

Thus the Fourier transform on tempered distributions is an extension of the classical definition of the Fourier transform. If on the other hand $f \in L^p(\mathbb{R}^n)$ for some $2 < p \leq \infty$ then f is a tempered L^p function and thus λ_f is a tempered distribution. This allows us to define the Fourier transform of f by looking at f as a tempered distribution. The discussion that followed the Hausdorff-Young theorem however suggests that $\hat{\lambda}_f$ will not be a function in general.

EXERCISE 3.69 (Poisson summation formula). For $f \in \mathcal{S}(\mathbb{R}^n)$ we define

$$\Lambda(f) = \sum_{k=(k_1,\ldots,k_n)\in\mathbb{Z}^n} f(k).$$

Note that Λ can be identified with the sum of a unit masses positioned on every point of the integer lattice $\mathbb{Z}^d$

$$\Lambda = \sum_{k\in\mathbb{Z}^n} \delta_k.$$

Show that $\Lambda \in \mathcal{S}'(\mathbb{R}^n)$ and that $\mathcal{F}\Lambda = \Lambda$.

Hints:

(a) First prove the case of dimension $n = 1$ by proving the following intermediate statements.

(i) Show that Λ satisfies the invariances $\tau_1\Lambda = \Lambda$ and $\mathrm{Mod}_1\Lambda = \Lambda$.

(ii) Consider a Schwartz function $g \in \mathcal{S}(\mathbb{R})$ with support in the interval $(-\frac{1}{4}, \frac{1}{4})$ and $g(0) = 1$. If $f \in \mathcal{S}(\mathbb{R})$ has compact support show that the function

$$h(x) = \frac{f(x) - \sum_{m\in\mathbb{Z}} f(m)\tau_m g(x)}{1 - e^{2\pi i x}}$$

is a smooth function with compact support.

(iii) Let Λ' be a tempered distribution which satisfies the invariances $\tau_1\Lambda' = \Lambda'$ and $\mathrm{Mod}_1\Lambda' = \Lambda'$. Show that

$$\Lambda'(f - \sum_{k\in\mathbb{Z}} f(k)\tau_k(g)) = 0$$

whenever f, g are as in step (ii). Conclude that

$$\Lambda'(f) = c\Lambda(f)$$

for some $c \in \mathbb{C}$, whenever f is a Schwartz function with compact support. Extend this equality to all $f \in \mathcal{S}(\mathbb{R})$ by a density argument.

(iv) Step (iii) essentially shows that any tempered distribution that has the symmetries in (i) must agree with Λ up to a multiplicative constant. Observe that $\mathcal{F}\Lambda$ satisfies the same invariances. Conclude that $\Lambda = c\hat{\Lambda}$ by step (i). Determine the numerical constant $c \in \mathbb{C}$ by testing against the Schwartz function $f(x) = e^{-\pi x^2}$. This concludes the proof for the one dimensional case.

(b) For general n use Fubini's theorem to show that

$$\mathcal{F} = \mathcal{F}_{x_1}\mathcal{F}_{x_2}\cdots\mathcal{F}_{x_n},$$

where $\mathcal{F}_{x_j}$ denotes the (one-dimensional) Fourier transform in the $j-th$ direction. Thus step (a) implies that

$$\mathcal{F}_{x_j}\Lambda = \Lambda,$$

for every $j = 1, 2, \ldots, n$. Conclude the proof by iterating this identity.

EXERCISE 3.70 (Equivalent form of Poisson summation formula). If $f \in \mathcal{S}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$ then we have that

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

3.6. The Uncertainty principle

A typical manifestation of the uncertainty principle is the following "inverse Hölder" type of bound.

LEMMA 3.71 (Bernstein's Inequality for a ball). *Let $f \in L^1 + L^2$ and $\hat{f}$ is supported in a Euclidean ball $B(0,R) \subset \mathbb{R}^n$. Then*

(i) *For any multi-index α we have*

$$\|\partial^\alpha f\|_{L^p(\mathbb{R}^n)} \lesssim_\alpha R^{|\alpha|}\|f\|_{L^p(\mathbb{R}^n)}.$$

(ii) *For $1 \le p \le q \le +\infty$ we have*

$$\|f\|_{L^q(\mathbb{R}^n)} \lesssim R^{n(\frac{1}{p}-\frac{1}{q})}\|f\|_{L^p(\mathbb{R}^n)}.$$

In other words, if a function f is frequency localized then the lower L^p norms control the higher L^p norms.

PROOF. For the first part let us consider some function $\phi \in \mathcal{S}(\mathbb{R}^n)$ such that $\hat{\phi}$ is identically 1 on $B(0,1)$. Setting $\psi := \operatorname{Dil}^1_{R^{-1}}\phi$ we have that $\hat{\psi}(\xi) = \operatorname{Dil}^\infty_R \hat{\phi}(\xi) = \phi(\xi/R)$ thus $\hat{\psi}$ is identically 1 on $B(0,R)$. We then have $\hat{f} = \hat{\psi}\hat{f}$ so that $f = f * \psi$. Furthermore for any $r \in [1,\infty]$ we have

$$\|\psi\|_{L^r(\mathbb{R}^n)} = \|\operatorname{Dil}^1_{R^{-1}}\phi\|_{L^r(\mathbb{R}^n)} = \|\phi\|_{L^r(\mathbb{R}^n)}R^{\frac{n}{r'}}.$$

For the derivatives of ψ an easy application of the chain rule shows that

$$\|\nabla\psi\|_{L^1(\mathbb{R}^n)} = R\|\nabla\phi\|_{L^1(\mathbb{R}^n)}.$$

In the case $|\alpha| = 1$ of (i) we then have $\alpha = (0,\ldots,0,\alpha_j,0\ldots,0)$ for some $j \in \{1,2,\ldots,n\}$ so that

$$\begin{aligned}\|\partial^\alpha f\|_{L^p(\mathbb{R}^n)} &= \|\partial^\alpha(f*\psi)\|_{L^p(\mathbb{R}^n)} = \|f*\partial^\alpha\psi\|_{L^p(\mathbb{R}^n)}\\ &\le \|f\|_{L^p(\mathbb{R}^n)}\|\partial^\alpha\psi\|_{L^1(\mathbb{R}^n)} \le \|\nabla\phi\|_{L^1(\mathbb{R}^n)}R\|f\|_{L^p(\mathbb{R}^n)}\\ &\lesssim R\|f\|_{L^p(\mathbb{R}^n)}.\end{aligned}$$

The case of a general multi-index α follows now by induction.

For (ii) let $\frac{1}{p}+\frac{1}{r}=1+\frac{1}{q}$. Since $f=f*\psi$ we can apply Young's inequality from Proposition 2.28 to estimate

$$\begin{aligned}\|f\|_{L^q(\mathbb{R}^n)} &= \|f*\psi\|_{L^q(\mathbb{R}^n)} \leq \|\psi\|_{L^r(\mathbb{R}^n)}\|f\|_{L^p(\mathbb{R}^n)}\\ &= R^{\frac{n}{r'}}\|\phi\|_{L^r(\mathbb{R}^n)}\|f\|_{L^p(\mathbb{R}^n)}\\ &\simeq R^{n(\frac{1}{p}-\frac{1}{q})}\|f\|_{L^p(\mathbb{R}^n)}\end{aligned}$$

which is the desired estimate. □

3.7. Translation invariant operators

Let V, W be vector spaces of functions on $\mathbb{R}^n$ and suppose that T is an operator that maps V into W. We will say that T *commutes with translations* or that T *is translation invariant* if $T\tau_y=\tau_y T$ for all $y\in\mathbb{R}^n$. To see an example of such an operator, consider $K\in L^1(\mathbb{R}^n)$ and define $T_K(f)(x)=(f*K)(x)$ for all $f\in L^p(\mathbb{R}^n)$, $1\leq p\leq\infty$. We have seen that T_K is well defined and furthermore that

$$\|T_K(f)\|_{L^p(\mathbb{R}^n)}\leq\|K\|_{L^1(\mathbb{R}^n)}\|f\|_{L^p(\mathbb{R}^n)},$$

that is, T_K is of strong type (p,p). We have seen that the convolution commutes with translations which implies that T_K commutes with translations. Actually the opposite is also true, namely, all translation invariant operators are given by a convolution with an appropriate kernel K (which might not be a function in general).

THEOREM 3.72. *Let $T:L^p(\mathbb{R}^n)\to L^q(\mathbb{R}^n)$, $1\leq p,q\leq\infty$, be a bounded linear operator that commutes with translations. Then there exists a unique tempered distribution K such that*

$$T(f)=f*K,\quad \textit{for all}\quad f\in\mathcal{S}(\mathbb{R}^n).$$

Thus, translation invariant bounded linear operators of strong type (p,q) are in a one to one correspondence with the subclass of tempered distributions K which satisfy

$$\|K*f\|_q\lesssim\|f\|_p,$$

for all $f\in\mathcal{S}(\mathbb{R}^n)$. In this case we will slightly abuse language and say that the tempered distribution K is of type (p,q). It would be desirable to characterize this class of tempered distribution for all $1\leq p,q\leq\infty$ but such a characterization is not known in general and probably does not exist. Here we gather some partial results in this direction:

PROPOSITION 3.73 ("The high exponents are on the left"). *Suppose that T is a linear operator which is translation invariant and of strong type (p,q). Then we must have that $p\leq q$. In particular the class of tempered distributions of type (p,q) is empty whenever $p>q$.*

EXERCISE 3.74. Prove Proposition 3.73 above.

Hint: Suppose that a that T is translation invariant and of strong type (p,q) with $p<\infty$. Let $f\in L^p(\mathbb{R}^n)$ and consider the function

$$g(x)=\sum_{k=1}^{N}f(x-x_n),$$

for some large positive integer N and points $x_1, \ldots, x_n$ that will be chosen appropriately. Show that by choosing the points $x_1, \ldots, x_n$ to be far apart from each other we have that $\|g\|_p \simeq_{f,p} N^{\frac{1}{p}} \|f\|_p$ while the left hand side will be of the order $\|Tg\|_q \simeq_{q,f} N^{\frac{1}{q}}$ for N large. However, if T is of strong type (p, q) this is only possible if $q \geq p$.

We also have a precise characterization of translation invariant operators in the following two special cases.

THEOREM 3.75 ($p = q = 2$). *A distribution K is of type $(2, 2)$ if and only if there exists $m \in L^\infty(\mathbb{R}^n)$ such that $\hat{K} = m$. In this case, the norm of the operator*

$$T_K : L^2(\mathbb{R}^n) \cap \mathcal{S}(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$$

defined on $\mathcal{S}(\mathbb{R}^n)$ as

$$T_K(f) = f * K, \quad f \in \mathcal{S}(\mathbb{R}^n),$$

is equal to $\|m\|_{L^\infty(\mathbb{R}^n)}$. Moreover, $\widehat{T_K f} = m\hat{f}$.

THEOREM 3.76 ($p = q = 1$). *A distribution K is of type $(1, 1)$ if and only if it is a finite Borel measure. In this case, the norm of the operator*

$$T_K : L^1(\mathbb{R}^n) \cap \mathcal{S}(\mathbb{R}^n) \to L^1(\mathbb{R}^n),$$

defined on $\mathcal{S}(\mathbb{R}^n)$ as

$$T_K(f) = f * K, \quad f \in \mathcal{S}(\mathbb{R}^n),$$

is equal to the total variation $\|K\|$ of the measure K.

For the proofs of these theorems and more details see [SW].

In this course we will not actually need that every translation invariant operator is a convolution operator since we will mostly consider specific examples where this is obvious. We will focus instead on the following case.

3.7.1. Multiplier Operators. Let $m \in L^\infty(\mathbb{R}^n)$. For $f \in L^2(\mathbb{R}^n)$ we define

$$\widehat{T_m(f)}(\xi) := m(\xi)\hat{f}(\xi), \quad \xi \in \mathbb{R}^n.$$

We will say that T_m is a *multiplier operator associated to the (Fourier) multiplier m*.

Observe that T_m is a well defined linear operator on $L^2(\mathbb{R}^n)$ and in fact it is bounded. Rather than relying on Theorem 3.75 let us see this directly:

$$\begin{aligned} \|T_m(f)\|_{L^2(\mathbb{R}^n)} &= \|\widehat{T_m(f)}\|_{L^2(\mathbb{R}^n)} = \|m\hat{f}\|_{L^2(\mathbb{R}^n)} \\ &\leq \|m\|_{L^\infty(\mathbb{R}^n)} \|\hat{f}\|_{L^2(\mathbb{R}^n)} = \|m\|_{L^\infty(\mathbb{R}^n)} \|f\|_{L^2(\mathbb{R}^n)}. \end{aligned}$$

In fact it is not hard to check that the opposite inequality is true so that $\|T_m\|_{L^2 \to L^2} = \|m\|_{L^\infty(\mathbb{R}^n)}$.

EXERCISE 3.77. If T_m is a multiplier operator associated to the multiplier $m \in L^\infty(\mathbb{R}^n)$ show that

$$\|T_m\|_{L^2 \to L^2} \geq \|m\|_{L^\infty(\mathbb{R}^n)}.$$

Thus T_m is a linear operator of type $(2, 2)$. If T_m extends to a linear operator of type (p, p), that is if there is an estimate of the form

$$\|Tf\|_{L^p(\mathbb{R}^n)} \leq c_{p,T} \|f\|_{L^p(\mathbb{R}^n)},$$

for all $f \in \mathcal{S}(\mathbb{R}^n)$, then we will say that m is a (Fourier) *multiplier on L^p*.

REMARK 3.78. The previous discussion and in particular Theorem 3.75 shows that T_m is in fact given in the form

$$T_m(f) = f * K,$$

for some $K \in \mathcal{S}'(\mathbb{R}^n)$. In particular K is the inverse Fourier transform of m in the sense of distributions.

CHAPTER 4

The Hardy-Littlewood maximal function

4.1. Averages and maximal operators

In this chapter we will be discuss the Hardy-Littlewood maximal function and some of its closely related variants. Let us first of all define the averages of a locally integrable function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ around the point $x \in \mathbb{R}^n$:

$$A_r(f)(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)dy,$$

where $B(x,r)$ is the Euclidean ball with center $x \in \mathbb{R}^n$ and radius $r > 0$ and $|B(x,r)|$ denotes its Lebesgue measure. Note that since Lebesgue measure is translation invariant we have

$$|B(x,r)| = |B(0,r)| = r^n|B(0,1)| = \Omega_n r^n,$$

where Ω_n denotes the Lebesgue measure (or volume in this case) of the n-dimensional unit ball $B(0,1) \subset \mathbb{R}^n$. Denoting by χ the indicator function of the normalized unit ball

$$\chi(x) := \frac{1}{|B(0,1)|}\mathbf{1}_{B(0,1)}(x),$$

and noting that Euclidean balls centered at zero are 0-symmetric, we can write

$$\begin{aligned} A_r(f)(x) &= \frac{1}{|B(0,1)|r^n} \int_{B(0,r)} f(x-y)dy \\ &= \int_{\mathbb{R}^n} f(x-y)\frac{1}{|B(0,1)|r^n}\mathbf{1}_{B(0,1)}(y/r)dy \\ &= (f * \chi_r)(x). \end{aligned}$$

Thus

$$A_r(f)(x) = (f * \chi_r)(x),$$

and of course χ_r is an approximation to the identity since $\int_{\mathbb{R}^n} |\chi| = \int_{\mathbb{R}^n} \chi = 1$ and χ_r is just the *dilation* of the function χ:

$$\chi_r(x) = \frac{1}{r^n}\chi\Big(\frac{x}{r}\Big) = \mathrm{Dil}^1_r\,\chi(x).$$

Remembering the discussion that followed the definition of the convolution in Chapter 2, the convolution of a locally integrable function f with the dilations of an L^1-function ϕ was viewed as an averaging operator. We now see that when $\phi = \chi$ this is a precise statement, that is, $f * \chi_r$ is the average of f with respect to a ball around x of radius r. A similar conclusion follows if we start with any set K that is say a bounded convex set in $\mathbb{R}^n$ with non empty interior, that is 0-symmetric, and normalized to volume $|K| = 1$. We then have that

$$f * (\mathbf{1}_K)_r = \frac{1}{|rK|}\int_{x+rK} f(y)dy = A_r^K(f)(x),$$

that is, $(f * (\mathbf{1}_K)_r)(x)$ are the averages of f with respect to the dilations of the fixed convex body K at every point $x \in \mathbb{R}^n$. Here we denote by rK the dilations of K

$$rK := \{rx : x \in K\}.$$

It is an easy exercise to show that all these averages are uniformly bounded in size. For all $1 \leq p \leq \infty$ we have

$$\|A_r^K(f)\|_{L^p(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)}.$$

One of course could consider more general sets K instead of convex sets which are 0-symmetric and in fact this leads to one of the most interesting family of problems in harmonic analysis. This however falls outside the scope of this course and we will mostly focus on the case of the normalized unit ball which in some sense is the prototypical example.

The Hardy-Littlewood maximal operator (with respect to Euclidean balls) is defined as

$$Mf(x) := \sup_{r>0} \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|dy = \sup_{r>0}(A_r|f|)(x) = \sup_{r>0}(|f| * \chi_r)(x).$$

Observe that this is a sublinear operator that is well defined at least when f is locally integrable. It is not hard to check that for $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ the function Mf is slightly more regular than the function f itself. We need the following definition:

DEFINITION 4.1. Let $f : \mathbb{R}^n \to \mathbb{C}$ and $x_o \in \mathbb{R}^n$. Then f is said to be *lower semicontinuous* at x_o if

$$\liminf_{x\to x_o} f(x) \geq f(x_o).$$

The function f is called lower semicontinuous if it is lower semicontinuous at every $x \in \mathbb{R}^n$. A function f is called upper semicontinuous if $-f$ is lower semicontinuous

PROPOSITION 4.2. *Let $f : \mathbb{R}^n \to \mathbb{C}$ where $\subset \mathbb{R}^n$ is measurable.*

(i) *If $f(x_o) = -\infty$ then f is lower semicontinuous at x_o.*
(ii) *If $f(x_o) > -\infty$ then f is lower semicontinuous at x_o if and only if, for every $M < f(x_o)$, there exists $\delta > 0$ such that $f(x) > M$ if $|x - x_o| < \delta$.*
(iii) *The function f is continuous at $x_o \in \mathbb{R}^n$ if and only if $|f(x_o)| < +\infty$ and f is both upper and lower semicontinuous at x_o.*
(iv) *The function f is lower semicontinuous if and only the set $\{x \in E : f(x) > a\}$ is open, for all finite a.*
(v) *If f is lower semicontinuous then f is measurable.*

EXAMPLE 4.3. Let $f : \mathbb{R} \to \mathbb{C}$ with

$$f(x) := \begin{cases} 0, & \text{if} \quad x < 0 \\ -1, & \text{if} \quad x \geq 0. \end{cases}$$

Then f is lower semicontinuous.

Now define $g : \mathbb{R} \to \mathbb{C}$ with

$$g(x) := \begin{cases} 0, & \text{if} \quad x < 0 \\ 1, & \text{if} \quad x \geq 0. \end{cases}$$

Then g is upper semicontinuous. Finally, consider the *Dirchlet function*

$$D(x) := \begin{cases} 1, & \text{if} \quad x \in \mathbb{Q} \\ 0, & \text{if} \quad x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Then D is upper semicontinuous at the rational numbers and lower semicontinuous at the irrational numbers.

As a consequence of the previous definitions we get the following.

PROPOSITION 4.4. *Let Mf denote the Hardy-Littlewood maximal function of a function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then Mf is lower semicontinuous at every $x \in \mathbb{R}^n$. In particular Mf is measurable and the sets $\{x \in \mathbb{R}^n : Mf(x) > \lambda\}$ are open for all finite $\lambda > 0$.*

Although maximal operators are interesting in their own right, there are some very specific applications we have in mind. The first has to do with pointwise convergence of averages of a function and is a consequence of the following simple proposition.

PROPOSITION 4.5. *Let $\{T_t\}_{t>0}$ be a family of sub-linear operators on $L^p(X, \mu)$ and define the maximal operator*

$$T^*(f)(x) = \sup_{t>0} |T_t(f)(x)|.$$

If T^* is of weak type (p, q) then for any $t_o > 0$ the set

$$\{f \in L^p(X, \mu) : \lim_{t \to t_o} T_t f(x) = f(x) \ \textit{a.e.}\}$$

is closed in $L^p(X, \mu)$

PROOF. In order to show that the set

$$E_{T^*} := \{f \in L^p(X, \mu) : \lim_{t \to t_o} T_t f(x) = f(x) \ \text{a.e.}\}$$

is closed, consider a sequence of functions $\{f_n\} \subset E_{T^*}$ with $f_n \to f$ in $L^p(\mathbb{R}^n)$. We need to show that $f \in E_{T^*}$. To see this observe that for almost every $x \in \mathbb{R}^n$ we have

$$\begin{aligned} \limsup_{t \to t_o} |T_t f(x) - f(x)| &\leq |T_t(f - f_n)(x) - (f - f_n)(x)| \\ &\leq \sup_{t>0} |T_t(f - f_n)(x)| + |(f - f_n)(x)| \\ &= T^*(f - f_n)(x) + |(f - f_n)(x)|. \end{aligned}$$

Thus for any $\lambda > 0$ we can write

$$\begin{aligned} &\mu(\{x \in X : \limsup_{t \to t_o} |T_t f(x) - f(x)| > \lambda\}) \\ &\leq \mu(\{x \in X : T^*(f - f_n)(x) > \lambda/2\}) + \mu(\{x \in X : |(f - f_n)(x)| > \lambda/2\}) \\ &\lesssim_{T^*} \frac{\|f - f_n\|_p^q}{\lambda^q} + \frac{\|f_n - f\|_p^p}{\lambda^p}. \end{aligned}$$

Since the right hand side tends to 0 as $n \to \infty$ and the left hand side does not depend on n we conclude that for every $\lambda > 0$

$$\mu(\{x \in X : \limsup_{t \to t_o} |T_t f(x) - f(x)| > \lambda\}) = 0.$$

Now we have that

$$\begin{aligned}&\mu(\{x \in X : \limsup_{t \to t_o} |T_t f(x) - f(x)| > 0\}) \\ &\leq \sum_{k=1}^{\infty} \mu\Big(\Big\{x \in X : \limsup_{t \to t_o} |T_t f(x) - f(x)| > \frac{1}{k}\Big\}\Big) = 0.\end{aligned}$$

Thus $\lim_{t \to t_o} T_t(f)(x) = f(x)$ for almost every $x \in \mathbb{R}^n$ so that $f \in E_{T^*}$. □

REMARK 4.6. We have indexed the family T_t in $t \in \mathbb{R}_+$ for the sake of definitiveness but one can of course consider more general index sets and the previous proposition remains valid. In every case that the index set is uncountable some attention should be given in assuring the measurability of $T_t^*(f)$.

REMARK 4.7. To get a clearer picture of what this proposition says consider the family of operators

$$T_t(f)(x) = (f * \phi_t)(x),$$

for some $\phi \in L^1(\mathbb{R}^n)$ with integral $\int \phi = 1$. As we have seen already many times, these averages of f converge to f in many different senses for different classes of functions f. In particular if $f \in C_c^\infty(\mathbb{R}^n)$ then $f * \phi_t$ converges to f even uniformly as $t \to 0$. Thus we have

$$C_c^\infty(\mathbb{R}^n) \subset \{f \in L^p(X, \mu) : \lim_{t \to 0} T_t f(x) = f(x) \ \text{ a.e.}\}.$$

Since $C_c^\infty(\mathbb{R}^n)$ is dense in L^p, Proposition 4.5 implies that if T^* is of weak type (p, q) then

$$\lim_{t \to 0} (f * \phi_t)(x) = f(x),$$

for almost every $x \in \mathbb{R}^n$. Thus in order to show that approximations to the identity converge to the function almost everywhere, for all $f \in L^p$ it is enough to show that the corresponding maximal operator is of weak type (p, q), for some $q \geq 1$. In what follows we will show that the Hardy-Littlewood maximal operator is of weak type $(1, 1)$ and this already implies the corresponding statement for a wide class of approximations to the identity.

To avoid confusion, remember that in Theorem 3.26 we have already exhibited that

$$\lim_{t \to 0} (f * \phi_t)(x) = f(x)$$

for every Lebesgue point x of f. However this is only interesting if we already know that f has “many” Lebesgue points (in particular almost every point in $\mathbb{R}^n$). In Theorem 15 of Chapter we took for granted that the integral of a locally integrable function is almost everywhere differentiable and this in turn implied that almost every point in $\mathbb{R}^n$ is a Lebesgue point of f. In this part of the course we will fill in this gap by showing that the integral of a locally integrable function is almost everywhere differentiable.

EXERCISE 4.8. Let $T^*(f)(x) := \sup_{t>0}|T_t f(x)|$ be of weak type (p,q). Show that for every $t_o > 0$ the set

$$\{f \in L^p(X,\mu) : \lim_{t\to t_o} T_t f(x) \text{ exists a.e. }\}$$

is closed in $L^p(X,\mu)$.

Hint: The proof is very similar to that of Proposition 4.5. Observe that it suffices to show that

$$\mu(\{x \in X : \limsup_{t\to t_o} T_t f(x) - \liminf_{t\to t_o} f(x) > \lambda\}) = 0,$$

for every $\lambda > 0$.

4.2. The Hardy-Littlewood maximal theorem

We focus our attention to the Hardy-Littlewood maximal operator; for $f \in L^1_{\text{loc}}$ we have defined

$$M(f)(x) = \sup_{r>0} \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|dy, \quad x \in \mathbb{R}^n.$$

The discussion in the previous section suggests that one should try to prove weak (p,q) bounds for the operator M. In fact we will prove the following theorem which summarizes the boundedness properties of M.

THEOREM 4.9 (Hardy-Littlewood maximal theorem). *Let M denote the Hardy-Littlewood maximal operator as above.*

(ii) *The Hardy-Littlewood maximal operator if of weak type $(1,1)$:*

$$|\{x \in \mathbb{R}^n : M(f)(x) > \lambda\}| \lesssim_n \frac{\|f\|_{L^1(\mathbb{R}^n)}}{\lambda}, \quad \lambda > 0,$$

for all $f \in L^1(\mathbb{R}^n)$.

(i) *The Hardy-Littlewood maximal operator is of strong type (p,p); for $1 \leq p < \infty$ we have:*

$$\|M(f)\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} \|f\|_{L^p(\mathbb{R}^n)},$$

for all $1 < p \leq \infty$ and $f \in L^p(\mathbb{R}^n)$.

REMARK 4.10. The Hardy-Littlewood maximal operator is *not* of strong type $(1,1)$. To see this note that for any $f \in L^1(\mathbb{R}^n)$ we have that

$$M(f)(x) \gtrsim_f \frac{1}{|x|^n}, \quad |x| \to \infty,$$

which shows in particular that $M(f)$ is never integrable whenever $f \in L^1(\mathbb{R}^n)$ is not identically 0. Moreover, no strong estimates of type (p,q) are possible whenever $p \neq q$ as can be seen by examining the dilations of f and Mf.

EXERCISE 4.11. Prove the assertions in the previous remark.

EXERCISE 4.12. Let $f \in L^1(\mathbb{R}^n)$ and let B be a ball such that $M(f)(x) > \lambda$ for every $x \in B$. Let B^* be the ball with the same center and twice the radius of B. Show that $M(f)(x) \gtrsim_n \lambda$ for every $x \in B^*$.

PROOF OF THEOREM 4.9. First of all let us observe that M is of strong type (∞,∞). This is just a consequence of the general fact that an average never exceeds a maximum. In view of the Marcinkiewicz interpolation theorem it then suffices to show the assertion (i) of the theorem, namely that M is of weak type $(1,1)$.

A further reduction can be made by observing that, by homogeneity, it suffices to show that

$$|\{x\in\mathbb{R}^n : Mf(x)>1\}|\lesssim_n \|f\|_{L^1(\mathbb{R}^n)}.$$

We now fix some $f\in L^1(\mathbb{R}^n)$ and set

$$E:=\{x\in\mathbb{R}^n : |Mf(x)|>1\},$$

and let $K\subset E$ be any compact subset of E. Our task is to obtain an estimate of the form $|K|\lesssim_n \|f\|_{L^1(\mathbb{R}^n)}$, uniformly in $K\subset E$.

For every $x\in K$ there is a ball $B_x=B(x,r(x))$ such that

$$\int_{B_x}|f(y)|dy>|B_x|.$$

The family $\{B_x\}_{x\in K}$ clearly covers the compact set K so we can extract a finite subcollection of balls $\{B_m\}_{m=1}^N$ which still covers K. Since $K\subset\cup_{m=1}^N B_m$ we get that

$$|K|\le\sum_{m=1}^N|B_m|<\sum_{m=1}^N\int_{B_m}|f(y)|dy=\int|f(y)|\sum_{m=1}^N\mathbf{1}_{B_m}(y)dy.$$

Observe on the other hand that

$$\int|f(y)|\mathbf{1}_{\cup_m B_m}(y)dy\le\|f\|_{L^1(\mathbb{R}^n)},$$

so if we managed to show that

$$\sum_{m=1}^N\mathbf{1}_{B_m}(y)\lesssim_n\mathbf{1}_{\cup_m B_m}(y)$$

almost everywhere, we would be done. The main obstruction to such an estimate is that the balls B_m may overlap a lot. On the other hand, if the balls B_m were disjoint (or "almost disjoint") then there would be no problem. Although we cannot directly claim that the family $\{B_m\}$ is non-overlapping, the following lemma will allow us to extract a subcollection of balls which has this property, without losing too much of the measure of the union of the balls in the collection.

LEMMA 4.13 (Vitali covering lemma). *Let $B_1,\ldots,B_N$ be a finite collection of balls. Then there exists a subcollection $B_{n_1},\ldots,B_{n_M}$ of* disjoint balls *such that*

$$\sum_{j=1}^M|B_{n_j}|=|\cup_{j=1}^M B_{n_j}|\ge 3^{-n}|\cup_{i=1}^N B_i|.$$

Before giving the proof of this covering lemma let us see how we can use it to conclude the proof of Theorem 4.9. Recall that we have extracted a finite collection of balls $\{B_m\}_{m=1}^N$ which cover the set K and which satisfy

$$\int_{B_m}|f(x)|dx>|B_m|,\quad m=1,2,\ldots,N.$$

Now applying the covering lemma we can extract a subcollection of disjoint balls $\{B_{m_j}\}_{j=1}^M$ so that the measure of their union exceeds a multiple of the measure of the union of the original family of balls. Thus, we can write

$$\begin{aligned}|K| &\le |\cup_{m=1}^N B_m| \le 3^n|\cup_{j=1}^M B_{m_j}| = 3^n \sum_{j=1}^M |B_{m_j}| \\ &< 3^n \sum_{j=1}^M \int_{B_{m_j}} |f(y)|dy = 3^n \int |f(y)| \sum_{j=1}^M \mathbf{1}_{B_{m_j}}(y)dy \\ &= 3^n \int |f(y)|\mathbf{1}_{\cup_{j=1}^M B_{m_j}}(y)dy \le 3^n \int_{\mathbb{R}^n} |f(x)|dx = 3^n\|f\|_{L^1(\mathbb{R}^n)}.\end{aligned}$$

Observe that this estimate is uniform over all compact sets $K \subset E$ so taking the supremum over such sets and using the inner regularity of the Lebesgue measure we conclude that

$$|E| \le 3^n\|f\|_{L^1(\mathbb{R}^n)},$$

which concludes the proof. □

We now give the proof of the Vitali covering lemma which was in the heart of the previous argument.

Proof of the Vitali covering Lemma 4.13. First of all let us assume that the balls $B_1, \ldots, B_N$ are numbered so that their size is decreasing (thus B_1 is the largest ball). We will choose the subcollection $B_{m_1}, \ldots, B_{m_M}$ based on the *greedy principle*. The first ball we choose in the subcollection is the largest ball, thus $B_{m_1} := B_1$. Now assume we have chosen the balls $B_{m_1}, B_{m_2}, \ldots, B_{m_i}$ for some $i \ge 1$. We choose the ball $B_{m_{i+1}}$ to be the largest ball which doesn't intersect any of the balls already chosen. Observe that this amounts to choosing

$$\begin{aligned}m_{i+1} &:= \max\{j : 1 \le j \le N, B_j \cap B_{m_\ell} = \emptyset \quad \text{for all} \quad \ell = 1, 2, \ldots, i\} \\ &= \max\{j : m_i < j \le N, B_j \cap B_{m_\ell} = \emptyset \quad \text{for all} \quad \ell = 1, 2, \ldots, i\}.\end{aligned}$$

Since the original collection was finite the selection process will end in a finite number of M steps. It is clear that the resulting subcollection $\{B_{m_j}\}_j$ consists of disjoint balls. On the other hand, every ball B of the original collection is either selected or it intersects one of the selected balls, say B_{m_ℓ}, of greater or equal radius. Indeed, if B is not selected then by the selection process there is at least one $k \in \{1, 2, \ldots, M\}$ such that $B \cap B_{m_k} \neq \emptyset$. Let $\ell := \min\{k : B \cap B_{m_k} \neq \emptyset\}$. Then necessarily the radius of B_{m_ℓ} is greater or equal than the radius of B. To see this observe that B does not intersect any of the balls $B_{m_1}, \ldots, B_{m_{\ell-1}}$. Since B_{m_ℓ} was selected instead of B it means that B_{m_ℓ} had a greater or equal radius.

We can then conclude that $B \subset B^*_{m_\ell}$, where $B^*_{m_\ell}$ is the ball with the same center as B_{m_ℓ} and three times its radius. Thus we have that

$$B_1 \cup \cdots \cup B_N \subset B^*_{m_1} \cup \cdots \cup B^*_{m_M}.$$

Taking the Lebesgue measure of both unions we conclude

$$|B_1 \cup \cdots \cup B_N| \le 3^n|B_{n_1} \cup \cdots B_{n_M}|,$$

as desired. □

EXERCISE 4.14 (The maximal function on the class $L \log L$). We saw that if f is a non-trivial integrable function then $\mathcal{M}(f)$ is never integrable. Suppose however that f is supported in a finite ball $B \subset \mathbb{R}^n$ and that it is a "bit better" than being integrable, namely it satisfies

$$\|f\|_{L\log L(B)} := \int_B |f(x)|(1 + \log^+ |f(x)|)dx < +\infty.$$

where $\log^+ x = \max(\log x, 0)$. We say in this case that $f \in L \log L(B)$. Then we have that $M(f) \in L^1(B)$ and

$$\|Mf\|_{L^1(B)} \lesssim |B| + \|f\|_{L\log L(B)}.$$

Hints:

(a) For $\lambda > 0$ show that

$$|\{x \in B : M(f)(x) > 2\lambda\}| \lesssim \frac{1}{\lambda} \int_{\{x\in B:|f(x)|>\lambda\}} |f(x)|.$$

It will help you to split the function f as

$$f = f\mathbf{1}_{\{|f|>\lambda\}} + f\mathbf{1}_{\{|f|<\lambda\}} =: f_2 + f_1,$$

and observe that $\|M(f_1)\|_{L^\infty(B)} < \lambda$.

(b) Show that

$$\int_B M(f)(x)dx \le 2|B| + 2\int_1^\infty |\{x \in B : M(f)(x) > 2\lambda\}|d\lambda.$$

From this, (a) and Fubini's theorem you can conclude the proof.

4.3. Consequences of the maximal theorem

Our first application of the maximal theorem has to do with the differentiability of the integral of a locally integrable function. Indeed, using Theorem 4.9 and Proposition 4.5 we immediately get the following.

COROLLARY 4.15 (Lebesgue differentiation theorem). *Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ be a locally integrable function. Then, for almost every $x \in \mathbb{R}^n$ we have that*

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)dy = f(x).$$

For the proof just observe that $|A_t(f)(x)| \le M(f)(x)$ and that the claimed convergence property is a local property thus one can restrict any locally integrable function to a ball around the point x which turns f into an L^1 function. As we have already seen in Corollary 3.25, the previous statement also implies the following:

COROLLARY 4.16. *Let $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then almost every point in $\mathbb{R}^n$ is a* Lebesgue point *if f, that is, we have that*

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(x) - f(y)|dy = 0,$$

for almost every $x \in \mathbb{R}^n$.

Lebesgue's differentiation theorem generalizes to more general averages. A manifestation of this is already presented in Theorem 3.26 which asserts that for "nice" approximations to the identity ϕ, the means $f * \phi_t$ converge to f at every Lebesgue point of f. Here we will give an alternative proof of this theorem by controlling the maximal operator $\sup_{t>0} f * \phi_t$ by the Hardy-Littlewood maximal function.

PROPOSITION 4.17. *Let $\phi \in L^1(\mathbb{R}^n)$ be a positive and radially decreasing function with $\int_{\mathbb{R}^n} \phi(x)dx = 1$. Then we have that*

$$\sup_{t>0}(f * \phi_t)(x) \leq \|\phi\|_{L^1(\mathbb{R}^n)} M(f)(x).$$

PROOF. First suppose that ϕ is of the form $\phi(x) = \sum_{j=1}^N a_j \chi_{B_j}$ where $a_J > 0$ and B_j are Euclidean balls centered at 0 for all $j = 1, 2, \ldots, N$. Then we have

$$\begin{aligned}
\phi * f(x) &= \sum_{j=1}^N a_j (f * \chi_{B_j})(x) = \sum_{j=1}^N a_j |B_j| \frac{1}{|B_j|}(f * \chi_{B_j})(x) \\
&\leq \sum_{j=1}^N a_j |B_j| \, M(f)(x) = \int_{\mathbb{R}^n} \phi(x)dx \, M(f)(x) \\
&= \|\phi\|_{L^1(\mathbb{R}^n)} M(f)(x).
\end{aligned}$$

However, any function ϕ which is positive and radially decreasing can be approximated monotonically from below by a sequence of simple functions of the form $\sum a_j \chi_{B_j}$ so we are done. □

As an immediate corollary we get the same control for approximations to the identity which are controlled by positive radially decreasing functions. Observe that this gives an alternative proof of Theorem 3.26.

COROLLARY 4.18. *Let $|\phi(x)| \leq \psi(x)$ almost everywhere where $\psi(x)$ is positive, radially decreasing and integrable. Then we have that*

$$T^*(f)(x) := \sup_{t>0}(f * \phi_t)(x) \leq \int_{\mathbb{R}^n} \psi(y)dy \, M(f)(x).$$

In particular T^* is of weak type $(1,1)$ and strong type (p,p) for all $1 < p \leq \infty$. We conclude that

$$\lim_{t \to 0}(f * \phi_t)(x) = \int_{\mathbb{R}^n} \phi(y)dy \, f(x),$$

for almost every $x \in \mathbb{R}^n$.

REMARK 4.19. The qualitative conclusion of the previous corollaries is that maximal averages of f with radially decreasing integrable kernels are controlled by the Hardy-Littlewood maximal function. A typical radially decreasing integrable kernel is the Gaussian kernel

$$W(x) = e^{-\pi|x|^2}.$$

By dilating W by $\sqrt{2\pi t}$ we get

$$W_t(x) = \frac{1}{(2\pi t)^{\frac{n}{2}}} e^{-\frac{|x|^2}{4t}}.$$

The function $e^{-|x|^2/4t}$ can be viewed as smooth approximation of the indicator function of a ball of radius $\sim \sqrt{t}$ (up to constants). Indeed, for $|x| < \sqrt{t}$ say, we have that $e^{-|x|^2/4t} \simeq 1$, while for $|x| \gtrsim \sqrt{t}$ the function $e^{-|x|^2/4t}$ decays very fast. Thus the kernel W_t is not so different from $\chi_{\sqrt{t}} = t^{-\frac{n}{2}}\mathbf{1}_{B(0,\sqrt{t})}$.

4.3.1. Points of density and the Marcinkiewicz Integral. A direct consequence of Lebesgue's differentiation theorem is that almost every point of a measurable set is "completely" surrounded by other points of the set. To make this precise, let us give a definition.

DEFINITION 4.20. Let E be be a measurable set in $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$. We say that x is a *point of density* of the set E, if

$$\lim_{r\to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = 1.$$

Of course the limit in the previous definition might not exist in general or not be equal to 1. Observe however that if the previous limit is equal to 0 then x is a point of density of the set E^c, the complement of E. On the other hand, applying Lebesgue's differentiation theorem to the function $\mathbf{1}_E$ which is obviously locally integrable we get

$$\lim_{r\to 0} \frac{1}{|B(x,r)|}\int_{B(x,r)} \mathbf{1}_E(y)dy = \lim_{r\to 0} \frac{|E\cap B(x,r)|}{|B(x,r)|} = \mathbf{1}_E(x),$$

for almost every $x \in \mathbb{R}^n$. Thus we immediately get the following

PROPOSITION 4.21. *Let $E \subset \mathbb{R}^n$ be a measurable set. Then almost every point of E is a point of density of E. Likewise, almost every point $x \in E^c$ is a point of density of E^c.*

Thus a point of density is in a measure theoretic sense completely surrounded by other points of E. The measure of the set E in the ball $B(x,r)$ is proportional to the measure of the ball as $r \to 0$ and x is a point of density.

Another way to describe this notion is the following. Let F be a closed set and define $\delta(x) = \operatorname{dist}(x,F)$. Of course $\delta(x) = 0$ if $x \in F$. Now think of y in a neighborhood of zero so that the vector $x + y$ is in the neighborhood of x. If $x \in F$ then the distance of the point $x + y$ from F is at most $|y|$ since $x \in F$ and $|(x + y) - x| = |y|$. Thus we have that $\delta(x + y) \leq |y|$ whenever $x \in F$. That is, when the point $x + y$ approaches $x \in F$, the distance $\delta(x + y)$, that is the distance of $x + y$ from F approaches zero. In fact the estimate above can be improved.

PROPOSITION 4.22. *Let F be a closed set. Then for almost every $x \in F$, $\delta(x + y) = o(|y|)$ as $|y| \to 0$. This is true in particular if x is a point of density of the set F.*

EXERCISE 4.23. Prove Proposition 4.22 above. The $o(|y|)$ is interpreted as follows: For every $\epsilon > 0$ there exists some $\delta > 0$ such that $\delta(x+y) \leq \epsilon|y|$ whenever $|y| \leq \delta$.

We will be mostly interested in another instance of this principle that is reflected in the Marcinkiewicz integral. This will also come in handy in our study of oscillatory integrals in the next chapter.

For F a closed set as before we define the *Marcinkiewicz integral associated to* F, $I(x)$, as

$$I(x) := \int_{|y|\leq 1} \frac{\delta(x+y)}{|y|^{n+1}}, \quad x \in \mathbb{R}^n.$$

THEOREM 4.24. *Let I be the Marcinkiewicz integral as above.*

(i) *If $x \in F^c$ then $I(x) = +\infty$.*
(ii) *For almost every $x \in F$ we have that $I(x) < +\infty$.*

REMARK 4.25. The previous theorem shows that, in average, $\delta(x+y)$ is small enough whenever $x \in F$ to make the integral converge locally. This can be seen as a variation of Proposition 4.22 though no direct quantitative connection is claimed.

Part (i) is obvious and is left as an exercise. For (ii) it will be enough to show the following:

LEMMA 4.26. *Let F be a closed set whose complement F^c has finite measure. Then we set*

$$I_*(x) = \int_{\mathbb{R}^n} \frac{\delta(x+y)}{|y|^{n+1}}.$$

Then $I_*(x) < +\infty$ for almost every $x \in F$. In particular we have

$$\int_F I_*(x) \lesssim_n |F^c|.$$

PROOF. It is enough to show

$$\int_F I_*(x) \lesssim |F^c|,$$

since then $I*(x)$ is finite for almost every $x \in F$. To that end we write

$$\begin{aligned}\int_F I_*(x)dx &= \int_F \int_{\mathbb{R}^n} \frac{\delta(x+y)}{|y|^{n+1}} dy\, dx = \int_F \int_{\mathbb{R}^n} \frac{\delta(y)}{(x+y)^{n+1}} dy\, dx \\ &= \int_F \int_{F^c} \frac{\delta(y)}{|x-y|^{n+1}} dy\, dx = \int_{F^c} \Big(\int_F \frac{1}{|x-y|^{n+1}} dx \Big) \delta(y) dy.\end{aligned}$$

Now fix a $y \in F^c$. As $x \in F$ we obviously have that $|x-y| \geq \delta(y)$ thus $F \subset \{x \in \mathbb{R}^n : |x-y| \geq \delta(y)\}$. Since all the quantities under the integral signs are positive the previous estimate implies

$$\int_F \frac{1}{|x-y|^{n+1}} dy \leq \int_{\{x\in\mathbb{R}^n : |x-y|\geq\delta(y)\}} \frac{1}{|x|^{n+1}} \lesssim_n \frac{1}{\delta(y)},$$

whenever $y \in F^c$. Integrating for $y \in F$ we get

$$\int_F I_*(x)dx \lesssim_n \int_{F^c} \delta(y)\delta(y)^{-1} dy \leq |F^c|$$

which is the desired estimate. $\square$

To get the proof of Theorem 4.24 we now use the previous lemma as follows. Let F be a closed set and let B_m be a ball of radius m centered at 0. Let $F_m =$

$F \cup B_m^{\mathsf{c}}$. Then F_m is closed and $F_m^{\mathsf{c}} \subset B_m$ so that $|F_m^{\mathsf{c}}| < \infty$. Thus the previous lemma applies to F_m and we get that

$$\int_{|y|\leq 1} \frac{\delta_m(x+y)}{|y|^{n+1}} dy < +\infty,$$

for almost every $x \in F_m$ where we denote by δ_m the distance from the set F_m. Now observe that for $x \in F \cap B_{m-2}$ and $|y| \leq 1$ we have that $\delta_m(x+y) = \delta(x+y)$; indeed $\delta_m(x+y) \leq |y| \leq 1$ and $|x+y| \leq m-1$ thus $\operatorname{dist} x+y, B_m^{\mathsf{c}} \geq 1$. We conclude that

$$\int_{|y|\leq 1} \frac{\delta(x+y)}{|y|^{n+1}} dy < \infty,$$

for almost every $x \in F \cap B_{m-2}$. Since every $x \in \mathbb{R}^n$ eventually belongs to some B_{m-2} for some large m we get the conclusion of the theorem.

EXERCISE 4.27. This exercise concerns strengthened estimates for the Marcinkiewicz integral.

(i) Show the following strengthened form of Lemma 4.26: For $\psi \geq 0$ and locally integrable then

$$\int_F I_*(x)\psi(x)dx \leq \int_{F^{\mathsf{c}}} (M\psi)(x)dx,$$

whenever F is closed and $|F^{\mathsf{c}}| < +\infty$.

(ii) Use (i) and the maximal theorem to conclude that $I_*(x) \in L^p(F)$ for all $1 \leq p < \infty$.

4.4. Variants of the Hardy-Littlewood maximal function

4.4.1. The dyadic maximal function. We now come to a different approach to the maximal function theorem. On the one hand the "dyadic" approach we will follow here already implies the maximal theorem presented in the previous paragraph. It is however interesting in its own right and it will give us the chance to present a dyadic structure on the Euclidean space which will come in handy in many different cases.

Consider the basic cube $Q_{0,0} = [0,1)^n \subset \mathbb{R}^n$. A dyadic dilation of this cube is the cube $Q_{m,0} := 2^m Q_{0,0} := [0, 2^m)^n$ where $m \in \mathbb{Z}$. Now we also consider integer translations of this cube of the form $Q_{m,k} := k + Q_{m,0}$ for some integer vector $k \in \mathbb{Z}^n$. We have the following definition:

DEFINITION 4.28. A dyadic cube of *generation* m is a cube of the form

$$Q_{m,k} = 2^m(k + [0,1)^n) = \{2^m(k+x) : x \in [0,1)^n\},$$

where $m \in \mathbb{Z}$ and $k \in \mathbb{Z}^n$. The family of disjoint cubes

$$\mathcal{D}_m := \{Q_{m,k}\}_{k\in\mathbb{Z}}$$

defines the m-th generation of dyadic cubes. We denote by $\mathcal{D}$ the collection of all dyadic cubes in $\mathbb{R}^n$.

The dyadic cubes have the following basic properties.

(d1) The dyadic cubes in the generation m are disjoint and their union is $\mathbb{R}^n$. Thus any point $x \in \mathbb{R}^n$ belongs to unique dyadic cube in the m-th generation.

(d2) Two (different) dyadic cubes are either disjoint or one contains the other.

(d3) A dyadic cube in $\mathcal{D}_m$ consists of exactly 2^n dyadic cubes of the generation $\mathcal{D}_{m-1}$. On the other hand, for any dyadic cube $Q \in \mathcal{D}_m$ and any $j > m$ there is a unique dyadic cube in the collection $\mathcal{D}_j$ that contains Q.

As a first instance of how things simplify and get sharper in the dyadic world, let us see the analogue of the Vitali covering lemma in the dyadic case.

LEMMA 4.29 (Dyadic Vitali covering lemma). *Let $Q_1, \ldots, Q_N$ be a finite collection of dyadic cubes. There exists a subcollection $Q_{m_1}, \ldots, Q_{m_M}$ of disjoint dyadic cubes such that*

$$Q_1 \cup \cdots \cup Q_N = Q_{m_1} \cup \cdots \cup Q_{m_M}.$$

PROOF. Let Q_{m_i} be the maximal cubes among $Q_1, \ldots, Q_N$, that is, the cubes that are not contained in any other cube of the collection $Q_1, \ldots, Q_N$. Then the cubes $\{Q_{m_j}\}_{j=1}^M$ are disjoint (otherwise they wouldn't be maximal). Also any cube that is not maximal is contained in the union $Q_{m_1} \cup \cdots \cup Q_{m_M}$. □

Given a function $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$ we set

$$\mathbb{E}_m f(x) := \sum_{Q \in \mathcal{D}_m} \Big(\frac{1}{|Q|} \int_Q f \Big) \mathbf{1}_Q(x).$$

Observe that given x there is a unique cube $Q_x \in \mathcal{D}_m$ that contains x and then the value of $\mathbb{E}_m f$ at x equals the average of the function f over the cube Q_x. In fact, $\mathbb{E}_m f$ is the *conditional expectation* of f with respect to the σ-algebra generated by the family $\mathcal{D}_m$. Observe that for every generation m, if Ω is a union of cubes in $\mathcal{D}_m$ then

$$\int_\Omega \mathbb{E}_m f = \int_\Omega f.$$

The operator $\mathbb{E}_M$ is the discrete dyadic analogue of an approximation to the identity dilated at scale 2^k. A difference however is that the averages here are not "centered". Indeed, $\mathbb{E}_m f(x)$ is the average of f with respect to the cube Q whenever $x \in Q$ for some $Q \in \mathcal{D}_m$. However x is not the center of the cube Q.

The *dyadic maximal function* is defined as

$$M_{\mathcal{D}}(f)(x) := \sup_{m \in \mathbb{Z}} \mathbb{E}_m |f|(x) = \sup_{\substack{Q \in \mathcal{D} \\ Q \ni x}} \frac{1}{|Q|} \int_Q |f(y)| dy.$$

Thus the supremum is taken over all dyadic cubes that contain x or, equivalently, over all generations of dyadic cubes. We have the analogue of the maximal theorem:

THEOREM 4.30 (Dyadic Maximal Theorem). *Let $M_{\mathcal{D}}$ denote the dyadic maximal function.*

(i) *The dyadic maximal function is of weak type* $(1,1)$ *with weak type norm at most* 1*:*

$$|\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\}| \leq \frac{\|f\|_{L^1(\mathbb{R}^n)}}{\lambda},$$

for all $f \in L^1(\mathbb{R}^n)$.

(ii) *The dyadic maximal function is of strong type* (p,p), *for all* $1 < p \leq \infty$; *for all* $f \in L^p(\mathbb{R}^n)$ *we have*

$$\|M_{\mathcal{D}}(f)\|_{L^p(\mathbb{R}^n)} \lesssim_p \|f\|_{L^p(\mathbb{R}^n)},$$

where the implied constant depends only on p.

(iii) *We conclude using Proposition 4.5 that for every* $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ *we have that*

$$\lim_{m\to-\infty} \mathbb{E}_m(f)(x) = f(x) \quad \textit{for a.e.} \quad x \in \mathbb{R}^n.$$

EXERCISE 4.31. Show the pointwise estimate

$$M_{\mathcal{D}}(f)(x) \lesssim_n M(f)(x)$$

where the implied constant depends only on the dimension n. On the other hand, show that the opposite estimate cannot be true. For example when $n = 1$ test against the function $\mathbf{1}_{[0,1)}$. Conclude that the dyadic maximal theorem follows from the non-dyadic one (with a different constant though).

Hint: Observe that if $x \in Q$ and Q is a dyadic cube, there exists a ball $B(x,r)$ which contains Q and $|B(x,r)| \simeq_n |Q|$.

EXERCISE 4.32. Give the proof of Theorem 4.30 above. Observe that the proof is essentially identical to that of Theorem 4.9 using the dyadic version of the Vitali covering Lemma instead of the non-dyadic one. For (ii) you need to observe that the statement is true for continuous functions (for example) and use Proposition 4.5.

4.4.2. The maximal function with respect to cubes. For $x \in \mathbb{R}^n$ and $r > 0$ let $Q(x,r)$ denote the cube of sidelength r, centered at $x \in \mathbb{R}^n$, that is $Q(x,r) := [x - \frac{r}{2}, x + \frac{r}{2})^n$. The maximal function with respect to cubes is

$$M_{\square}(f)(x) := \sup_{r>0} \frac{1}{r^n} \int_{[-\frac{r}{2},\frac{r}{2}]^n} |f(x-y)|dy = \sup_{r>0} (|f| * \psi_r)(x),$$

where ψ is the indicator function of the cube $[-\frac{1}{2},\frac{1}{2}]^n$. Thus $M_{\square}f$ is the maximal average of f with respect to Euclidean cubes. In the current setup of the Euclidean space equipped with the Lebesgue measure, the operators $M, M_{\square}$ are pointwise equivalent and thus they have exactly the same mapping properties, with norms comparable up to constants.

EXERCISE 4.33 (Pointwise equivalence of $M, M_{\square}$). Show that for every $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ we have

$$Mf(x) \simeq_n M_{\square}f(x),$$

with the implied constants depending only upon the dimension n.

4.4.3. The non-centered maximal function. Another common variant of the Hardy-Littlewood maximal function is the *non-centered* maximal function with respect to either Euclidean balls of cubes. For example we have

$$M'(f)(x) := \sup_{B\ni x} \frac{1}{|B|} \int_B |f(y)|dy,$$

where the supremum is taken over all Euclidean balls containing x. Likewise

$$M'_{\square}(f)(x) := \sup_{Q\ni x} \frac{1}{|Q|} \int_Q f(y)dy,$$

where the supremum is taken over all cubes (with sides parallel to the coordinate axes) that contain x. Again these maximal operators are pointwise equivalent with say $M(f)$ and thus they have they same mapping properties as M, given in Theorem 4.9.

EXERCISE 4.34. Consider the *non-centered maximal functions* $M', M'_\square$ and a $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Show the pointwise equivalences

$$M(f)(x) \simeq_n M'(f)(x) \simeq_n M'_\square(f)(x),$$

with the implied constants depending only upon the dimension n.

4.4.4. Maximal functions with respect to a non-negative Borel measure. Given a non-negative Borel measure μ on $\mathbb{R}^n$ and $f \in L^1_{\text{loc}}(\mathbb{R}^n, d\mu)$ we consider the averages of f over Euclidean balls

$$A_r^\mu f(x) := \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y)| d\mu(y),$$

with respect to the measure μ. The corresponding maximal operator with respect to μ is then

$$M^\mu f(x) := \sup_{r>0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y)| d\mu(y).$$

Completely analogously we can consider the maximal averages of f with respect to μ and the family of Euclidean cubes in $\mathbb{R}^n$ as

$$M_\square^\mu f(x) := \sup_{r>0} \frac{1}{\mu(Q(x,r))} \int_{Q(x,r)} |f(y)| d\mu(y).$$

It is essential to note here that these are *centered* maximal operators with respect to an essentially arbitrary measure in $\mathbb{R}^n$. The following theorem summarizes the mapping properties of these maximal functions.

THEOREM 4.35. *Let μ be a non-negative Borel measure on $\mathbb{R}^n$.*

(i) *The operators $M^\mu, M_\square^\mu$ map $L^1(\mathbb{R}^n, d\mu)$ to $L^{1,\infty}(\mathbb{R}^n, d\mu)$ with weak-type norms depending only on the dimension n.*
(ii) *For $1 < p \leq +\infty$ the operators $M^\mu, M_\square^\mu$ map $L^p(\mathbb{R}^n, d\mu)$ to $L^p(\mathbb{R}^n, d\mu)$ with norms depending only on n and p.*

The proof of this theorem depends on an especially strong covering lemma due to Besicovitch:

LEMMA 4.36 (Besicovitch covering lemma). *Let A be a bounded set in $\mathbb{R}^n$ and suppose that we are given a collection $\{B_x\}_{x\in A}$ of (open, half open, or closed) Euclidean balls, where each B_x is centered at x and has radius $r(x) > 0$. Then there exists a (possibly finite) sequence $\{B_j\}_j \subset \{B_x\}_{x\in A}$ such that*

(i) *$A \subseteq \cup_j B_j$.*
(ii) *We have the estimate $\left\|\sum_j \mathbf{1}_{B_j}\right\|_{L^\infty(\mathbb{R}^n)} \lesssim_n 1$ with the implied constant depending only on the dimension n.*
(iii) *The collection $\{B_j\}_j$ can be divided into at most $O(n)$ subcollections, each of which consists of pairwise disjoint balls.*

REMARK 4.37. A completely analogous statement is valid if one replaces Euclidean balls by Euclidean cubes in the statement of the Besicovitch covering lemma above. It is important to note however that the cubes, as well as the balls, are assumed to be *centered* in points of the set A.

Before giving the proof of the Besicovitch covering lemma let us see how we can use it in order to prove Theorem 4.35 above.

PROOF OF THEOREM 4.35. We will give the proof for Mf since the proof for $M_\square f$ follows by completely analogous arguments. Note however that for $M_\square f$ one needs to use the version of the Besicovitch covering lemma for centered Euclidean cubes instead of balls. Furthermore, by Marcinkiewicz interpolation it will be enough to show the weak $(1,1)$ bound for Mf, namely that

$$\mu(\{x \in \mathbb{R}^n : Mf(x) > \lambda\}) \lesssim_n \frac{\|f\|_{L^1(\mathbb{R}^n,d\mu)}}{\lambda}, \quad \lambda > 0.$$

By homogeneity we can assume that $\lambda = 1$. In order to have bounded sets, for $m \in \mathbb{N}$ we set

$$E_m := \Big\{x \in \mathbb{R}^n : |x| \leq m \text{ and } Mf(x) > 1\Big\}.$$

Now for $x \in E_m$ let Q_x be a cube centered at x and such that $\frac{1}{\mu(B_x)}\int_{B_x}|f(y)|d\mu(y)$. We apply the Besicovitch covering lemma to the family $\{B_x\}_{x\in E_m}$ producing a sequence $\{B_j^m\}_j$ with properties (i)-(iii), as in Lemma 4.36. Thus for every $m \in \mathbb{N}$ we get

$$\mu(E_m) \leq \sum_j \mu(B_j^m) \leq \int |f(y)|\Big(\sum_j \mathbf{1}_{B_j^m}\Big)d\mu(y) \lesssim_n \int |f(y)|d\mu(y),$$

with the first estimate following from (i) and the last estimate following from (ii) of Lemma 4.36. The conclusion now follows by letting $m \to +\infty$ and using the monotone convergence theorem. □

PROOF OF LEMMA 4.36. Let $B_x = B(x, r(x))$ for each $x \in A$. First suppose that $a_0 := \sup\{r(x) : x \in A\} = +\infty$. Since A is bounded a moment's reflection will allow us to pick a single ball (with large enough radius) that covers the whole set A. We can thus assume that $a_0 < +\infty$. We then choose $x_1 \in A$ such that $r(x_1) > \frac{3}{4}a_0$ and set $B_1 := B(x_1, r(x_1))$. Assuming we have chosen $B_1, \ldots, B_j$ let $a_j := \sup\{r(x) : x \in A \setminus \cup_{k\leq j}B_k\}$. We choose $x_{j+1} \in A \setminus \cup_{k\leq j}B_k$ so that $r(x_{j+1}) > \frac{3}{4}a_j$ and set $B_{j+1} := B(x_{j+1}, r(x_{j+1}))$. The balls selected above satisfy

$$B\Big(x_j, \frac{1}{3}r(x_j)\Big) \cap B\Big(x_k, \frac{1}{3}r(x_k)\Big) = \emptyset \quad \text{whenever} \quad j \neq k. \tag{4.1}$$

To see this observe that for $j < k$ (so B_j was chosen first) we have that $r(x_j) > \frac{3}{4}r(x_k)$. So if $x \in B(x_j, \frac{1}{3}r(x_j)) \cap B(x_k, \frac{1}{3}r(x_k))$ we would get that

$$|x_k - x_j| \leq |x_k - x| + |x_j - x| \leq \frac{1}{3}r(x_k) + \frac{1}{3}r(x_j) \leq \frac{7}{9}r(x_j) < r(x_j).$$

But then $x_k \in B(x_j, r(x_j))$ which is impossible by the selection process.

We have thus construct a sequence $\{B_j\}_j$ that can be finite or infinite. If $\{B_j\}_j$ is finite this means that the selection process terminated because $A \subset \cup_j B_j$. Thus $\{Q_j\}_j$ satisfies (i) in this case. If $\{B_j\}_j$ is infinite then necessarily $r(x_j) \to 0$ as $j \to +\infty$. Indeed, if not then we would have that $r(x_\tau) > \delta$ for some $\delta > 0$

and infinitely many τ's. However all the balls $B(x_\tau, \frac{1}{3}r(x_\tau))$ are contained in the bounded set $\{x \in \mathbb{R}^n : \operatorname{dist}(x, A) \leq a_0\}$ and by (4.1) they are disjoint which is clearly a contradiction. Thus $r(x_j) \to 0$ as $j \to +\infty$.

Assume now that there exists some $y \in A \setminus \cup_j B_j$. By the selection process we have for every k that $r(x_k) > \frac{3}{4}\sup\{r(x) : x \in A \setminus \cup_j B_j\} > r(y)$ which is impossible since $r(x_k) \to 0$. This proves (i) in the case that $\{B_j\}$ is infinite as well.

Let us now verify property (ii). For some fixed $k \geq 1$ let us consider the collection $\mathfrak{G}_k$ as

$$\mathfrak{G}_k := \{B_j : j < k,\, B_j \cap B_k \neq \emptyset\}.$$

We divide the collection $\mathfrak{G}_k$ into two subcollections $\mathfrak{G}_k^1, \mathfrak{G}_k^2$ as follows

$$\mathfrak{G}_k^1 := \{B = B_j(x_j, r(x_j)) \in \mathfrak{G}_k : r(x_j) \leq \frac{3}{4}Mr(x_k)\} = \{B_j\}_{j \in J_1},$$

$$\mathfrak{G}_k^2 := \{B = B_j(x_j, r(x_j)) \in \mathfrak{G}_k : r(x_j) > \frac{3}{4}Mr(x_k)\} = \{B_j\}_{j \in J_2},$$

where $M > 3$ is a positive integer to be chosen later in the proof. Let $|\mathfrak{G}_k^i|$, $i = 1, 2$, denote the number of balls in the collections $\mathfrak{G}_k^1$, $\mathfrak{G}_k^2$, respectively. We estimate the size of these two collection in two lemmas below.

LEMMA 4.38. *Let $\mathfrak{G}_k^1 = \{B_j(x_j, r(x_j)\}_{j \in J_1}$ be the first collection of balls from the Besicovitch covering lemma, as above. We have the estimate $|\mathfrak{G}_k^1| \leq 4^n(M+1)^n$.*

PROOF. As we have already shown in the proof of the Besicovitch lemma so far, the balls $B(x_j, \frac{1}{3}r(x_j))$ are disjoint. We also claim that these balls are contained in $B(x_k, (M+1)r(x_k)$. Indeed, since $B(x_k, r_k(x)) \cap B(x_j, r_j(x)) \neq \emptyset$ for all $j \in J_1$ we have for $j \in J_1$

$$|x_j - x_k| \leq r(x_j) + r(x_k) \leq (1 + \frac{3}{4}M)r(x_k).$$

Now let $x \in B(x_j, \frac{1}{3}r(x_j))$ for some $j \in J$. Remember that $r(x_j) \leq \frac{3}{4}Mr(x_k)$ for $j \in J_1$. We have

$$|x - x_k| \leq \frac{1}{3}r(x_j) + (1 + \frac{3}{4}M)r(x_k) \leq \frac{1}{3}\frac{3}{4}Mr(x_k) + (1 + \frac{3}{4}M)r(x_k) = (1 + M)r(x_k).$$

This proves the claim. We get

$$\sum_{j \in J_1} |B_j(x_j, \frac{1}{3}r(x_j)| \leq |B(x_k, (M+1)r(x_k)$$

and hence

$$\sum_{j \in J_1} \frac{r(x_j)^n}{3^n} \leq (M+1)^n r(x_k)^n.$$

Since $j < k$ it follows from the selection algorithm of the Besicovitch covering lemma that $r(x_j)/3 > \frac{1}{4}r(x_k)$. Thus

$$\frac{r(x_k)^n}{4^n}|\mathfrak{G}_k^1| \leq (M+1)^n r(x_k)^n$$

which obviously implies the bound in the statement of the lemma. ◻

We now move to the study of the collection $\mathfrak{G}_k^2$. For this let $j, j' \in J_2$ with $j' < j < k$ and consider the segments $\ell_{k,j}$ and $\ell_{k,j'}$ connecting x_k with x_j and x_k with $x_{j'}$, respectively. We next show that the angle between the segments $\ell_{k,j}$ and $\ell_{k,j'}$ is bounded away from zero. This will allow us to obtain an upper bound for $|\mathfrak{G}_k^2|$.

LEMMA 4.39. *Let $\theta := \{\text{angle between } \ell_{k,j} \text{ and } \ell_{k,j'}\}$. Then there is a choice of a positive integer M (that does not depend on any of the parameters of the problem) such that $\theta \geq \theta_0 = \arccos\frac{5}{6}$.*

PROOF. Let $j, j' \in J_2$ with $j' < j < k$ as above. By the Besicovitch selection algorithm we have $x_j \notin B(x_{j'}, r(x_{j'})$ so that

$$|x_j - x_{j'}| \geq r(x_{j'}).$$

Similarly we have that $x_k \notin B(x_j, r(x_j)) \cup B(x_{j'}, r(x_{j'}))$ thus

$$|x_k - x_{j'}| \geq r(x_{j'}) \quad \text{and} \quad |x_k - x_j| \geq r(x_j).$$

Now by the definition of the selection $\mathfrak{G}_k^2$ we have that both $B(x_j, r(x_j)), B(x_{j'}), r_(x_{j'})$ intersect $B(x_k, r(x_k))$ and $r(x_{j'}), r(x_j) > \frac{3}{4}Mr(x_k)$. Combining these facts with the observations above we can write

$$\frac{3}{4}Mr(x_k) < r(x_{j'}) \leq |x_{j'} - x_k| \leq r(x_{j'}) + r(x_k),$$
$$\frac{3}{4}Mr(x_k) < r(x_j) \leq |x_j - x_k| \leq r(x_j) + r(x_k).$$

Now we use the simple consequence of the parallelogram law that for two vectors $\vec{a}$ and $\vec{b}$ in the plane (in fact in any Hilbert space) we have

$$\vec{a} \cdot \vec{b} = \frac{|\vec{a}|^2 + |\vec{b}|^2 - |\vec{a} - \vec{b}|^2}{2}$$

Applying this to $\vec{a} := x_k - x_{j'}$ and $\vec{b} := x_k - x_j$ we get

$$\cos\theta = \frac{|x_k - x_{j'}|^2 + |x_k - x_j|^2 - |x_j - x_{j'}|^2}{2|x_k - x_{j'}||x_k - x_j|}$$

We can clearly assume that $\cos\theta > 0$ otherwise there is nothing to show. We estimate

$$\begin{aligned}
\cos\theta &\leq \frac{(r(x_{j'}) + r(x_k))^2 + (r(x_j) + r(x_k))^2 - r(x_{j'})^2}{2r(x_{j'})r(x_j)} \\
&\leq \frac{r(x_j)^2 + 2r(x_k)^2 + 2r(x_k)(r(x_{j'}) + r(x_k))^2}{2r(x_{j'})r(x_j)} \\
&\leq \frac{r(x_j)}{2r(x_{j'})} + \frac{r(x_k)}{r(x_j)}\frac{r(x_k)}{r(x_{j'})} + \frac{r(x_k)}{r(x_{j'})} + \frac{r(x_k)}{r(x_j)} \\
&\leq \frac{r(x_j)}{2r(x_{j'})} + \frac{4}{3M}\frac{4}{3M} + \frac{4}{3M} + \frac{4}{3M}.
\end{aligned}$$

Finally we remember that $j' < j$ which by the Besicovitch selection algorithm implies that $r(x_{j'}) > \frac{3}{4}r(x_j)$. Plugging this into the previous estimate we get

$$0 < \cos\theta \leq \frac{2}{3} + \frac{16}{9M^2} + \frac{8}{3M}.$$

Now it is clear that choosing M to be a large positive integer we get $\cos\theta \le \frac{5}{6}$. For example the value $M = 20$ will do. □

Having a lower bound on the angle between the angle of the segments $\ell_{k,j}$ and $\ell_{k,j'}$ easily implies an upper bound for $|\mathfrak{G}_k^2|$.

LEMMA 4.40. *Let* $\mathfrak{G}_k^2 = \{B_j(x_j, r(x_j))\}_{j\in J_2}$ *be the second collection of balls from the Besicovitch covering lemma. There exists a positive integer* M *(for example* $M = 20$*) such that* $|\mathfrak{G}_k^2| \le c_n$ *where* c_n *depends only on the dimension* n*.*

PROOF. Let us fix M large, say $M = 20$ so that $\theta \ge \arccos 5/6 =: \theta_0$ as in the previous lemma. To see how the proof of the lemma works let us first consider the two-dimensional case. The number of balls in the collection $\mathfrak{G}_k^2$ is bounded by the maximum number of rays connecting x_k with points in the plane such that, the minimum angle between any two rays is at least θ_0. There are at most $\theta_0/2\pi$ such rays so in this case we immediately get $|\mathfrak{G}_k^2| \lesssim 1$.

Now consider the general n-dimensional case with $n \ge 3$. Consider a cone $C(\theta_0$ in $\mathbb{R}^n$ with vertex at x_k and aperture θ_0. The number of balls in $\mathfrak{G}_k^2$ is bounded by the maximum number of disjoint such cones. This number can be estimated from above by

$$|\mathfrak{G}_k^2| \le \frac{\sigma_{n-1}(S^{n-1})}{\sigma_{n-1}(C(\theta_0) \cap S^{n-1})},$$

where S^{n-1} denotes the unit sphere in $\mathbb{R}^n$ and σ_{n-1} is the induced Lebesgue measure on S^{n-1}. What is important here is that the upper bound in the last estimate depends only upon dimension and the angle θ_0, which is an absolute number. □

It remains to verify property (iii) in the Besicovitch covering lemma. However this is just a consequence of (ii) and is left as an exercise. □

4.4.5. Maximal functions with respect to a doubling measure in $\mathbb{R}^n$. We now describe a setup that will allow us to extend the theory towards two different directions. For the first one we are still in the Euclidean setup but the Lebesgue measure is replaced by a locally finite measure μ which is assumed to be *doubling*.

DEFINITION 4.41. Let μ be a locally finite, non-negative measure on $\mathbb{R}^n$. We say that μ is *doubling* if there exists a constant c_μ, which may depend on the dimension, such that

$$\mu(B(x, 2r)) \le c_\mu \mu(B(x, r))$$

for every $x \in \mathbb{R}^n$ and $r > 0$. The constant c_μ will be called the *doubling constant* of μ.

The definition of a doubling measure above was given with respect to Euclidean balls. However it is an easy exercise to show that μ is doubling with respect to Euclidean balls if and only if it is doubling with respect to Euclidean cubes, possibly with a different doubling constant.

EXERCISE 4.42. Show that μ is doubling with respect to Euclidean balls if and only if it is doubling with respect to Euclidean cubes. The doubling constants however might be different.

We saw in the previous section that the *centered* maximal function $M^\mu f$ is always of weak type $(1,1)$ and strong (p,p), $1<p\leq\infty$ whenever μ is just locally finite and non-negative. This is no longer true in such generality if we consider the non-centered version of the maximal operator. We define

$$M'^\mu f(x) := \sup_{\substack{B \text{ ball} \\ B\ni x}} \frac{1}{|B|}\int_B |f(y)|d\mu(y).$$

For this non-centered version of the maximal operator we need some extra assumptions on the measure μ. The doubling assumption is one such hypothesis which allow us to extend all the theory for the maximal operator to the operator M'^μ as defined above.

THEOREM 4.43. *Let μ be a doubling measure on $\mathbb{R}^n$ (non-negative, locally finite) and let M'^μ denote the non-centered maximal operator with respect to μ as above. Then M'^μ is of weak type $(1,1)$ and strong type (p,p) for all $1<p\leq+\infty$. The operator norms depend only on the doubling constant of μ, the dimension, and p.*

PROOF. The idea of the proof is that for doubling measures μ one can easily prove that the operators M^μ and M'^μ are pointwise equivalent. Indeed, let $B=B(z,r)$ denote any Euclidean ball in $\mathbb{R}^n$ and let $x\in B$. Letting $\tilde{B}:=B(x,2r)$ we readily see that $B\subset\tilde{B}\subset 3B$, where $3B=B(z,3r)$. By the doubling property of the measure μ we see that $\mu(\tilde{B})\leq\mu(2^2B)\leq c_\mu^2\mu(B)$. Thus for any B with $B\ni x$ we have

$$\frac{1}{|B|}\int_B |f(y)|d\mu(y)\leq c_\mu^2\frac{1}{\mu(\tilde{B})}\int_{\tilde{B}}|f(y)|d\mu(y)\leq M^\mu f(x).$$

The previous estimate shows that $M'^\mu f(x)\leq c_\mu^2 M^\mu f(x)$ while we obviously also have $M^\mu f(x)\leq M'^\mu f(x)$. Thus $M'^\mu f(x)\simeq_\mu M^\mu f(x)$ and Theorem 4.35 completes the proof. □

REMARK 4.44. Under the doubling assumption one can essentially repeat the proof given for the Hardy-Littlewood maximal function, using an obvious analogue of the Vitali covering lemma. Thus the previous theorem has a more elementary proof that goes through in a more abstract setting, as we shall see below.

4.5. The Calderón-Zygmund decomposition

Let (X,μ) be a measure space and $f:X\to\mathbb{C}$ be a measurable function (say) in $L^p(X,\mu)$. For a level $\lambda>0$ we have many times used the decomposition of f at level $\lambda>0$:

$$f=f\mathbf{1}_{\{x\in X:|f(x)|\leq\lambda\}}+f\mathbf{1}_{\{x\in X:|f(x)|>\lambda\}}=:g+b.$$

The function $g=f\mathbf{1}_{\{x\in X:|f(x)|\leq\lambda\}}$ is the "good" part of f; indeed we have that

$$\|g\|_{L^p}\leq\|f\|_{L^p}\quad\text{and}\quad\|g\|_{L^\infty}\leq\lambda.$$

Thus the good part g adopts the L^p-integrability of f and furthermore it is bounded. On the other hand the "bad" part b satisfies

$$\|b\|_{L^p}\leq\|f\|_{L^p}\quad\text{and}\quad\mu(\operatorname{supp}(b))\leq\frac{\|f\|_{L^p}^p}{\lambda^p}.$$

Thus the bad part b also inherits the L^p-integrability of f but it also has "small" support.

In a general measure space one cannot do much more than that in terms of decomposing f in a good part and a bad part. If however there is also a metric structure in the space which is compatible with the measure, one can do a bit better and also get some control on the local oscillation of the bad part b. Various forms of this decomposition are usually referred to as Calderón-Zygmund decompositions. We present here the basic example in the dyadic Euclidean setup.

PROPOSITION 4.45 (Dyadic Calderón-Zygmund decomposition). *Let $f \in L^1(\mathbb{R}^n)$ and $\lambda > 0$. There exists a decomposition of f of the form*

$$f = g + \sum_{Q\in\mathcal{B}} b_Q,$$

where $\mathcal{B}$ is a collection of disjoint dyadic cubes and the sum is taken over all the cubes $Q \in \mathcal{B}$. This decomposition satisfies the following properties:

(i) *The "good part" g satisfies the bound*

$$\|g\|_{L^1(\mathbb{R}^n)} \le \|f\|_{L^1(\mathbb{R}^n)} \quad \textit{and} \quad \|g\|_{L^\infty(\mathbb{R}^n)} \le 2^n\lambda.$$

(ii) *The "bad part" is $b = \sum_{Q\in\mathcal{B}} b_Q$; each function b_Q is supported on Q and*

$$\int_Q b_Q = 0, \quad \|b_Q\|_{L^1(\mathbb{R}^n)} \le 2^{n+1}\lambda|Q|, \quad \textit{for all} \quad Q\in\mathcal{B}.$$

(iii) *For each $Q \in \mathcal{B}$ we have*

$$\lambda \le \frac{1}{|Q|}\int_Q |f(y)|dy \le 2^n\lambda.$$

Furthermore we have that

$$\bigcup_{Q\in\mathcal{B}} Q = \{x\in\mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\} \subset \{x\in\mathbb{R}^n : M(f)(x) > \lambda\}.$$

In particular, from the dyadic maximal theorem we have

$$\sum_{Q\in\mathcal{B}} |Q| \le \frac{\|f\|_{L^1(\mathbb{R}^n)}}{\lambda}.$$

PROOF. The proof is very similar to the proof of the dyadic covering lemma. We fix some level $\lambda > 0$ and let us call a dyadic cube Q *bad* if

$$\frac{1}{|Q|}\int_Q |f| > \lambda.$$

If a dyadic cube is not bad we call it *good*. A bad cube will be called *maximal* if Q is bad and also there is no dyadic cube strictly containing Q.

Observe that every bad cube is contained in some maximal bad cube. Indeed, if Q' is bad cube then $\mathbf{1}_{2^mQ'} \to 1$ as $m\to\infty$ so monotone convergence implies that $\int_{2^mQ'}|f| \nearrow \|f\|_{L^1(\mathbb{R}^n)}$. It follows that there is a large enough M such that

$$\frac{1}{|2^MQ'|}\int_{2^MQ'}|f| > \lambda \quad \text{and} \quad \frac{1}{|2^mQ'|}\int_{2^mQ'}|f| < \lambda,$$

for all $m > M$. Thus the dyadic cube $2^MQ'$ is maximal and bad.

Let us denote by $\mathcal{B}$ the collection of maximal bad cubes. Since the cubes in the collection $\mathcal{B}$ are dyadic and maximal, they are disjoint. Also, for any bad cube Q', let $x \in Q'$. We have that

$$M_{\mathcal{D}}(f)(x) = \sup_{\substack{Q\in\mathcal{D} \\ Q\ni x}} \frac{1}{|Q|}\int_Q f \geq \frac{1}{|Q'|}\int_{Q'} |f| > \lambda.$$

The previous estimate implies that $\cup_{Q\in\mathcal{B}} Q \subset \{x : M_{\mathcal{D}} f(x) > \lambda\}$. For the opposite inclusion let $x \in \{x : M_{\mathcal{D}} f(x) > \lambda\}$. Then there is a cube $Q \in \mathcal{D}$ such that

$$\frac{1}{|Q|}\int_Q |f(y)|dy > \lambda$$

so that Q is bad. Since every bad cube is contained in a maximal bad cube we get that $\{x : M_{\mathcal{D}} f(x) > \lambda\} \subset \cup_{Q \text{ bad}} Q \subset \cup_{Q\in\mathcal{B}} Q$.

Now let Q be a maximal bad cube and consider the parent of Q, $Q^{(1)}$, that is the unique dyadic cube with twice its sidelength that contains Q. Since Q is maximal, $Q^{(1)}$ has to be good so we have

$$\frac{1}{|Q^{(1)}|}\int_{Q^{(1)}} |f| \leq \lambda$$

and thus

$$\frac{1}{|Q|}\int_Q |f| \leq 2^n\lambda.$$

for all maximal bad cubes Q. We set

$$b_Q := \Big(f - \frac{1}{|Q|}\int_Q f\Big)\mathbf{1}_Q,$$

whenever $Q \in \mathcal{B}$ is a maximal bad cube. We also set

$$g := (1 - \mathbf{1}_{\cup_{Q\in\mathcal{B}} Q})f + \sum_{Q\in\mathcal{B}} \Big(\frac{1}{|Q|}\int_Q f\Big)\mathbf{1}_Q = f - \sum_{Q\in\mathcal{B}} b_Q.$$

It is not hard to verify all the required properties of b, g except maybe that $\|g\|_{L^\infty(\mathbb{R}^n)} \leq 2^n\lambda$. It is easy to see that

$$\sup_{x\in Q} |g(x)| \leq \frac{1}{|Q|}\int_Q |f| \leq 2^n\lambda,$$

whenever $Q \in \mathcal{B}$ is a bad cube. If $x \notin \bigcup_{Q\in\mathcal{B}} Q$ and $x \in Q' \in \mathcal{D}$, then necessarily Q' is good. We thus have that

$$\frac{1}{|Q'|}\int_{Q'} |g| = \frac{1}{|Q'|}\int_{Q'} |f| < \lambda,$$

since Q' is good. It follows $\frac{1}{|Q'|}\int_{Q'} f(y)dy \to f(x)$ as $|Q'| \to 0$ with $x \in Q'$, by the dyadic maximal theorem. Since $x \notin \bigcup_{Q\in\mathcal{B}} Q$ we conclude that $|g(x)| = |f(x)| \leq \lambda$ and we are done in this case as well. □

Observe that in the previous decomposition of $f = b + g$, the "bad set", that is the set where b lives, is given in the form

$$\cup_{Q\in\mathcal{B}} Q = \{x \in \mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\}.$$

One could prove the Calderón-Zygmund decomposition starting from the set $\{x \in \mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\}$ and decomposing it as a disjoint union of dyadic cubes. This sort of decomposition is interesting in its own right. Let us see how this can be done.

PROPOSITION 4.46 (Dyadic Whitney decomposition). *Let $\Omega \subset \mathbb{R}^n$ be an open set which is not all of $\mathbb{R}^n$. Then there exists a decomposition*

$$\Omega = \bigcup_{Q \in \mathcal{Q}} Q,$$

where $\mathcal{Q}$ is a collection of disjoint dyadic cubes. For each $Q \in \mathcal{Q}$ we have

$$\operatorname{dist}(Q, \mathbb{R}^n \setminus \Omega) \simeq \operatorname{diam}(Q).$$

PROOF. Let $\mathcal{Q}$ denote the dyadic cubes inside Ω such that

$$\operatorname{diam}(Q) \leq \operatorname{dist}(Q, \mathbb{R}^n \setminus \Omega) \leq 5 \operatorname{diam}(Q). \tag{4.2}$$

Obviously $\cup_{Q \in \mathcal{Q}} Q \subset \Omega$ but the opposite inclusion is also true. Indeed, if $x \in \Omega$ note that x is contained in some dyadic cube $Q \subset \Omega$ since Ω is open. Now for Q a dyadic cube let Q' be its "parent", that is the unique dyadic cube of side twice the side-length of Q, containing Q. Considering successive parents of Q there will be a dyadic cube Q'' containing x with diameter greater than $\operatorname{dist}(x, \mathbb{R}^n \setminus \Omega)/4$ and less than $\operatorname{dist}(x, \mathbb{R}^n \setminus \Omega)/2$. Thus $Q'' \subset \Omega$ and $\operatorname{diag}(Q'') \simeq \operatorname{dist}(Q'', \mathbb{R}^n \setminus \Omega)$. The collection of dyadic cubes $\mathcal{Q}$ is not necessarily disjoint so we only choose the cubes in $\mathcal{Q}$ which are maximal with respect to set inclusion and call this collection again $\mathcal{Q}$. Now maximal and dyadic means disjoint so we are done. $\square$

Using the Whitney decomposition lemma one can give an alternative proof of the Calderón-Zygmund decomposition by taking

$$\Omega = \{x \in \mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\},$$

and noting that the latter set is open.

As a corollary we get a control of the level sets of the Hardy-Littlewood maximal function by the level sets of the dyadic maximal function.

LEMMA 4.47. *For all $\lambda > 0$ we have that*

$$|\{x \in \mathbb{R}^n : M_{\square}(f)(x) > 4^n \lambda\}| \leq 2^n |\{x \in \mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\}|.$$

PROOF. Let $\mathcal{B}$ be the collection of dyadic cubes obtained by the Calderón-Zygmund decomposition at level $\lambda > 0$. We have that

$$\cup_{Q \in \mathcal{B}} Q = \{x \in \mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\}.$$

We write Q^* for the cube with the same center as Q and twice its side-length. We claim that

$$\{x \in \mathbb{R}^n : M_{\square}(f)(x) > 4^n \lambda\} \subset \cup_{Q \in \mathcal{B}} Q^*. \tag{4.3}$$

Indeed, let $x \notin \bigcup_{Q \in \mathcal{B}} Q^*$ and R be any cube centered at x. Denoting by r the side-length of R, we choose $k \in \mathbb{Z}$ so that $2^{k-1} \leq r < 2^k$. Then R intersects $m \leq 2^n$ cubes in the k-th generation $\mathcal{D}_k$, and let us call them $R_1, \ldots, R_m$. Observe that none of these cubes can be contained in any of the $Q \in \mathcal{B}$ because otherwise we would have that $x \in \bigcup_{Q\in\mathcal{B}} Q^*$. Thus the average of f on each R_j is at most λ so

$$\frac{1}{|R|}\int_R |f| \le \frac{1}{|R|}\sum_{j=1}^m \int_{R_j\cap R} |f| \le \sum_{j=1}^m \frac{2^{kn}}{|R|}\frac{1}{|R_j|}\int_{R_j} |f| \le \lambda m 2^n \le 4^n\lambda.$$

This proves the claim (4.3) and thus the corollary. □

EXERCISE 4.48. Using the dyadic maximal theorem only, conclude that the operators $M_\square, M$ are of weak type $(1,1)$.

4.5.1. The Fefferman-Stein inequality. We give a first application of the Calderón-Zygmund decomposition which in some sense is the prototype of a weighted norm inequality. It is a variation of the maximal theorem where the Lebesgue measure is replaced by a measure of the form $w(x)dx$ for some non-negative measurable function w. It then turns out that the maximal function maps $L^p(\mathbb{R}^n, Mw(x)dx)$ to $L^p(\mathbb{R}^n, w(x)dx)$ boundedly for all $1 < p < \infty$ and that it also satisfies a weak endpoint analogue for $p = 1$. In particular we have

THEOREM 4.49 (Fefferman-Stein inequality). *Let w be a non-negative locally integrable function (a weight).*

(i) *The operator M maps $L^1(Mw)$ to $L^{1,\infty}(w)$:*

$$\int_{\{x\in\mathbb{R}^n : M(f)(x)>\lambda\}} w(x)dx \lesssim_n \int_{\mathbb{R}^n} |f(x)|Mw(x)dx,$$

for all $f \in L^1(\mathbb{R}^n, Mw)$.

(ii) *We have that*

$$\int_{\mathbb{R}^n} [Mf(x)]^p w(x)dx \lesssim_{p,n} \int_{\mathbb{R}^n} |f(x)|^p Mw(x)dx,$$

for all $f \in L^p(\mathbb{R}^n, Mw)$ with $1 < p \le \infty$.

PROOF. We will show that $\|M(f)\|_{L^\infty(w)} \le \|f\|_{L^\infty(Mw)}$ and that the weak $(1,1)$ inequality in (i) holds. Then the Marcinkiewicz interpolation theorem will give (ii) as well.

The bound

$$\|M(f)\|_{L^\infty(w)} \le \|f\|_{L^\infty(Mw)},$$

is trivial and is left as an exercise. We turn our attention to the $(1,1)$-bound. Let $\mathcal{B}$ be the collection of the dyadic cubes obtained from the Calderón-Zygmund decomposition at level $\lambda > 0$. By the proof of Lemma 4.47 we have that

$$\{x \in \mathbb{R}^n : M_\square(f) > 4^n\lambda\} \subset \cup_{Q\in\mathcal{B}} Q^*,$$

where Q^* is the cube with the same center as Q and twice its side-length. We have

$$\int_{\{x\in\mathbb{R}^n : M_\square(f)(x)>4^n\lambda\}} w(x)dx \le \sum_{Q\in\mathcal{B}} \int_{Q^*} w(x)dx = \sum_{Q\in\mathcal{B}} 2^n|Q|\frac{1}{|Q^*|}\int_{Q^*} w(x)dx.$$

Again, from the Calderón-Zygmund decomposition (at level λ) we have that

$$|Q| < \frac{1}{\lambda}\int_Q |f(y)|dy = \frac{1}{\lambda}\int_{\mathbb{R}^n} |f(y)|\mathbf{1}_Q(y)dy,$$

for all $Q \in \mathcal{B}$ of the decomposition. Combining the last two estimates we can write

$$\int_{\{x\in\mathbb{R}^n: M_\square(f)(x)>4^n\lambda\}} w(x)dx \leq \frac{2^n}{\lambda}\sum_{Q\in\mathcal{B}}\int_{\mathbb{R}^n}|f(y)|\left(\frac{1}{|Q^*|}\int_{Q^*}w(x)dx\right)\mathbf{1}_Q(y)dy.$$

For fixed $Q \in \mathcal{B}$ the term $|f(y)|\left(\frac{1}{|Q^*|}\int_{Q^*}w(x)dx\right)\mathbf{1}_Q(y)$ is non-zero if and only if $y \in Q \subset Q^*$. Thus the previous estimate implies

$$\int_{\{x\in\mathbb{R}^n: M_\square(f)(x)>4^n\lambda\}} w(x)dx \leq \frac{2^n}{\lambda}\int_{\mathbb{R}^n}|f(y)|M'_\square(w)(y)dy,$$

where $M'_\square$ is the non-centered maximal function associated to cubes. See Exercise 4.33. Since $M'_\square(f)(x) \lesssim_n M(f)(x)$ this concludes the proof. □

EXERCISE 4.50 (Heldberg's inequality and Hardy-Littlewood-Sobolev theorem). Let $0<\gamma<n$, $1<p<q<\infty$ and $\frac{1}{q}=\frac{1}{p}-\frac{n-\gamma}{n}$.

(i) Show Heldberg's inequality: If $f \in L^p(\mathbb{R}^n)$ then

$$|(f*|y|^{-\gamma})(x)| \lesssim_{\gamma,n,p} [M(f)(x)]^{\frac{p}{q}}\|f\|_{L^p(\mathbb{R}^n)}^{1-\frac{p}{q}}.$$

(ii) Use the Hardy-Littlewood maximal theorem and (i) to conclude the Hardy-Littlewood-Sobolev theorem: For every $f \in L^p(\mathbb{R}^n)$ we have that

$$\|f*|y|^{-\gamma}\|_{L^q(\mathbb{R}^n)} \lesssim_{\gamma,n,p} \|f\|_{L^p(\mathbb{R}^n)}.$$

Hint: In order to show (i) split the integral

$$\begin{aligned}|(f*|y|^{-\gamma})(x)| &= \left|\int_{\mathbb{R}^n} f(x-y)|y|^{-\gamma}dy\right| \\ &\leq \left|\int_{|y|<R} f(x-y)|y|^{-\gamma}dy\right| + \left|\int_{|y|\geq R} f(x-y)|y|^{-\gamma}dy\right| =: I_1+I_2,\end{aligned}$$

where $R>0$ is a parameter to be chosen later on. For I_1 observe that

$$I_1 = f*(|y|^{-\gamma}\chi_{B(0,R)}).$$

Observe that $|y|^{-\gamma}\chi_{B(0,R)}$ is decreasing, radial, non-negative and integrable (since $\gamma<n$). Use Proposition 4.17 and the calculation in its proof to show the bound

$$|I_1| \lesssim R^{n-\gamma}M(f)(x).$$

For I_2 use Hölder's inequality to show

$$|I_2| \lesssim R^{-\frac{n}{q}}\|f\|_{L^p(\mathbb{R}^n)}.$$

Choose the parameter $R>0$ to minimize the sum I_1+I_2. Part (ii) is a trivial consequence of (i).

CHAPTER 5

The Hilbert transform

In this chapter we initiate the study of singular integral operators, that is operators of the form

$$Tf(x) = \int K(x,y)f(y)dy, \quad x \in \mathbb{R}^n, \tag{5.1}$$

defined initially for "nice" functions $f \in \mathcal{S}(\mathbb{R}^n)$. Here we typically want to include the case where K has a singularity close to the diagonal

$$\Delta = \{(x,x) : x \in \mathbb{R}^n\} \subset \mathbb{R}^{2n},$$

which is not locally integrable. Typical examples are

$$K(x,y) = \frac{1}{|x-y|^n}, \quad x,y \in \mathbb{R}^n,$$

$$K(x,y) = \frac{x_j - y_j}{|x-y|^{n+1}}, \quad x,y \in \mathbb{R}^n$$

and in one dimension

$$K(x,y) = \frac{1}{x-y}, \quad x,y \in \mathbb{R},$$

and so on. Observe that these kernels have a non-integrable singularity both at infinity as well as on the diagonal Δ. It is however the local singularity close to the diagonal that is the most crucial and that will characterize a kernel as a singular kernel. For example, the kernel

$$K(x-y) = \frac{1}{|x-y|^{n-\epsilon}}, \quad \epsilon > 0$$

is not a singular kernel since its singularity is locally integrable. Observe that for Schwartz functions $f \in \S(\mathbb{R}^n)$ it makes perfect sense to define

$$T(f)(x) = \int_{\mathbb{R}^n} \frac{f(y)}{|x-y|^{n-\epsilon}} dy,$$

and in fact the previous integral operator was already considered in the Hardy-Littlewood-Sobolev inequality of Exercise 4.50 and can be treated via the standard tools we have seen so far.

Thus, if one insists on writing the representation formula (5.1) throughout $\mathbb{R}^n$ then K will not be a function in general. Indeed, the discussion in § 3.7 reveals that if the operator T is translation invariant then the kernel K must necessarily be of the form $K(x-y)$ for an appropriate tempered distribution $K \in \mathcal{S}(\mathbb{R}^n)$:

$$T(f) = K * f.$$

Bearing in mind that there are tempered distributions which do not arise from functions or measures we see that (5.1) does not make sense in general and it should be understood in a different way. To give a more concrete example, think of the principal value distribution $K = \text{p.v.}\frac{1}{x} \in \mathcal{S}'(\mathbb{R})$ and write

$$Tf := f * \text{p.v.}\frac{1}{x}.$$

Here we would like to rewrite this in the form

$$Tf = \int_{\mathbb{R}} \frac{f(y)}{x-y} dy,$$

but this does not make sense even for $f \in \mathcal{S}(\mathbb{R})$ since the function $\frac{1}{x-y}$ is not locally integrable on the diagonal $x = y$.

In fact, the representation (5.1) of the operator will not be true in general but we will satisfy ourselves with its validity for functions $f \in L^2(\mathbb{R}^n)$, of compact support, and whenever x does not lie in the support of f. Indeed, if f has compact support and $x \notin \text{supp}(f)$ then $|y - x| > \epsilon$ in (5.1) and thus we are away from the diagonal. Indeed, returning to the principal value example, observe that the integral

$$\int_{\mathbb{R}} \frac{f(y)}{x-y} dy,$$

makes perfect sense when f has compact support and $x \notin \text{supp}(f)$.

Eventually the theory of singular integral operators does not depend on translation invariance; singular kernels of the type $K(x - y)$ can be viewed as a special cases of the more general class of singular kernels $K(x, y)$ which satisfy appropriate growth and regularity assumptions. It is however instructive to consider the translation invariant case first. In the Calderón-Zygmund theory of singular integral operators we will start with more or less assuming that the operator T is well defined and bounded on $L^2(\mathbb{R}^n)$ and that its kernel K satisfies certain growth and regularity conditions. Alternatively, assumptions on K will allow us to show the L^2-boundedness. We will see that under these conditions T will extend to a bounded operator on $L^p(\mathbb{R}^n)$ for $1 < p < \infty$ and of weak type $(1, 1)$.

5.1. The definition of the Hilbert transform on $\mathcal{S}(\mathbb{R}^n)$.

In order to illustrate the general ideas let us consider what is probably the primordial example of a singular integral operator, the *Hilbert transform*, given in the form

$$\begin{aligned} Hf(x) &:= \text{p. v.}\, \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(y)}{x-y} dy = \text{p. v.}\, \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(x-y)}{y} dy \\ &= \lim_{\epsilon \to 0} \frac{1}{\pi} \int_{|x|>\epsilon} \frac{f(x-y)}{y} dy. \end{aligned}$$

Remembering the principal value distribution we can rewrite this in the form

$$Hf(x) = (\text{p. v.}\, \frac{1}{\pi y} * f)(x),$$

at least whenever $f \in \mathcal{S}(\mathbb{R})$. The previous formula makes sense just because the principal value of $1/\pi y$ is a well defined tempered distribution. Alternatively, we can repeat the argument we used for $\mathrm{p.v.}\,\frac{1}{\pi y}$ to write for any $\epsilon > 0$ and a Schwartz function $f \in \mathcal{S}(\mathbb{R})$

$$\int_{|y|>\epsilon} \frac{f(x-y)}{y} dy = \int_{\epsilon<|y|<1} \frac{f(x-y)-f(x)}{y} dy + \int_{1<|y|<\infty} \frac{f(x-y)}{y} dy.$$

Observe that we heavily rely on the fact that the kernel $\frac{1}{y}$ has zero mean on symmetric intervals around (and away from) the origin:

$$\int_{a<|y|<b} \frac{1}{y} dy = 0, \quad 0 < a < b < +\infty.$$

The mean value theorem now shows that $\frac{f(x-y)-f(x)}{y}$ is uniformly bounded by $\|f'\|_{L^\infty(\mathbb{R})}$ thus the limit of the first summand as $\epsilon \to 0$ exists and we have that

$$Hf(x) = \int_{0<|y|<1} \frac{f(x-y)-f(x)}{y} dy + \int_{|y|>1} \frac{f(x-y)}{y} dy, \tag{5.2}$$

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

REMARK 5.1. Trying to write the Hilbert transform as an integral operator with respect to a kernel K,

$$T(f)(x) = \int_{\mathbb{R}} K(x,y) f(y) dy,$$

we immediately run into the problem that the principal value distribution does not arise from a function. The previous discussion allows us however to write

$$Hf(x) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(y)}{x-y} dy = \frac{1}{\pi} \int_{\mathbb{R}} \frac{f(x-y)}{y} dy,$$

whenever f is a compactly supported function in $\mathcal{S}(\mathbb{R}^n)$ or $L^2(\mathbb{R})$ and $x \notin \operatorname{supp}(f)$. This is essentially equivalent to the fact that the integrals

$$\frac{1}{\pi} \int_{|x-y|>\epsilon} \frac{f(y)}{x-y} dy,$$

are absolutely convergent whenever $f \in L^2(\mathbb{R})$ and $\epsilon > 0$ is fixed.

Thus we see that the Hilbert transform is a linear operator which is at least well defined on the Schwartz class $\mathcal{S}(\mathbb{R})$. This is quite promising since we know that $\mathcal{S}(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for $p < \infty$. Of course, in order to extend the action of H to say $L^2(\mathbb{R})$ we need to exhibit the continuity of H on the dense subclass $\mathcal{S}(\mathbb{R})$. In the abstract theory of singular integrals it will be a "given" that our operator is bounded on L^2. To make this general assumption meaningful we have to exhibit that it is indeed satisfied in the model case of the Hilbert transform. We begin this investigation by first showing a simple asymptotic relationship.

LEMMA 5.2. *Let $f \in \mathcal{S}(\mathbb{R}^n)$. Then we have*

$$\lim_{|x|\to+\infty} xHf(x) = \frac{1}{\pi} \int_{\mathbb{R}} f(y) dy.$$

Before giving the proof of this Lemma let us discuss its consequences. Already the expression (5.2) shows that Hf is a bounded function whenever $f \in \mathcal{S}(\mathbb{R})$. Indeed, using the mean value theorem for the first term in (5.2) and Hölder's inequality for the second term we have that

$$|Hf(x)| \lesssim \|f'\|_{L^\infty(\mathbb{R})} + \|f\|_{L^2(\mathbb{R})}.$$

As a result, the integrability of Hf for $f \in \mathcal{S}(\mathbb{R})$ solely depends on the behavior of Hf at infinity. Now the lemma just stated shows that

$$Hf(x) \simeq_f \frac{1}{|x|}, \quad |x| \to \infty,$$

whenever $f \in \mathcal{S}(\mathbb{R})$ with $\int_{\mathbb{R}} f(y)dy \neq 0$. Thus for a general $f \in \mathcal{S}$ with non-zero mean, Hf fails to be in $L^1(\mathbb{R})$ since it doesn't decay fast enough at infinity. It is however in $L^p(\mathbb{R})$ for any $p > 1$. As we shall see the failure of continuity of H on L^1 has a weak substitute, namely that H is of weak type $(1,1)$ and this is the typical behavior of all singular integral operators we want to consider.

PROOF OF LEMMA 5.2. The proof is a variation of the idea used in (5.2). For any $\epsilon > 0$ and $|x|$ large we can write

$$\begin{aligned}
\lim_{\epsilon\to 0} x \int_{|y|>\epsilon} \frac{f(x-y)}{y} dy &= x \int_{0<|y|\leq \frac{|x|}{2}} \frac{f(x-y)-f(x)}{y} dy \\
&\quad + x \int_{\frac{|x|}{2}<|y|\leq 2|x|} \frac{f(x-y)}{y} dy \\
&\quad + x \int_{|y|>2|x|} \frac{f(x-y)}{y} dy \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

For I_1 observe that $|x|/2 \leq |x-y| \leq 3|x|/2$ whenever $|y| \leq |x|/2$ thus we have that

$$|I_1| \lesssim |x|^2 \sup_{|\xi|\simeq|x|} |f'(\xi)| \simeq \sup_{|\xi|\simeq|x|} |\xi^2 f'(\xi)| \to 0$$

as $|x| \to \infty$ since f is a Schwartz function. On the other hand, for I_3 we have that $|x-y| \geq |x|$ whenever $|y| > 2|x|$. We get

$$|I_3| \lesssim \int_{|x-y|\geq|x|} |f(x-y)|dy = \int_{|z|\geq|x|} |f(z)|dz \to 0$$

as $|x| \to \infty$ since f is integrable, f being a Schwartz function. Now consider the expression

$$I_2 - \int_{\mathbb{R}} f(x-y)dy = \int_{\frac{|x|}{2}<|y|\leq 2|x|} (x/y-1)f(x-y)dy - \int_{\{|y|<|x|/2\}\cup\{|y|>2|x|\}} f(x-y)dy,$$

thus

$$\left| I_2 - \int_{\mathbb{R}} f \right| \lesssim \frac{1}{|x|} \int_{\mathbb{R}} |yf(y)|dy + \int_{|z|>|x|/2} |f(z)|dy \to 0,$$

as $|x| \to \infty$. □

EXERCISE 5.3. Let $f \in \mathcal{S}(\mathbb{R}^n)$. Show that $Hf \in L^1(\mathbb{R})$ if and only if $\int_{\mathbb{R}} f(y)dy = 0$.

Hint: One way to show this is to rework the proof of Lemma 5.2 in order to estimate $x^2 Hf(x)$.

5.2. The Hilbert transform on $L^2(\mathbb{R})$

Having exhibited that $Hf \in L^2(\mathbb{R})$ whenever $f \in \mathcal{S}(\mathbb{R})$ our next task is to show that H is bounded as an operator $H : \mathcal{S}(\mathbb{R}) \cap L^2(\mathbb{R}) \to L^2(\mathbb{R})$, that is to show that

$$\|Hf\|_{L^2(\mathbb{R})} \lesssim \|f\|_{L^2(\mathbb{R})},$$

for all $f \in \mathcal{S}(\mathbb{R})$. Remember that since $\mathcal{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$ such an estimate will allow us to extend H to a bounded linear operator on $L^2(\mathbb{R})$. There are several different approaches to such a theorem, most of them connected to the significance of the Hilbert transform in complex analysis and the theory of holomorphic functions. A simple way to show the boundedness of H on $L^2(\mathbb{R})$ is use Example 3.66 for the principal value distribution according to which

$$\mathcal{F}\Big(\mathrm{p.v.}\,\frac{1}{\pi x}\Big)(\xi) = -i\,\mathrm{sgn}(\xi)$$

It immediately follows that $\|Hf\|_{L^(\mathbb{R})} = \|f\|_{L^(\mathbb{R})}$. Here we describe an alternative approach based on the connection of the Hilbert transform with Cauchy integrals.

PROPOSITION 5.4. *Let f be a function on $\mathbb{R}$ such that Hf is well defined, say $f \in C^1(\mathbb{R})$ and $|f(x)| \lesssim (1+|x|)^{-1}$ for $|x|$ large. Then*

$$\lim_{\epsilon\to 0} \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{f(y)}{y-(x\pm i\epsilon)}dy = \frac{\pm f(x) + iHf(x)}{2},$$

for every $x \in \mathbb{R}$.

PROOF. By the translation invariance of H and taking complex conjugate in both sides of the identity it suffices to show that

$$\lim_{\epsilon\to 0} \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{f(y)}{y-i\epsilon}dy = \frac{f(0) + iHf(0)}{2}, \tag{5.3}$$

which is equivalent to

$$\lim_{\epsilon\to 0} \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{f(y)}{y-i\epsilon}dy - \frac{1}{2}f(0) - \frac{i}{2\pi}\int_{|y|>\epsilon} \frac{f(y)}{-y}dy = 0.$$

Changing variables $y = \epsilon u$ this is equivalent to

$$\lim_{\epsilon\to 0} \int_{\mathbb{R}} \Big(\frac{1}{u-i} - \chi_{\{|u|>1\}}(u)\frac{1}{u}\Big) f(\epsilon u)du = \pi i f(0).$$

Now let

$$h(u) = \frac{1}{u-i} - \chi_{\{|u|>1\}}(u)\frac{1}{u}.$$

For $|u| \leq 1$ we have that

$$|h(u)| = \frac{1}{|u-i|} = \frac{1}{(1+u^2)^{\frac{1}{2}}} \leq 1,$$

while for $|u| > 1$ we can calculate

$$|h(u)| = \frac{1}{|u^2 - iu|} = \frac{1}{(u^2+u^4)^{\frac{1}{2}}} \leq \frac{1}{u^2}.$$

The previous estimates obviously imply that h is absolutely integrable on $\mathbb{R}$. Furthermore

$$\int_{\mathbb{R}} h(u)du = \int_{\mathbb{R}} \Big(\frac{1}{u-i} - \mathbf{1}_{\{|u|>1\}}(u)\frac{1}{u} \Big) du = i\pi,$$

as can be seen by a direct calculation. Thus by the previous calculations it suffices to show that

$$\lim_{\epsilon\to 0} \int_{\mathbb{R}} (f(\epsilon u) - f(0))h(u)du = 0, \tag{5.4}$$

which follows by dominated convergence since $h \in L^1(\mathbb{R})$ and f is bounded. $\square$

EXERCISE 5.5. Show that for $f \in C^1(\mathbb{R})$ satisfying $|f(x)| \le (1+|x|)^{-1}$ for $|x| \to \infty$ the Hilbert transform Hf is indeed well defined. Furthermore, show that it indeed suffices to show (5.3) in the previous proposition. In particular exhibit how the full statement of the previous follows from (5.3).

THEOREM 5.6. *If $f \in \mathcal{S}(\mathbb{R})$ then*

$$\widehat{Hf}(\xi) = -i \operatorname{sgn}(\xi)\hat{f}(\xi).$$

PROOF. Let us define the Cauchy-type integral

$$C_\epsilon(f)(x) = \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f(y)}{y-(x-i\epsilon)} dy.$$

Then Proposition 5.4 shows that

$$\lim_{\epsilon\to 0} C_\epsilon(f)(x) = \frac{-f(x) + iHf(x)}{2}.$$

Observe by the proof of the proposition applied to the function $\tau_{-x}f$ that

$$C_\epsilon(f)(x) - \frac{-f(x)+iHf(x)}{2} = \int_{\mathbb{R}} (\tau_{-\epsilon u}f(x) - f(x))h(u)du$$

for all $x \in \mathbb{R}$. Thus by Minkowski's integral inequality we get that

$$\Big\| C_\epsilon(f) - \frac{-f+iHf}{2} \Big\|_{L^2(\mathbb{R})} \le \int_{\mathbb{R}} \|\tau_{-\epsilon u}f - f\|_{L^2(\mathbb{R})} |h(u)| du.$$

By dominated convergence we conclude that $C_\epsilon(f)$ converges to $\frac{-f+iH(f)}{2}$ in L^2 as well. By Plancherel's theorem we get that we must also have that

$$\widehat{C_\epsilon(f)} \to \frac{1}{2}(-\hat{f} + \widehat{Hf}),$$

in L^2, as $\epsilon \to 0$. Note here that the Fourier transform $\widehat{Hf}$ is well defined since $f \in \mathcal{S}(\mathbb{R})$ and in this case we have exhibited that $Hf \in L^2(\mathbb{R})$. The problem now reduces to calculating the Fourier transform of $C_\epsilon(f)$ for $\epsilon > 0$ and see what happens in the limit. Consider the truncations $C_{\epsilon,R}(f)$

$$C_{\epsilon,R}(f)(x) = \frac{1}{2\pi i} \int_{|x-y|<R} \frac{f(y)}{y-(x-i\epsilon)} dy.$$

Let us write

$$g_\epsilon(t) = \frac{1}{2\pi i} \frac{1}{-t+i\epsilon}, \quad g_{\epsilon,R}(t) = \frac{1}{2\pi i} \frac{1}{-t+i\epsilon} \chi_{\{|t|<R\}}.$$

Then $g_{\epsilon,R}(t) \to g_\epsilon$ as $R \to \infty$ in L^2 by dominated convergence and thus

$$\|C_{\epsilon,R}(f) - C_\epsilon(f)\|_{L^2(\mathbb{R})} = \|f * g_{\epsilon,R} - f * g_\epsilon\|_{L^2(\mathbb{R})} \leq \|f\|_{L^1(\mathbb{R})}\|g_{\epsilon,R} - g_\epsilon\|_{L^2(\mathbb{R})} \to 0,$$

as $R \to 0$. We now have that

$$\widehat{C_{\epsilon,R}(f)}(\xi) = \hat{f}(\xi)\widehat{g_{\epsilon,R}}(\xi).$$

However we have that

$$\widehat{g_{\epsilon,R}}(\xi) = -\frac{1}{2\pi i}\int_{|x|<R} \frac{e^{-2\pi i x\xi}}{-x + i\epsilon}dx.$$

Now Cauchy's theorem from Complex analysis shows that $\lim_{R\to\infty} \widehat{g_{\epsilon,R}}(\xi) = 0$ whenever $\xi > 0$.

The previous definitions allow us to conclude that the Fourier transform

$$\widehat{C_\epsilon(f)}(\xi) = 0,$$

whenever $\xi > 0$ and thus that

$$\frac{1}{2}(-\hat{f}(\xi) + i\widehat{Hf)(\xi}) = 0$$

whenever $\xi > 0$. We conclude that

$$\widehat{Hf}(\xi) = -i\hat{f}(\xi), \quad \xi > 0.$$

Now not that the Hilbert transform satisfies

$$Hf(-x) = \lim_{\epsilon\to 0}\int_{|y|>\epsilon} \frac{f(-x-y)}{y}dy = -H(\tilde{f})(x),$$

where remember that $\tilde{f}(x) = f(-x)$. So for $\xi > 0$ we can write

$$\begin{aligned}\widehat{Hf}(-\xi) &= \int_{\mathbb{R}} Hf(x)e^{2\pi i x\xi}dx = -\int_R H(\tilde{f})(x)e^{-2\pi i x\xi}dx \\ &= -\widehat{H(\tilde{f})}(\xi) = i\hat{\tilde{f}}(\xi) = i\hat{f}(-\xi).\end{aligned}$$

In other words for $\xi \in \mathbb{R}$ we get that $\widehat{Hf}(\xi) = -i\,\mathrm{sgn}(\xi)\hat{f}(\xi)$. □

The previous theorem shows in particular that $\|Hf\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})}$ for all $f \in \mathcal{S}(\mathbb{R})$. This allows us to extend the Hilbert transform to a bounded linear operator on $L^2(\mathbb{R})$. In fact H is an isometry by Plancherel's theorem and the fact that $|-i\mathrm{sgn}(\xi)| = 1$. Furthermore, although at the current stage it is not clear that our original definition makes sense on $L^2(\mathbb{R})$, we can directly define the Hilbert transform on $L^2(\mathbb{R})$ by means of

$$\widehat{Hf}(\xi) := -i\mathrm{sgn}(\xi)\hat{f}(\xi),$$

which is a good definition whenever $f \in L^2(\mathbb{R})$. In fact, recalling the discussion on multiplier transformations it is clear that the operator H on L^2 is the multiplier transformation associated with the multiplier $m(\xi) = -i\mathrm{sgn}(\xi)$ which is obviously a bounded function. We also have that $\|H\|_{L^2\to L^2} = \|m\|_{L^\infty} = 1$ which is also obvious from the fact that H is an isometry.

COROLLARY 5.7. *The Hilbert transform extends to an isometry on $L^2(\mathbb{R})$. We have that*

$$\|Hf\|_{L^2(\mathbb{R})} = \|f\|_{L^2(\mathbb{R})},$$

for all $f \in L^2(\mathbb{R})$. Furthermore, for $f \in L^2(\mathbb{R})$ the Hilbert transform can be defined by means of

$$\widehat{Hf}(\xi) := -i\mathrm{sgn}(\xi)\hat{f}(\xi), \quad f \in L^2(\mathbb{R}).$$

COROLLARY 5.8. *Consider the Hilbert transform $H : L^2(\mathbb{R}) \to L^2(\mathbb{R})$. Then we have the following properties*

(i) *The Hilbert transform H commutes with translations and dilations (but not modulations).*

$$H\tau_{x_o} = \tau_{x_o}H, \quad \mathrm{Dil}^p_\lambda H = H\mathrm{Dil}^p_\lambda, \quad x_o \in \mathbb{R},\ \lambda > 0,\ 1 \le p \le +\infty.$$

(ii) *The Hilbert transform is skew-adjoint on $L^2(\mathbb{R})$*

$$\int_{\mathbb{R}} Hf\overline{g} = -\int_{\mathbb{R}} f\overline{H(g)}, \quad f, g \in L^2(\mathbb{R}).$$

(iii) *We have the identity $H^2 = -\mathrm{id}$ on $L^2(\mathbb{R})$:*

$$H(Hf) = -f, \quad f \in L^2(\mathbb{R}).$$

EXERCISE 5.9. Prove Corollary 5.8 above.
Hint: Use the formula of Theorem 5.6.

EXERCISE 5.10. Let $f(x) := \mathbf{1}_{[0,1]}(x)$. Show that

$$Hf(x) = \frac{1}{\pi}\log\left|\frac{x}{x-1}\right|.$$

Conclude that the Hilbert transform is not of strong type $(1,1)$ nor of strong type (∞,∞).

5.3. The Hilbert transform on $L^p(\mathbb{R})$

So far we have defined our first singular integral operator, the Hilbert transform. This is an operator that is bounded on $L^2(\mathbb{R})$ and that has the representation

$$Hf(x) = \int_{\mathbb{R}} f(y)\frac{1}{x-y}dy,$$

whenever $f \in L^2(\mathbb{R})$ has compact support and $x \notin \mathrm{supp}(f)$. The function

$$K(x,y) = \frac{1}{x-y}$$

is the singular kernel associated with the Hilbert transform. Although we have seen that the Hilbert transform can be described for all $x \in \mathbb{R}$, at least for nice functions $f \in \mathcal{S}(\mathbb{R})$, the restricted representation just described is all we really need to execute our program. Furthermore, this approach will serve as a good introduction to the general case of Calderón-Zygmund operators. From the previous discussion we know that the Hilbert transform is not of strong type $(1,1)$ nor of type (∞,∞). The following theorem is the main result of the theory.

THEOREM 5.11. *Let H denote the Hilbert transform, initially defined on $L^2(\mathbb{R})$.*

(i) *The Hilbert transform is of weak type $(1,1)$; for $f \in L^1(\mathbb{R})$ we have that*
$$|\{x \in \mathbb{R} : |Hf(x)| > \lambda\}| \lesssim \frac{\|f\|_{L^1(\mathbb{R})}}{\lambda}, \quad \lambda > 0.$$
(ii) *For $1 < p < \infty$, the Hilbert transform is of strong type (p,p); for $f \in L^p(\mathbb{R})$ we have*
$$\|Hf\|_{L^p(\mathbb{R})} \lesssim_p \|f\|_{L^p(\mathbb{R})}.$$

PROOF. We will divide the proof in several steps. The most important one however is the proof of the weak type $(1,1)$. All the rest really relies on exploiting the symmetries of the Hilbert transform, interpolation and duality.

step 1; the weak $(1,1)$ bound: We fix a level $\lambda > 0$ and a function $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and write the Calderón-Zygmund decomposition of the function f at level λ in the form
$$f = g + b.$$
Recall that the "bad part" b is described as
$$b = \sum_{Q \in \mathcal{B}} b_Q$$
where $\mathcal{B}$ is a collection of disjoint dyadic intervals (since $n = 1$) and each b_Q is supported on Q. Furthermore we have that
$$\int_Q b_Q = 0,$$
and
$$\frac{1}{|Q|} \int_Q |b_Q| \lesssim \lambda.$$
Recall also that
$$|\cup_{Q \in \mathcal{B}} Q| \leq \frac{\|f\|_1}{\lambda},$$
by the maximal theorem. On the other hand the 'good part' g is bounded
$$\|g\|_\infty \lesssim \lambda$$
and its L^1 norm is controlled by the L^1 norm of f:
$$\|g\|_1 \leq \|f\|_1.$$
Observe that $g \in L^1 \cap L^\infty$ thus $g \in L^2(\mathbb{R})$ and by the log-convexity of the norm we have
$$\|g\|^2_{L^2(\mathbb{R})} \leq \|g\|_{L^1(\mathbb{R})} \|g\|_{L^\infty(\mathbb{R})} \lesssim \lambda \|f\|_{L^1(\mathbb{R})}. \tag{5.5}$$

REMARK 5.12. Since $f, g \in L^2(\mathbb{R})$ it follows that $b \in L^2(\mathbb{R})$ as well. Also, by the definition of the pieces b_Q it is easy to see that $b_Q \in L^2(Q)$ as well. However, we will not use the L^2 bounds on b nor on b_Q, the fact that they belong to L^2 being merely a technical assumption that allows us to define their Hilbert transforms. Overall, the hypothesis that $f \in L^2(\mathbb{R})$ cannot be used in any quantitative way if we ever want to extend our results to $L^1(\mathbb{R})$.

Since $f = b + g$ and H is linear, we have the following basic estimate

$$\{|Hf(x)| > \lambda\}| \leq |\{x \in \mathbb{R} : |H(g)(x)| > \lambda/2\}| + |\{|H(b)(x)| > \lambda/2\}|. \tag{5.6}$$

The part that corresponds to g is the easy one to estimate. This is not surprising since g is the good part. Since we already know that H is of strong type $(2,2)$ it is certainly of weak type $(2,2)$ thus we have

$$|\{x \in \mathbb{R} : |H(g)(x)| > \lambda/2\}| \lesssim \frac{\|g\|^2_{L^2(\mathbb{R})}}{\lambda^2} \leq \frac{\|f\|_{L^1(\mathbb{R})}}{\lambda},$$

by (5.5). This estimate takes care of the good part. Let's move now to the estimate for the bad part. The main ingredient for the estimate of the bad part is the following statement which we formulate as a lemma for future reference.

LEMMA 5.13. *Let $I = (x_o - \epsilon, x_o + \epsilon)$ be any interval in $\mathbb{R}$ and denote by I^* the interval with the same center as I and twice its length. For $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ supported in I and with zero mean on I, $\int_I f = 0$, we have*

$$|Hf(x)| \lesssim \frac{|I|}{|x - x_o|^2} \int_I |f|,$$

for all $x \notin I^*$. We conclude that

$$\int_{\mathbb{R}\setminus I*} |Hf(x)|dx \lesssim \int_I |f|.$$

REMARK 5.14. Here we require that f is also in $L^2(\mathbb{R})$ just in order to make sure that $Hf(x)$ is well defined. Note that in the case of the Hilbert transform it can be verified directly that $Hf(x)$ is well defined for $f \in L^1(I)$ and $x \notin I^*$. However we prefer this formulation since for more general Calderón-Zygmund operators we will only have a formula available to us for $f \in L^2(\mathbb{R})$ with compact support and $x \notin \mathrm{supp}(f)$.

PROOF. Using the zero mean value hypothesis for f we can write for $x \notin I^*$

$$\begin{aligned} |Hf(x)| &= \Big| \int_I \frac{f(y)}{x-y} dy \Big| = \Big| \int_I \Big(\frac{1}{x-y} - \frac{1}{x - x_o} \Big) f(y) dy \Big| \\ &\leq \int_I \frac{|y - x_o|}{|x - x_o||x - y|} |f(y)| dy. \end{aligned}$$

Now since $x \notin I^*$ we have that

$$|x - y| \geq |x - x_o| - |y - x_o| = |x - x_o| - \epsilon \geq |x - x_o| - |x - x_o|/2 = |x - x_o|/2$$

so we can write

$$|Hf(x)| \lesssim \frac{|I|}{|x - x_o|^2} \int_I |f(y)| dy,$$

as we wanted to show. The second claim of the lemma follows easily by integrating this estimate. □

We now go back to the estimate of b. First of all note that

$$|H(b)(x)| \leq \sum_{Q \in \mathcal{B}} |H(b_Q)(x)|,$$

for almost every $x \in \mathbb{R}$. Indeed, if we enumerate the intervals in $\mathcal{B}$ as $Q_1, \ldots, Q_N, \ldots$ then we have that $b_N(x) := \sum_{j=1}^N b_{Q_j}(x) \nearrow b(x)$ for every $x \in \mathbb{R}$ thus $b_N \to b$ in

$L^2(\mathbb{R})$. Since H is an isometry on $L^2(\mathbb{R})$ it follows that $H(b_N)$ converges to $H(b)$ in L^2 as well. Taking subsequences we then have that $H(b_{N_j})(x) \to H(b)(x)$ almost everywhere. Thus

$$|H(b_{N_j})(x)| = |\sum_{m=1}^{N_j} H(b_{Q_m})(x)| \le \sum_{Q\in\mathcal{B}} |H(b_Q)(x)|,$$

almost everywhere and we get the claim by letting $j \to +\infty$.

For each $Q \in \mathcal{B}$ let Q^* denote the interval with the same center and twice the side-length. We now estimate the 'bad part' as follows

$$|\{x \in \mathbb{R} : |H(b)(x)| > \lambda/2\}| \le |\cup_{Q\in\mathcal{B}} Q^*| + |\{x \notin \cup_{Q\in\mathcal{B}} Q^* : \sum_{Q\in\mathcal{B}} |H(b_Q)(x)| > \lambda/2\}|.$$

By the Calderón-Zygmund decomposition we have that

$$|\cup_{Q\in\mathcal{B}} Q^*| = 2|\cup_{Q\in\mathcal{B}} Q| \lesssim \frac{\|f\|_1}{\lambda},$$

which takes care of the first summand. For the second we use Lemma 5.13 to write

$$\int_{\mathbb{R}\setminus Q^*} |H(b_Q)(x)|dx \lesssim \int |b_Q(x)|dx \lesssim |Q|\lambda,$$

again by the Calderón-Zygmund decomposition. Observe that each $b_Q \in L^1(Q)\cap L^2(Q)$ and has mean zero on Q so the appeal to Lemma 5.13 is legitimate. Summing up the estimates for all the bad intervals in $\mathcal{B}$ we get

$$\Big\| \sum_{Q\in\mathcal{B}} |H(b_Q)| \Big\|_{L^1(\mathbb{R}\setminus\cup_{Q\in\mathcal{B}} Q^*)} \lesssim \lambda \sum_{Q\in\mathcal{B}} |Q| = \lambda \frac{\|f\|_1}{\lambda} = \|f\|_1.$$

By Chebyshev's inequality we thus get

$$|\{x \in \mathbb{R} \setminus \cup_{Q\in\mathcal{B}} Q^*) : \sum_{Q\in\mathcal{B}} |H(b_Q)(x)| > \lambda/2\}| \lesssim \frac{\|f\|_1}{\lambda}.$$

Summing up the estimates for the bad part we conclude that

$$|\{x \in \mathbb{R} : |H(b)(x)| > \lambda/2\}| \lesssim \frac{\|f\|_1}{\lambda}.$$

By (5.6) now we conclude that

$$|\{x \in \mathbb{R} : |Hf(x)| > \lambda\}| \lesssim \frac{\|f\|_1}{\lambda},$$

whenever $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

We have a priori assumed that $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ in order to have a good definition of H. However, the weak $(1,1)$ inequality on $L^1 \cap L^2$ allows us to extend the Hilbert transform to a linear operator on $L^1(\mathbb{R})$ which is also of weak type $(1,1)$. The details are left as an exercise.

EXERCISE 5.15. Let $T : L^1(\mathbb{R}^n) \cap \mathcal{S}(\mathbb{R}^n) \to L^1(\mathbb{R}^n)$ be a linear operator which is of weak type $(1,1)$. Show that T extends to a linear operator on $L^1(\mathbb{R}^n)$ which is of weak type $(1,1)$, with the same $(1,1)$ constant.

step 2; the strong (p,p) bound: As promised, the difficult part of the proof was the weak $(1,1)$ bound. The rest is routine. First of all observe that since H is of weak type $(1,1)$ and strong type $(2,2)$, the Marcinkiewicz interpolation theorem allow us to show that H is of strong type (p,p) for any $1<p<2$. To treat the interval $2<p<+\infty$ we argue by duality, exploiting the fact that H is almost self-adjoint (in fact it is skew adjoint as we have seen in Corollary 5.8). Indeed, let $f\in\mathcal{S}(\mathbb{R})$ and $2<p<\infty$. Now for any $g\in L^{p'}(\mathbb{R})$ we have

$$\Big|\int_{\mathbb{R}} Hf\bar{g}\Big| = \Big|\int_{\mathbb{R}} f\overline{H(g)}\Big| \leq \|f\|_{L^p(\mathbb{R})}\|H(g)\|_{L^{p'}(\mathbb{R})} \lesssim_p \|g\|_{L^{p'}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})},$$

using the fact that H is of strong type (p',p') since $1<p'<2$. Taking the supremum over all $g\in L^{p'}(\mathbb{R})$ with $\|g\|_{L^{p'}}\leq 1$ we get

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

for $2<p<\infty$ as well, whenever $f\in\mathcal{S}(\mathbb{R})$. Using standard arguments again this shows that H extends to a bounded linear operator on $L^p(\mathbb{R})$, $1<p<\infty$. □

REMARK 5.16. In fact, tracking the constants in the previous argument we see that

$$\|H\|_{L^p\to L^p} \lesssim \frac{1}{p-1} \quad \text{as} \quad p\to 1$$

and

$$\|H\|_{L^p\to L^p} \lesssim \frac{1}{p'-1} = \frac{p}{p'} \simeq p \quad \text{as} \quad p\to\infty.$$

Overall we have proved that H is of strong type (p,p) with a norm bound of the order

$$\|H\|_{L^p\to L^p} \lesssim \max((p-1)^{-1},p), \quad 1<p<\infty.$$

REMARK 5.17. We have exhibited that H extends to a bounded linear operator to L^p for $1<p<\infty$ and that it is of weak type $(1,1)$. However, for a general $f\in L^p(\mathbb{R})$, $1\leq p<+\infty$, there is no reason why Hf should by given by the same formula by which it was initially defined; remember that

$$Hf = \lim_{\epsilon\to 0}\int_{|y|>\epsilon}\frac{f(x-y)}{y}dy =: \lim_{\epsilon\to 0} H_\epsilon(f), \quad f\in\mathcal{S}(\mathbb{R}).$$

Thus the question whether $H_\epsilon(f)(x)\to Hf(x)$ a.e., for $f\in L^p(\mathbb{R})$, is very natural. Since we know this convergence is true for the dense subset §($\mathbb{R}$), the study of the pointwise convergence amounts to studying the boundedness properties of the corresponding maximal operator

$$H^*(f)(x) := \sup_{\epsilon>0}\int_{|y|>\epsilon}\frac{f(x-y)}{y}dy.$$

Thus if one can show that H^* is of weak type $(1,1)$ for example, the pointwise convergence of $H_\epsilon(f)$ to Hf would follow by Proposition 4.5. Such an estimate is actually true and thus this formula extends to all L^p functions for $1\leq p<\infty$. We will however see this in the general theory of Calderón-Zygmund operators of which the Hilbert transform is a special case and so we postpone the proof until then.

5.3.1. The Hilbert transform and the boundary values of holomorphic functions. In this section we briefly discuss the connection of the Hilbert transform with the boundary values of holomorphic functions in the upper half plane. Let us write

$$\mathbb{R}^2_+ = \mathbb{C}_+ = \{(x, y) : x \in \mathbb{R}, y > 0\} = \{x + iy : x \in \mathbb{R}, y > 0\},$$

for the upper half plane. Two function u, v on $\mathbb{R}_+$ are called *conjugate harmonic functions* if they are the real and imaginary part respectively of a holomorphic function $F(z)$ in the upper half plane, where $z = x + iy$. Thus we have that

$$F(z) = F(x + iy) = u(x, y) + iv(x, y).$$

By definition both u, v are real and harmonic. Moreover, they satisfy the *Cauchy-Riemann* equations (since F is holomorphic). Now assume that F has a boundary value $F_o(x) = u_o(x) + iv_o(x)$ on the real line $x \in \mathbb{R}$. Then

$$v_o(x) = H(u_o)(x), \quad \text{and} \quad u_o(x) = -H(v_o)(x).$$

Of course, some technical assumptions are needed to make all these claims rigorous as for example assuming that the holomorphic function F has some decay of the form $|F(z)| \lesssim (1 + |z|)^{-1}$ in the upper half plane.

Conversely, Let $f \in L^p(\mathbb{R})$ be a real function and $P_y(x)$ be the Poisson kernel for the upper half plane

$$P_y(x) = \frac{1}{\pi}\frac{y}{y^2 + x^2}.$$

As we have seen, the convolution $u(x, t) = (f * P_y)(x)$ is a harmonic function in the upper half plane $\mathbb{R}_+ = \{(x, t) : x \in \mathbb{R}, t > 0\}$. Observe that

$$u(x, y) = \frac{y}{\pi}\int_{\mathbb{R}} \frac{f(t)}{y^2 + (x - t)^2} dt.$$

Consider now the *conjugate Poisson kernel*

$$Q_t(x, y) = \frac{1}{\pi}\frac{x}{y^2 + x^2}.$$

The name comes from the fact that both P_t, Q_t are both real harmonic functions and writing $z = x + iy$ we have

$$P_t(x) + iQ_t(x) = \frac{1}{\pi}\frac{ix + y}{x^2 + y^2} = \frac{i}{\pi}\frac{x - iy}{x^2 + y^2} = \frac{i}{\pi z},$$

which is holomorphic in the upper half plane. Thus P_t, Q_t are conjugate harmonic functions which is what makes the functions u, v conjugate harmonic functions as well. We conclude that the function

$$v(x, y) = (f * Q_t)(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{f(t)(x - t)}{y^2 + (x - t)^2} dt,$$

is harmonic in the upper half plane and that

$$F(z) = u(x, y) + iv(x, y), \quad z = x + iy \in \mathbb{C}_+,$$

is holomorphic in the upper half plane.

Finally observe that according to the previous formulae we have

$$F(z) = u(x, y) + iv(x, y) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{f(t)[y + i(x - t)]}{y^2 + (x - t)^2} dt = \frac{1}{\pi i}\int_{\mathbb{R}} \frac{f(t)}{t - x - iy} dt.$$

In this language, Proposition 5.4 just states that $F(x+iy)$ converges to its boundary value $f+iHf$ as $y\to 0$. We also see that the imaginary part of F converges to the Hilbert transform:

$$\lim_{y\to 0}(f*Q_t)(x)=Hf(x),$$

both in $L^p(\mathbb{R})$ and almost everywhere.

5.3.2. Frequency cut-off multipliers and partial Fourier integrals. Remember that for a bounded function $m\in L^\infty(\mathbb{R})$ the operator

$$T:L^2(\mathbb{R})\to L^2(\mathbb{R}),\quad \widehat{T(f)}(\xi)=m(\xi)\hat{f}(\xi)$$

is a multiplier operator (associated to the multiplier m) and that $\|T\|_{L^2\to L^2}=\|m\|_{L^\infty(\mathbb{R})}$. We also say that m is a multiplier on L^p if T extends to a bounded linear operator $T:L^p(\mathbb{R})\to L^p(\mathbb{R})$. Thus we see that the Hilbert transform is a multiplier operator on $L^p(\mathbb{R})$ associated with the multiplier

$$m(\xi)=-i\operatorname{sgn}(\xi),\quad \xi\in\mathbb{R},$$

which is obviously a bounded function with $\|m\|_{L^\infty(\mathbb{R})}=1$. A very closely related multiplier is the *frequency cutoff multiplier*. Given an interval (a,b) in the frequency space, where $a<b$, we define the operator $S_{(a,b)}:L^2(\mathbb{R})\to L^2(\mathbb{R})$ by means of the formula

$$\widehat{S_{(a,b)}f}(\xi):=\mathbf{1}_{(a,b)}(\xi)\hat{f}(\xi).$$

Thus the operator $S_{(a,b)}$ applied to f, localizes the function f in frequency, in the interval (a,b). Such operators as well as their multidimensional analogues turn out to be very important in harmonic analysis as well as in the theory of partial differential operators. Obviously $S_{(a,b)}$ is bounded on $L^2(\mathbb{R})$, since $\|S_{(a,b)}\|_{L^2\to L^2}=\|\chi_{(a,b)}\|_{L^\infty(\mathbb{R})}=1$. However, the corresponding estimate in $L^p(\mathbb{R})$ is far from obvious. After all the work we have done for the Hilbert transform though, we can get the L^p bounds for $S_{(a,b)}$ as a simple corollary. This is based on the observation that

$$S_{(a,b)}=\frac{i}{2}(\mathrm{Mod}_aH\mathrm{Mod}_{-a}-\mathrm{Mod}_bH\mathrm{Mod}_{-b}),\tag{5.7}$$

where the equality should be understood as an equality of operator in $L^2(\mathbb{R})$. Here remember that

$$\mathrm{Mod}_{x_o}(f)(x)=e^{2\pi i x_o x}f(x).$$

The verification of this formula is left as an exercise. Formula (5.7) is also true when $a=-\infty$ or $b=+\infty$ with obvious modifications.

EXERCISE 5.18. Prove formula (5.7).

A simple corollary of the L^p boundedness of the Hilbert transform is the corresponding statement for $S_{(a,b)}$.

LEMMA 5.19. *The operator $S_{(a,b)}$ is of strong type (p,p) for $1<p<\infty$:*

$$\|S_{(a,b)}(f)\|_{L^p(\mathbb{R})}\lesssim_p\|f\|_{L^p(\mathbb{R})}.$$

Note that the operator norm of $S_{(a,b)}$ does not depend on a,b.

Now for $N > 0$ and $f \in \mathcal{S}(\mathbb{R})$ define the partial Fourier integral operator

$$S_N(f)(x) := \int_{-N}^{N} \hat{f}(\xi) e^{2\pi i x\xi} d\xi \int_{\mathbb{R}} \mathbf{1}_{(-N,N)}(\xi) \hat{f}(\xi) e^{2\pi i x\xi} d\xi, \quad x \in \mathbb{R}.$$

Observe that these integrals are the $\mathbf{1}_{(-N,N)}$-means of the integral $\int \hat{f}(\xi)e^{2\pi i x\xi}d\xi$. We have seen that the Gauss-Weierstrass or Abel means of this integral converge to f, both almost everywhere as well as in the L^p sense. However the function $\mathbf{1}_{(-N,N)}$ is much rougher. We still have the following theorem as a consequence of the (p,p) bound for the Hilbert transform.

THEOREM 5.20. *For $1 < p < \infty$ the operator S_N has a unique extension to a bounded linear operator on $L^p(\mathbb{R})$ for $1 < p < \infty$.*

However the L^p boundedness of S_N controls the L^p convergence of partial Fourier integrals.

LEMMA 5.21. *The partial Fourier integrals $S_N(f)$ converge to f in the L^p norm for $1 < p < \infty$ if and only if S_N is of strong type (p,p) uniformly in N.*

Now Theorem 5.20 and Lemma 5.21 immediately imply:

COROLLARY 5.22. *For $1 < p < \infty$ the partial Fourier integrals $S_N(f)$ converge to f in the L^p norm.*

The question whether $S_N(f)$ converges to f almost everywhere is much harder. For $f \in L^p(\mathbb{R})$, $1 < p < +\infty$, the answer is positive and this is the content of the famous Carleson-Hunt theorem. This theorem was first proved by Carleson for L^2 and then extended to L^p by Hunt. A counterexample by Kolmogorov shows that both in the L^1 sense as well a almost everywhere, the convergence of the partial Fourier integrals fails for $f \in L^1$.

EXERCISE 5.23. Show that S_N extends to an operator of weak type $(1,1)$ on $L^1(\mathbb{R})$ and that the partial Fourier integrals converge to f *in measure* for $f \in L^1(\mathbb{R})$. Conclude that for almost every $x \in \mathbb{R}$ there is a subsequence $\{N_k\}_k$ such that $S_{N_k}(f)(x) \to f(x)$ as $k \to +\infty$.

CHAPTER 6

Calderón-Zygmund operators

After having studied the Hilbert transform in detail we now move to the study of general Calderón-Zygmund operators, that is operators given formally as

$$Tf(x) = \int K(x,y)f(y)dy,$$

for an appropriate kernel K. Let us quickly review what we used in order to show that the Hilbert transform H is of weak type $(1,1)$ and strong type (p,p). First of all we have essentially used the fact that the linear operator H is defined on L^2 and bounded, that is, that it is of strong type $(2,2)$. This information was used in two different ways. First of all, the fact that H is defined on L^2 means that it is defined on a dense subspace of L^p for every $1 \leq p < +\infty$. Furthermore, the boundedness of the Hilbert transform on L^2 allowed us to treat the set $\{|H(g)| > \lambda\}$ where g is the "good part" in the Calderón-Zygmund decomposition of a function f. Secondly, we used the fact that there is a specific representation of the operator H of the form

$$H(f)(x) = \int K(x,y)f(y)dy,$$

whenever $f \in L^2$ and has compact support and $x \notin \text{supp}(f)$. For the Hilbert transform we had that the kernel K is given as

$$K(x,y) = \frac{1}{x-y}.$$

We used the previous representation and the formula of K to prove a sort of restricted L^1 boundedness of H on functions which are localized and have mean zero, which is the content of Lemma 5.13. This in turn allowed us to treat the "bad part" of the Calderón-Zygmund decomposition of f. From the proof of that lemma it is obvious that what we really need for K is a Hölder type condition. Note as well that for the Hilbert transform we first proved the L^p bounds for $1 < p < 2$ and then the corresponding boundedness for $2 < p < \infty$ followed by the fact that H is essentially self-adjoint.

6.1. Singular kernels and Calderón-Zygmund operators

We will now define the class of Calderón-Zygmund operators in such a way that we will be able to repeat the schedule used for the Hilbert transform. We begin by defining an appropriate class of kernels K, namely the singular (or standard) kernels.

DEFINITION 6.1 (Singular or Standard kernels). A *singular* (or *standard*) kernel is a function $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$, defined away from the diagonal $x = y$, which satisfies the decay estimate

$$|K(x,y)| \lesssim_n |x-y|^{-n}, \tag{6.1}$$

for $x \neq y$ and the Hölder-type regularity estimates

$$|K(x,y_1) - K(x,y)| \lesssim_{n,\sigma} \frac{|y-y_1|^\sigma}{|x-y|^{n+\sigma}} \quad \text{if} \quad |y-y_1| < \frac{1}{2}|x-y|, \tag{6.2}$$

and

$$|K(x_1,y) - K(x,y)| \lesssim_{n,\sigma} \frac{|x-x_1|^\sigma}{|x-y|^{n+\sigma}} \quad \text{if} \quad |x-x_1| < \frac{1}{2}|x-y|, \tag{6.3}$$

for some Hölder exponent $0 < \sigma \leq 1$.

EXAMPLE 6.2. Let $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given as $K(x,y) = (x-y)^{-1}$ for $x, y \in \mathbb{R}$ with $x \neq y$. Then K is a singular kernel. Observe that K is the singular kernel associated with the Hilbert transform.

EXAMPLE 6.3. Let $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be given as

$$K(x,y) = \Omega\Big(\frac{x-y}{|x-y|}\Big)|x-y|^{-n},$$

where $\Omega : S^{n-1} \to \mathbb{C}$ is a Hölder-continuous function:

$$|\Omega(x') - \Omega(y')| \lesssim_{n,\sigma} |x'-y'|^\sigma,$$

for some $0 < \sigma \leq 1$. Then K is a singular kernel.

EXERCISE 6.4. Prove that the kernel K of example 6.3 is a singular kernel.

EXAMPLE 6.5. Let $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ satisfy the size estimate

$$|K(x,y)| \lesssim_n |x-y|^{-n},$$

and the regularity estimates

$$|\nabla_x K(x,y)| \lesssim_n |x-y|^{-(n+1)}, \quad |\nabla_y K(x,y)| \lesssim_n |x-y|^{-(n+1)},$$

away from the diagonal $x = y$. Then K is a singular kernel. In particular, the kernel $K : \mathbb{R}^n \to \mathbb{R}^n \to \mathbb{C}$ given as

$$K(x,y) = |x-y|^{-n},$$

is a singular kernel since the gradient of K is of the order $|x-y|^{-(n+1)}$. Thus the estimates (6.2) and (6.3) are consistent with (6.1) but of course do not follow from it.

REMARK 6.6. The constant $\frac{1}{2}$ appearing in (6.2), (6.3) is inessential. The conditions are equivalent with the corresponding conditions where $\frac{1}{2}$ is replaced by any constant between zero and one.

We are now ready to define Calderón-Zygmund operators.

DEFINITION 6.7 (Calderón-Zygmund operators). A *Calderón-Zygmund operator* (in short *CZO*) is a linear operator $T : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ which is bounded on $L^2(\mathbb{R}^n)$:

$$\|Tf\|_{L^2(\mathbb{R}^n)} \lesssim_{T,n} \|f\|_{L^2(\mathbb{R}^n)} \quad \text{for all} \quad f \in L^2(\mathbb{R}^n),$$

and such that there exists a singular kernel K for which we have

$$Tf(x) = \int_{\mathbb{R}^n} K(x,y) f(y) dy,$$

for all $f \in L^2(\mathbb{R}^n)$ with compact support and $x \notin \operatorname{supp}(f)$.

REMARK 6.8. Note that the integral $\int K(x,y)f(y)$ converges absolutely whenever $f \in L^2(\mathbb{R}^n)$ has compact support and x lies outside the support of f. Indeed,

$$\begin{aligned} \int_{\mathbb{R}^n} |K(x,y)||f(y)| dy &\leq \left(\int_{y \notin \operatorname{supp}(f)} |K(x,y)|^2 dy \right)^{\frac{1}{2}} \|f\|_{L^2(\mathbb{R}^n)} \\ &\leq \left(\int_{|x-y|\geq \delta} \frac{1}{|x-y|^{2n}} dy \right)^{\frac{1}{2}} \|f\|_{L^2(\mathbb{R}^n)} \end{aligned}$$

by (6.1), for some $\delta > 0$. Observe that the integral in the last estimate converges.

REMARK 6.9. For any singular kernel K one can define T_K by means of

$$T_K(f)(x) = \int_{\mathbb{R}^n} K(x,y) f(y) dy,$$

for $f \in L^2(\mathbb{R}^n)$ with compact support and $x \notin \operatorname{supp}(f)$. It is not necessary however that T_K is a CZO since it might fail to be bounded on $L^2(\mathbb{R}^n)$.

REMARK 6.10. It is not hard to see that T *uniquely* determines the kernel K. That is if

$$Tf(x) = \int_{\mathbb{R}^n} K(x,y) f(y) dy = \int_{\mathbb{R}^n} K_1(x,y) f(y) dy,$$

for all $f \in L^2(\mathbb{R}^n)$ with compact support, then $K = K_1$ almost everywhere (why?). The opposite is not true. Indeed, for any bounded function $b \in L^\infty(\mathbb{R}^n)$ the operator defined as $Tf(x) = b(x)f(x)$ is a Calderón-Zygmund kernel with kernel zero. A more specific example is the identity operator which also falls in the previous class, and is CZO with kernel 0. However, this is the only ambiguity. See Exercise 6.11.

EXERCISE 6.11. Let T_1, T_2 be two CZO's with the same singular kernel K. Show that there exists a bounded function $b \in L^\infty(\mathbb{R}^n)$ such that

$$T_1 f = T_2 f + bf,$$

for all $f \in L^2(\mathbb{R}^n)$.

If T is a CZO, the definition already contains the fact that T is defined and bounded on $L^2(\mathbb{R}^n)$, so we don't need to worry about that. The next step is to establish the restricted L^1 boundedness for L^1 functions with mean zero. The following lemma is the analogue of Lemma 5.13.

LEMMA 6.12. *Let $Q = Q(z, R)$ be a Euclidean cube in $\mathbb{R}^n$ with center $z \in \mathbb{R}^n$ and sidelength R, and denote by Q^* the cube with the same center and $1 + 2\sqrt{n}$ times the sidelength of Q, that is $Q^* = Q(z, (1 + 2\sqrt{n})R)$. Let $f \in L^1(Q)$ have mean zero, that is $\int_Q f = 0$. Then we have that*

$$|Tf(x)| \lesssim_{n,\sigma} \frac{R^\sigma}{|x-z|^{n+\sigma}} \int_Q |f(y)|dy,$$

for all $x \notin Q^*$. We conclude that

$$\|Tf\|_{L^1(\mathbb{R}^n \setminus Q^*)} \lesssim_{n,\sigma} \|f\|_{L^1(Q)}.$$

PROOF. Using the fact that f has zero mean on B, for $x \notin B^*$ we can estimate

$$\begin{aligned}|Tf(x)| &\leq \int_Q |K(x,y) - K(x,z)||f(y)|dy \leq \int_Q \frac{|y-z|^\sigma}{|x-y|^{n+\sigma}}|f(y)|dy \\ &\lesssim_{n,\sigma} \frac{R^\sigma}{|x-z|^{n+\sigma}} \int_Q |f(y)|dy.\end{aligned}$$

The last estimate follows since

$$|x-y| \geq |x-z| - |y-z| \geq |x-z| - \frac{1}{2}\sqrt{n}R \geq \frac{1}{2}|x-z|.$$

Now observe that $Q \subset B(z, (1+2\sqrt{n})R) \subset Q^*$. Integrating throughout $\mathbb{R}^n \setminus Q^*$ we get

$$\begin{aligned}\int_{\mathbb{R}^n \setminus Q^*} |Tf(x)|dx &\leq R^\sigma \int_Q |f(y)|dy \int_{\mathbb{R}^n \setminus B(z,(1+2\sqrt{n})R)} \frac{1}{|x-z|^{n+\sigma}}dx \\ &\lesssim_n R^\sigma \int_Q |f(y)|dy \int_{(1+2\sqrt{n})R}^\infty \frac{r^{n-1}}{r^{n+\sigma}}dr \simeq_{n,\sigma} \int_Q |f(y)|dy\end{aligned}$$

which is the desired estimate. □

The only thing missing in order to conclude the proof of the L^p bounds for CZO's is the the fact that they are self adjoint *as a class*. In particular, we need the following.

LEMMA 6.13. *Let T be a CZO. Consider the adjoint T^* defined by means of*

$$\int Tf\bar{g} = \int f\overline{T^*(g)}, \tag{6.4}$$

for all f, g in L^2. Then T^* is a CZO.

PROOF. It is immediate from (6.4) and the fact that T is bounded on L^2 that T^* is also bounded on L^2 with the same norm. Now let $f, g \in L^2(\mathbb{R}^n)$ have disjoint compact supports. We have

$$\int Tf\bar{g} = \int\int K(x,y)f(y)dy\, \bar{g}(x)dx = \int f(y)\overline{\int \overline{K(x,y)}g(x)dx}\, dy. \tag{6.5}$$

Let $z \notin \operatorname{supp}(g)$ and $\phi \in C_c^\infty(\mathbb{R}^n)$ have support inside $B(0,1)$ with $\int \phi = 1$. For $\epsilon > 0$, the functions $\phi_\epsilon(y-z)$ are supported in $B(z,\epsilon)$ so, for ϵ small enough, the support of ϕ_ϵ is disjoint from the support of g. By (6.5) we conclude that

$$\int \phi_\epsilon(z-y)\overline{T^*(g)(y)}dy = \int \phi_\epsilon(z-y)\overline{\int \overline{K(x,y)}g(x)dx}\, dy.$$

Letting $\epsilon \to 0$ we get

$$T^*(g)(z) = \int \overline{K(x,z)} g(x) dx,$$

for almost every $z \notin \operatorname{supp}(g)$. Since the conditions defining singular kernels are symmetric in the variables x, y, the kernel $S(x,y) := \overline{K(y,x)}$ is again a singular kernel so we are done. □

The discussion above leads to the main theorem for CZO's. Given Lemmas 6.12, 6.13, the proof follows by using the Calderón-Zygmund decomposition, exactly as in the case of the Hilbert transform in § 5.3

THEOREM 6.14. *Let T be a Calderón-Zygmund operator. Then T extends to a linear operator which is of weak type $(1,1)$ and of strong type (p,p) for all $1 < p < +\infty$, where the corresponding norms depend only on n, σ and p.*

6.2. Pointwise convergence and maximal truncations

Let T be a CZO. The example of the Hilbert transform suggests that we should have the almost everywhere convergence

$$Tf(x) = \lim_{\epsilon\to 0} \int_{|x-y|>\epsilon} K(x,y) f(y) dy,$$

at least for nice functions $f \in \mathcal{S}(\mathbb{R}^n)$. The truncated operators

$$T_\epsilon f(x) := \int_{|x-y|>\epsilon} K(x,y) f(y) dy,$$

certainly make sense for $f \in L^2(\mathbb{R}^n)$ because of (6.1). However, the pointwise limit of the truncations, $\lim_{\epsilon\to 0} T_\epsilon(f)(x)$, need not even exist in general or may exist and be different from $Tf(x)$. Here we can use the trivial example of the operator $Tf(x) = b(x)f(x)$. As we have already observed this is a CZO operator with kernel 0. Thus $T_\epsilon(f)(x) = 0$ for all $\epsilon > 0$ but clearly $Tf \neq 0$ in general.

The following lemma clears out the situation as far as the existence of the limit is concerned:

LEMMA 6.15. *The limit*

$$\lim_{\epsilon\to 0} T_\epsilon f(x),$$

exists almost everywhere for all $f \in \mathcal{S}(\mathbb{R}^n)$ if and only if the limit

$$\lim_{\epsilon\to 0} \int_{\epsilon<|x-y|<1} K(x,y) dy,$$

exists almost everywhere.

PROOF. First suppose that the limit $\lim_{\epsilon\to 0} T_\epsilon f(x)$ exists for all $f \in \mathcal{S}(\mathbb{R}^n)$ and let $\phi \in \mathcal{S}(\mathbb{R}^n)$ with $\phi \equiv 1$ on $B(0,1)$. Then

$$\lim_{\epsilon\to 0} T_\epsilon \phi(x) = \lim_{\epsilon\to 0} \int_{\epsilon<|x-y|<1} K(x,y) dy + \int_{|x-y|>1} K(x,y)\phi(y) dy.$$

Observe that by (6.1) the second integral on the right hands side converges absolutely. Since the limit on the left hand side exists we conclude that the limit on the right hand side exists as well. Conversely, suppose that the limit

$$\lim_{\epsilon\to 0}\int_{\epsilon<|x-y|<1} K(x,y)dy = L$$

exists and let $f\in\mathcal{S}(\mathbb{R}^n)$. We have that

$$\begin{aligned} T_\epsilon f(x) &= \int_{\epsilon<|x-y|<1} K(x,y)f(y)dy + \int_{|x-y|>1} K(x,y)f(y)dy \\ &= \int_{\epsilon<|x-y|<1} K(x,y)[f(y)-f(x)]dy + f(x)\int_{\epsilon<|x-y|<1} K(x,y)dy \\ &\quad + \int_{|x-y|>1} K(x,y)f(y)dy =: I_1(\epsilon)+I_2(\epsilon)+I_3. \end{aligned}$$

By the same considerations are before $|I_3|$ is a positive number that does not depend on ϵ. By the hypothesis we also have that $\lim_{\epsilon\to 0} I_2(\epsilon) = Lf(x)$. Finally for $I_1(\epsilon)$ observe that we have

$$\int_{0<|x-y|<1} |K(x,y)||x-y|dy \lesssim_n \int_{|x-y|<1} |x-y|^{-(n-1)}dy \lesssim_n 1,$$

by (6.1). Since

$$|K(x,y)[f(x)-f(y)]| \lesssim \|\nabla f\|_{L^\infty(\mathbb{R}^n)}|K(x,y)||x-y|,$$

dominated convergence implies that $\lim_{\epsilon\to 0} I_1(\epsilon)$ exists as well. □

Thus, for specific kernels K one has an easy criterion to establish whether the limit $\lim_{\epsilon\to 0} T_\epsilon(f)$ exists a.e. for "nice" functions f. For example, for the kernel $K(x,y)=(x-y)^{-1}$ of the Hilbert transform, the existence of the limit

$$\lim_{\epsilon\to 0}\int_{\epsilon<|x-y|<1}\frac{1}{x-y}dy = 0$$

is obvious. In order to extend the almost everywhere convergence to the class $L^p(\mathbb{R}^n)$ we need to consider the corresponding maximal function.

DEFINITION 6.16. Let T be a CZO and define the truncations of T as before

$$T_\epsilon f(x) := \int_{|x-y|>\epsilon} K(x,y)f(y)dy, \quad x\in\mathbb{R}^n, \quad f\in\mathcal{S}(\mathbb{R}^n).$$

The *maximal truncation* of T is the sublinear operator defined as

$$T_*(f)(x) := \sup_{\epsilon>0}|T_\epsilon(f)(x)|, \quad x\in\mathbb{R}^n.$$

The maximal truncation of a CZO has the same continuity properties as T itself.

THEOREM 6.17. *Let T be a CZO and T_* denote its maximal truncation. Then T_* is of weak type $(1,1)$ and strong type (p,p) for $1<p<\infty$.*

The proof of Theorem 6.17 depends on the following preliminary results.

LEMMA 6.18 (Kolmogorov). *Let S be an operator of weak type $(1,1)$ and $\nu \in (0,1)$. Then for every set $E \subset \mathbb{R}^n$ with $0 < |E| < +\infty$ we have that*

$$\int_E |Sf(x)|^\nu dx \lesssim_{\nu,S} |E|^{1-\nu} \|f\|_1^\nu.$$

The proof of this lemma is a simple application of the representation of the L^ν norm in terms of level sets and is left as an exercise.

EXERCISE 6.19. Prove Lemma 6.18 above.

The second result we need is the following lemma that gives a pointwise control of the maximal truncations of the CZO T by an expression that involves the maximal function of f and the maximal function of Tf.

LEMMA 6.20. *Let T be a CZO and $0 < \nu \leq 1$. Then for all $f \in C_c^\infty(\mathbb{R}^n)$ we have that*

$$T_* f(x) \lesssim_{\nu,n,\sigma} [M(T|f|^\nu)(x)]^{\frac{1}{\nu}} + Mf(x).$$

PROOF. Let us fix a function $f \in C_c^\infty(\mathbb{R}^n)$ and $\epsilon > 0$ and consider the balls $B = B(x, \epsilon/2)$ and its double $B^* = B(x, \epsilon)$. We decompose f in the form

$$f = f\mathbf{1}_{B^*} + f(1 - \mathbf{1}_{B^*}) =: f_1 + f_2.$$

Since $\operatorname{supp}(f_2) \cap B = \emptyset$ and obviously $f_2 \in L^2(\mathbb{R}^n)$ has compact support we can write

$$Tf_2(x) = \int_{\mathbb{R}^n} K(x,y) f_2(y) dy = \int_{|x-y|>\epsilon} K(x,y) f(y) dy = T_\epsilon f(x). \tag{6.6}$$

Also every $w \in B$ is not contained in the support of f_2 thus

$$\begin{aligned}
|Tf_2(w) - Tf_2(x)| &= \left| \int_{|x-y|>\epsilon} [K(x,y) - K(w,y)] f_2(y) dy \right| \\
&\leq \int_{|x-y|>\epsilon} \frac{|x-w|^\sigma}{|x-y|^{n+\sigma}} |f(y)| dy,
\end{aligned}$$

by (6.3), since $|x - w| < \frac{\epsilon}{2} < \frac{1}{2}|x - y|$ for y in the domain of integration above. By this estimate we get that

$$\begin{aligned}
|T(f_2)(w) - T(f_2)(x)| &\lesssim_\sigma \epsilon^\sigma \sum_{k=0}^\infty \int_{2^k\epsilon<|x-y|<2^{k+1}\epsilon} \frac{|f(y)|}{(2^k\epsilon)^{n+\sigma}} dy \\
&\lesssim_\sigma \sum_{k=0}^\infty \frac{1}{\epsilon^n} \frac{1}{2^{k(n+\sigma)}} \int_{|x-y|<2^{k+1}\epsilon} |f(y)| dy \\
&\lesssim_{\sigma,n} \sum_{k=0}^\infty \frac{1}{2^{k\sigma}} Mf(x) \simeq_{n,\sigma} Mf(x).
\end{aligned}$$

Combining the previous estimates we conclude that for any $w \in B$

$$|T_\epsilon f(x)| \leq AMf(x) + |Tf_2(w)| \leq AMf(x) + |Tf(w)| + |Tf_1(w)|, \tag{6.7}$$

for some constant A depending only on n and σ.

We now deal with the cases $\nu = 1$ and $\nu < 1$ separately.

Case $\nu = 1$. If $T_\epsilon(f)(x) = 0$ then we are done. If $|T_\epsilon(f)(x)| > 0$ then there is $\lambda > 0$ such that $|T_\epsilon(f)(x)| > \lambda$. Let

$$B_1 = \{w \in B : |Tf(w)| > \lambda/3\},$$
$$B_2 = \{w \in B : |Tf_1(w)| > \lambda/3\},$$

and

$$B_3 = \begin{cases} \emptyset, & \text{if} \quad M(f)(x) \leq A^{-1}\lambda/3, \\ B, & \text{if} \quad M(f)(x) > A^{-1}\lambda/3 \end{cases}.$$

Let $w \in B$. Then either $w \in B_1$ or $w \in B_2$ or $AMf(x) > \lambda/3$. In the last case $B_3 = B$ so in every case we conclude that $w \in B_1 \cup B_2 \cup B_3$ thus $B \subset B_1 \cup B_2 \cup B_3$. However we have that

$$|B_1| \lesssim \frac{1}{\lambda}\int_B |Tf(y)|dy \leq \frac{|B|}{\lambda}M(Tf)(x).$$

Also, by the $(1,1)$ type of T we get

$$|B_2| \lesssim \frac{1}{\lambda}\|f_1\|_{L^1(\mathbb{R}^n)} = \frac{1}{\lambda}\int_B |f(y)|dy \leq \frac{|B|}{\lambda}M(f)(x).$$

Finally, if $B_3 = B$ then $\lambda \lesssim_{n,\sigma} M(f)(x)$. Otherwise $B_3 = \emptyset$ so

$$|B| \leq |B_1| + |B_2| \lesssim_{n,\sigma} \frac{|B|}{\lambda}(M(Tf)(x) + M(f)(x)).$$

Thus in every case we get that

$$\lambda \lesssim_{n,\sigma} M(Tf)(x) + M(f)(x).$$

Since the previous estimate is true for any $\lambda < T_\epsilon(f)(x)$ we conclude that

$$T_\epsilon(f)(x) \lesssim_{n,\sigma} M(Tf)(x) + M(f)(x),$$

which gives the desired estimate in the case $\nu = 1$.

Case $\nu < 1$. For $\nu < 1$ estimate (6.7) implies that

$$|T_\epsilon(f)(x)|^\nu \lesssim_{\sigma,\nu,n} |M(f)(x)|^\nu + |T(f)(w)|^\nu + |T(f_1)(w)|^\nu,$$

and integrate in $w \in B$ to get

$$|T_\epsilon(f)(x)|^\nu \lesssim_{\sigma,\nu,n} |M(f)(x)|^\nu + \frac{1}{|B|}\int_B |Tf(w)|^\nu dw + \frac{1}{|B|}\int_B |Tf_1(w)|^\nu dw,$$

and thus

$$|T_\epsilon(f)(x)| \lesssim_{\sigma,\nu,n} |M(f)(x)| + \left(\frac{1}{|B|}\int_B |Tf(w)|^\nu dw\right)^{\frac{1}{\nu}} + \left(\frac{1}{|B|}\int_B |Tf_1(w)|^\nu dw\right)^{\frac{1}{\nu}},$$

Note that

$$\left(\frac{1}{|B|}\int_B |Tf(w)|^\nu dw\right) \leq [M(|Tf|^\nu)(x)]^{\frac{1}{\nu}},$$

and by Lemma 6.18 the last term is controlled by

$$\left(\frac{1}{|B|}\int_B |Tf_1(w)|^\nu dw\right)^{\frac{1}{\nu}} \leq \frac{1}{|B|}\|f_1\|_1 \leq M(f)(x),$$

since T is of weak type $(1,1)$. Gathering these estimates we get

$$T_\epsilon(f)(x) \lesssim_{\sigma,\nu,n} M(f)(x) + [M(|Tf|^\nu)(x)]^{\frac{1}{\nu}},$$

as we wanted to show. □

The operator $f \mapsto [M(|Tf|^\nu)]^{\frac{1}{\nu}}$ that appears in the previous lemma is obviously bounded on $L^p(\mathbb{R}^n)$ for $1 < p < +\infty$. The following lemma shows that it is also well behaved at the endpoint $p = 1$, at least when $f \in C_c^\infty(\mathbb{R}^n)$.

LEMMA 6.21. *Suppose that S is a sublinear operator which is of weak type $(1,1)$ and let $0 < \nu < 1$. Then for all $f \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ we have*

$$\left|\left\{x \in \mathbb{R}^n : [M(|Sf|^\nu)]^{\frac{1}{\nu}}\right\}\right| \lesssim_n \frac{1}{\lambda}\|f\|_{L^1(\mathbb{R}^n)}.$$

PROOF. We first argue that the statement of the lemma is true for the operator $f \mapsto [M_{\mathcal{D}}(|Sf|^\nu)]^{\frac{1}{\nu}}$, where $M_{\mathcal{D}}$ is the dyadic maximal operator. Indeed, using the Calderón-Zygmund decomposition it is not hard to see (Exercise 6.22) that

$$|\{x \in \mathbb{R}^n : M_{\mathcal{D}}g(x) > \lambda\}| \leq \frac{1}{\lambda}\int_{\{M_{\mathcal{D}}g(x)>\lambda\}} |g(x)|dx,$$

where $M_{\mathcal{D}}$ is the dyadic maximal function. Applying the last estimate to the function $g(x) := [M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}}$ we get

$$|\{x \in \mathbb{R}^n : [M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}| \leq \frac{1}{\lambda^\nu}\int_{\{[M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}}>\lambda\}} |Sf(x)|^\nu dx.$$

For $f \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ the set $\{[M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}$ has finite measure. This is because

$$[M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} \leq [M_{\mathcal{D}}(|Sf|^{p\nu})(x)]^{\frac{1}{p\nu}}$$

for any $p > 1$. Using this estimate with $p = q/\nu$ for some $q > \nu$ we have

$$\begin{aligned}|\{x \in \mathbb{R}^n : [M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}| &\leq |\{x \in \mathbb{R}^n : [M_{\mathcal{D}}(|Sf|^q)(x)]^{\frac{1}{q}} > \lambda\}|\\ &\leq \frac{1}{\lambda^q}\int |Sf(y)|^q dy \lesssim_T \frac{1}{\lambda^q}\int |f(y)|^q dy < +\infty.\end{aligned}$$

Since S is of weak type $(1,1)$ we conclude by Lemma 6.18 we conclude that

$$|\{x : [M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}| \leq \frac{1}{\lambda^\nu}\|f\|^\nu_{L^1(\mathbb{R}^n)}|\{x : [M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}|^{1-\nu}.$$

which shows the desired estimate for $[M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}}$:

$$|\{x : [M_{\mathcal{D}}(|Sf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}| \leq \frac{1}{\lambda}\|f\|_{L^1(\mathbb{R}^n)}. \tag{6.8}$$

In order to complete the proof remember that from Lemma 4.47 we have

$$|\{x \in \mathbb{R}^n : M_{\square}g(x) > 4^n\lambda\}| \leq 2^n|\{x \in \mathbb{R}^n : M_{\mathcal{D}}(g)(x) > \lambda\}| \tag{6.9}$$

for all functions g, and that $M_{\square}g(x) \simeq_n Mg(x)$. These observations together with (6.8) conclude the proof of the lemma. □

EXERCISE 6.22. Show that for all $f \in L^1(\mathbb{R}^n)$ we have that

$$|\{x \in \mathbb{R}^n : M_{\mathcal{D}}(f)(x) > \lambda\}| \lesssim_n \int_{\{x\in\mathbb{R}^n : M_{\mathcal{D}}(f)(x)>\lambda\}} |f(x)|dx.$$

We can now give the proof of the fact that maximal truncation of a CZO is of weak type $(1,1)$ and strong type (p,p) for $1 < p < \infty$.

PROOF OF THEOREM 6.17. By Lemma 6.20 for $\nu = 1$ we immediately get that T_* is of strong type (p, p) for $1 < p < \infty$ since both M and T are. In order to show that T_* is of weak type $(1,1)$ we argue as follows. By Lemma 6.20 we have that

$$\begin{aligned}|\{x \in \mathbb{R}^n : T_*(f)(x) > \lambda\}| \lesssim_{n,\nu,\sigma} & |\{x \in \mathbb{R}^n : M(f)(x) > \lambda/2\}| \\ & + |\{x \in \mathbb{R}^n : [M(|Tf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda/2\}| \\ \lesssim & \frac{1}{\lambda}\|f\|_{L^1(\mathbb{R}^n)} + |\{x \in \mathbb{R}^n : [M(|Tf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda/2\}|.\end{aligned}$$

Thus the proof will be complete if we show that

$$|\{x \in \mathbb{R}^n : [M(|Tf|^\nu)(x)]^{\frac{1}{\nu}} > \lambda\}| \lesssim \frac{1}{\lambda}\|f\|_1.$$

However, this follows immediately from Lemma 6.21. □

6.3. Vector valued Calderón-Zygmund singular integral operators

We close this chapter on CZO's by describing a vector valued setup in which all our results on CZO's go through almost verbatim. We will see an application of these vector valued results in our study of *Littlewood-Paley* inequalities.

So let $\mathcal{H}$ be a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$ and consider a function $f : \mathbb{R}^n \to \mathcal{H}$. All the well known facts about spaces of measurable scalar functions have almost obvious generalizations in this setup once we fix some analogies. For example, the function f will be called measurable if for every $h \in \mathcal{H}$ the function $\mathbb{R}^n \to x \mapsto \langle f(x), h\rangle$ is a measurable function of x. If f is measurable then $\|f\|$ is also measurable. We then denote $L^p(\mathbb{R}^n;\mathcal{H})$ the space of all measurable functions $f : \mathbb{R}^n \to \mathcal{H}$ such that

$$\|f\|_{L^p(\mathbb{R}^n;\mathcal{H})} := \Big(\int \|f(x)\|^p dx\Big)^{\frac{1}{p}} < +\infty, \quad 1 \leq p < +\infty,$$

and the usual corresponding definition for $p = \infty$

$$\|f\|_{L^\infty(\mathbb{R}^n;\mathcal{H})} := \operatorname*{ess\,sup}_{x\in\mathbb{R}^n} \|f(x)\|.$$

It is not hard to check the duality relations for these L^p spaces; for example

$$\|f\|_{L^p(\mathbb{R}^n;\mathcal{H})} = \sup\Big\{\Big|\int \langle f(x), g(x)\rangle dx\Big| : \|g\|_{L^{p'}(\mathbb{R}^n;\mathcal{H})} \leq 1\Big\},$$

for all $1 \leq p < \infty$. Also our interpolations theorems, the Marcinkiewicz interpolation theorem and the Riesz-Thorin interpolation theorem go through in this setup as well.

Moreover, if a function $f : \mathbb{R}^n \to \mathcal{H}$ is absolutely integrable, we can define its integral as an element of $\mathcal{H}$ by defining the functional $I_f : \mathcal{H} \to \mathbb{C}$

$$I_f(h) := \int_{\mathbb{R}^n} \langle f(x), h\rangle dx.$$

Note here that I_f is uniquely defined as a functional on $\mathcal{H}^*$. Indeed, I_f is obviously linear and by the Cauchy-Schwartz inequality we have

$$|I_f(h)| = \Big|\int_{\mathbb{R}^n} \langle f(x), h\rangle dx\Big| \leq \int_{\mathbb{R}^n} |\langle f(x), h\rangle| dx \leq \Big(\int_{\mathbb{R}^n} \|f(x)\| dx\Big)\|h\|.$$

By the Riesz representation theorem on Hilbert spaces, there is a unique element of $\mathcal{H}$, which we denote by $\int_{\mathbb{R}^n} f(x)dx$, such that $I_f = \langle \int_{\mathbb{R}^n} f(x)dx, \cdot \rangle$, that is

$$I_f(h) = \Big\langle \int_{\mathbb{R}^n} f(x)dx, h \Big\rangle, \quad h \in \mathcal{H}.$$

Finally, if $\mathcal{H}_1, \mathcal{H}_2$ are separable Hilbert spaces we denote by $B(\mathcal{H}_1, \mathcal{H}_2)$ to be the space of bounded linear operators $T : \mathcal{H}_1 \to \mathcal{H}_2$, equipped with the usual operator norm:

$$\|T\|_{\mathcal{H}_1 \to \mathcal{H}_2} := \sup_{x \in \mathcal{H}_1} \frac{\|Tx\|_{\mathcal{H}_2}}{\|x\|_{\mathcal{H}_1}}.$$

Again, a function $F : \mathbb{R}^n \to B(\mathcal{H}_1, \mathcal{H}_2)$ will be called measurable if for every $h \in \mathcal{H}_1$ the function

$$\mathbb{R}^n \ni x \mapsto F(x)h \in \mathcal{H}_2$$

is a measurable $\mathcal{H}_2$-valued function.

We are now ready to give the description of vector valued CZO's. We start with the definition of a singular kernel.

DEFINITION 6.23 (Vector valued singular Kernel). Let $\mathcal{H}_1, \mathcal{H}_2$ be two separable Hilbert spaces and $K : \mathbb{R}^n \times \mathbb{R}^n \to B(\mathcal{H}_1, \mathcal{H}_2)$ be a function defined away from the diagonal $\Delta := \{x = y\}$. Then K will be called a (vector-valued) *singular kernel* if it obeys the size estimate

$$\|K(x,y)\|_{\mathcal{H}_1 \to \mathcal{H}_2} \lesssim_n \frac{1}{|x-y|^n}, \quad (x,y) \in \mathbb{R}^n \times \mathbb{R}^n \setminus \Delta, \tag{6.10}$$

and the regularity estimates

$$\|K(x,y_1) - K(x,y)\|_{\mathcal{H}_1 \to \mathcal{H}_2} \lesssim_{n,\sigma} \frac{|y-y_1|^\sigma}{|x-y|^{n+\sigma}} \quad \text{if} \quad |y-y_1| < \frac{1}{2}|x-y|, \tag{6.11}$$

and

$$\|K(x_1,y) - K(x,y)\|_{\mathcal{H}_1 \to \mathcal{H}_2} \lesssim_{n,\sigma} \frac{|x-x_1|^\sigma}{|x-y|^{n+\sigma}} \quad \text{if} \quad |x-x_1| < \frac{1}{2}|x-y|, \tag{6.12}$$

for some Hölder exponent $0 < \sigma \leq 1$.

DEFINITION 6.24. Let $\mathcal{H}_1, \mathcal{H}_2$ be separable Hilbert spaces. An linear operator $T : L^2(\mathbb{R}^n; \mathcal{H}_1) \to L^2(\mathbb{R}^n; \mathcal{H}_2)$ is called a (vector valued) *Calderón-Zygmund* operator (vector valued CZO) from $\mathcal{H}_1$ to $\mathcal{H}_2$ if it is bounded from $L^2(\mathbb{R}^n; \mathcal{H}_1)$ to $L^2(\mathbb{R}^n; \mathcal{H}_2)$

$$\|Tf\|_{L^2(\mathbb{R}^n;\mathcal{H}_2)} \lesssim_{n,T} \|f\|_{L^2(\mathbb{R}^n;\mathcal{H}_1)},$$

for all $f \in L^2(\mathbb{R}^n; \mathcal{H}_1)$, and there exists a vector valued singular kernel $K : \mathbb{R}^n \times \mathbb{R}^n \to B(\mathcal{H}_1, \mathcal{H}_2)$ such that

$$Tf(x) = \int K(x,y)f(y)dy,$$

whenever $f \in L^2(\mathbb{R}^n; \mathcal{H}_1)$ has compact support and $x \notin \operatorname{supp}(f)$.

Adjusting the proof of the scalar case to this vector valued setup we get the corresponding statement of Theorem 6.14.

THEOREM 6.25. *Let $\mathcal{H}_1, \mathcal{H}_2$ be separable Hilbert spaces and T be a vector valued Calderón-Zygmund operator from $\mathcal{H}_1$ to $\mathcal{H}_2$.*
(i) The operator T is of weak type $(1,1)$

$$|\{x \in \mathbb{R}^n : \|Tf(x)\|_{\mathcal{H}_2} > \lambda\}| \lesssim_{n,\sigma} \frac{\|f\|_{L^1(\mathbb{R}^n;\mathcal{H}_1)}}{\lambda}, \quad \lambda > 0,$$

for all $f \in L^1(\mathbb{R}^n; \mathcal{H}_1)$.
(ii) For all $1 < p < \infty$, T is of strong type (p,p)

$$\|Tf\|_{L^p(\mathbb{R}^n;\mathcal{H}_2)} \lesssim_{n,\sigma} \|f\|_{L^p(\mathbb{R}^n;\mathcal{H}_1)},$$

for all $f \in L^p(\mathbb{R}^n; \mathcal{H}_1)$.

CHAPTER 7

The space of functions of bounded mean oscillation, BMO.

7.1. Singular integral operators on L^∞ and BMO.

The theory of Calderón-Zygmund operators developed so far is pretty satisfactory except for one point, the action of a CZO on L^∞. Exercise 5.10 shows for example that in general a CZO cannot be bounded on L^∞. Furthermore, it is at the moment unclear how to define the action of T on a general bounded function or even on a dense subset of L^∞. Observe here that if $f \in L^p \cap L^2$ then Tf is well defined since we apriori assume T to be well defined and bounded on L^2. Thus what we did so far is show the boundedness of T as an operator $L^p \cap L^2 \to L^p$ and show that this operator extends by density to a bounded linear operator on L^p.

Let us try to interpret the kernel formula of T for a bounded function f:

$$Tf(x) = \int K(x,y)f(y)dy. \tag{7.1}$$

As we have already mentioned several times, such a formula is not meaningful throughout $\mathbb{R}^n$. Indeed the previous formula is problematic when x is close to y since K is singular along the diagonal $x = y$. Furthermore, the kernel formula is also problematic when $|y| \to +\infty$. Indeed, the function f is just bounded so no decay will come from the term f, and $K(x,y)$ is bounded by $|x-y|^{-n}$ which is not integrable as $|y| \to +\infty$. In order to deal with the first problem, the local singularity close to the diagonal, we can try our usual solution: localize f "away from" the point x. What about the behavior of the kernel formula as $|y| \to +\infty$? In the L^p-case with $p < +\infty$ we never really run into this problem since Cauchy-Schwartz and the size condition on K showed that the kernel formula was meaningful close to infinity. However, looking at the difference of the values of Tf at two points x_1, x_2 with $x_1 \neq x_2$, we can formally write

$$Tf(x_1) - Tf(x_2) = \int [K(x_1,y) - K(x_2,y)]f(y)dy.$$

Using the regularity condition (6.3) we see that

$$|K(x_1,y) - K(x_2,y)| \lesssim_{n,\sigma} \frac{|x_1 - x_2|^\sigma}{|x-y|^{n+\sigma}}.$$

This is enough to assure that the previous integral converges absolutely as $|y| \to +\infty$ as long as f is bounded.

In order to implement the heuristics discussed above we first choose some cube Q centered at a point c_Q and set $Q^* := (1 + 2\sqrt{n})Q$. We write

$$f(x) = f\mathbf{1}_{Q^*}(x) + f\mathbf{1}_{\mathbb{R}^n \setminus Q^*}(x).$$

thus splitting f to a "local piece" and a "far-away" piece. Now if $f \in L^\infty(\mathbb{R}^n)$ and $x \in Q$ we define

$$Tf(x) := T(f\mathbf{1}_{Q^*})(x) + \int_{\mathbb{R}^n\setminus Q^*} [K(x,y) - K(c_Q,y)]f(y)dy. \tag{7.2}$$

Observe that our definition of $Tf(x)$ is only *local*, it defines $Tf(x)$ for x in some cube Q.

It is easy to see that the right hand side of the definition (7.2) makes sense. Indeed, $T(f\mathbf{1}_{Q^*})$ is well defined since $f\mathbf{1}_{Q^*}$ is in $L^2(\mathbb{R}^n)$. On the other hand, the integral in the second summand converges absolutely since we integrate away from $Q \ni x$ and f is bounded. More precisely we have

$$|x-y| \geq \frac{1}{2}(2\sqrt{n})\,\mathrm{side}(Q) = \mathrm{diag}(Q) \geq 2|x-c_Q|.$$

Thus the regularity assumption for K applies to show that

$$|K(x,y) - K(c_Q,y)| \leq \frac{|x-c_Q|^\sigma}{|x-y|^{n+\sigma}} \lesssim_{Q,n,\sigma} \frac{1}{|x-y|^{n+\sigma}}.$$

Since f is bounded this shows that the integral in the second summand of (7.2) converges absolutely. However there is a certain ambiguity in the definition (7.2). Indeed if Q_1, Q_2 are two different cubes and $x \in Q_1 \cap Q_2$ we have two possible definitions for $Tf(x)$. However the difference in the two definitions can be estimated to be a constant independent of x. Let us first see this in the special case that $x \in Q_1 \subset Q_2$. We have

$$\begin{aligned}(\text{dfn-}Q_1 : \ Tf(x)) &= (\text{dfn-}Q_2 : \ Tf(x)) \\ &\quad + \int_{\mathbb{R}^n\setminus Q_2^*} [K(c_{Q_2},y) - K(c_{Q_1},y)]f(y)dy - \int_{Q_2^*\setminus Q_1^*} K(c_{Q_1},y)f(y)dy.\end{aligned}$$

The last two terms in the previous display are finite constants independent of x. For the second term note that the function $g = \mathbf{1}_{Q_2^*\setminus Q_1^*} f$ is an L^2-function with compact support and $c_{Q_1} \notin \mathrm{supp}(g)$. For the first term we have

$$|c_{Q_1} - c_{Q_2}| \leq \frac{1}{2}\sqrt{n}\,\mathrm{side}(Q_2) \leq \frac{1}{2}|y - c_{Q_2}|$$

for $y \in \mathbb{R}^n \setminus Q_2^*$. Thus the regularity estimate for K assures that the integral

$$\int_{\mathbb{R}^n\setminus Q_2^*} [K(c_{Q_2},y) - K(c_{Q_1},y)]f(y)dy$$

converges absolutely. Thus, in the special case $x \in Q_1 \subset Q_2$ the difference in the two definitions of $Tf(x)$ is a finite constant, independent of x. Now let Q_1, Q_2 be two cubes and $x \in Q_1 \cap Q_2$. We can assume without loss of generality that Q_2 has the largest sidelength among the two. Since the two cubes intersect at x and $\mathrm{side}(Q_2) \geq \mathrm{side}(Q_1)$, the cube $3Q_2 \supset Q_1, Q_2$. Now the definitions of $Tf(x)$ with respect to Q_1 and Q_2, both differ by a constant independent of x with the definition of $Tf(x)$ with respect to $3Q_2$. Thus all the definitions differ by a constant independent of x.

Thus we only define Tf modulo constants. To deal with this ambiguity in the definition, we have to define the appropriate space.

DEFINITION 7.1. We say that two functions $f,g\in\mathbb{R}^n$ are *equivalent modulo a constant* if there exists a constant $c\in\mathbb{C}$ such that $f(x)-g(x)=c$ almost everywhere on $\mathbb{R}^n$. This is an equivalence relationship. By abuse of language and notation we will oftentimes identify an equivalence class with a representative of the class, much like we do with measurable functions.

DEFINITION 7.2 (Bounded Mean Oscillation). Let f be a locally integrable function f, defined modulo a constant. We set

$$f_Q := \frac{1}{|Q|}\int_Q f =: \fint_Q f,$$

to be the average of f on the Euclidean cube Q. The BMO norm of f is the quantity

$$\|f\|_{\mathrm{BMO}} := \sup_Q \frac{1}{|Q|}\int_Q |f-f_Q|,$$

where the supremum varies over all Euclidean cubes Q. The space $\mathrm{BMO}(\mathbb{R}^n)$ is the set of all locally integrable functions f, defined modulo a constant, such that $\|f\|_{\mathrm{BMO}}<+\infty$. Thus, an element of BMO is only defined up to a constant.

First of all observe that this is a good definition since replacing a function f by $f+c$ for any constant $c\in\mathbb{C}$ does not affect its BMO norm. Thus, all elements in the equivalence class of f have the same BMO norm. The previous quantity actually defines a norm, always keeping in mind that we identify functions that differ by a constant. For example any constant is equivalent to the function 0 in BMO and thus $\|f\|_{\mathrm{BMO}}=0$ if and only if $f=c$ almost everywhere for some $c\in\mathbb{C}$.

It is not hard to give the following alternative description of the BMO norm, which is maybe a bit more revealing:

PROPOSITION 7.3. *Let* $f\in\mathrm{BMO}(\mathbb{R}^n)$.

(i) *We have that*

$$\|f\|_{\mathrm{BMO}} \simeq_n \sup_Q \inf_{a\in\mathbb{C}} \frac{1}{|Q|}\int_Q |f-a|.$$

(ii) *For any locally integrable function f and a ball B set $f_B=\fint_B f$. We set*

$$\|f\|_{\mathrm{BMO}_\bigcirc} := \sup_B \frac{1}{|B|}\int_Q |f-f_B|,$$

where the supremum is taken over all balls $B\subset\mathbb{R}^n$. Then

$$\|f\|_{\mathrm{BMO}_\bigcirc} \simeq_n \sup_B \inf_{a\in\mathbb{C}} \frac{1}{|B|}\int_B |f-f_B|$$

as in (i). Moreover $\|f\|_{\mathrm{BMO}}\simeq_n \|f\|_{\mathrm{BMO}_\bigcirc}$.

PROOF. For (i) observe that for any cube Q we have

$$\inf_{a\in\mathbb{C}} \frac{1}{|Q|}\int_Q |f-a| \leq \frac{1}{|Q|}\int_Q |f-f_Q|.$$

On the other hand for any $a\in\mathbb{C}$ we have

$$\frac{1}{|Q|}\int_Q |f-f_Q| \leq \frac{1}{|Q|}\int_Q |f-a| + \frac{1}{|Q|}\int_Q |f_Q-a| \leq \frac{2}{|Q|}\int_Q |f-a|,$$

which gives the opposite inequality as well by taking the infimum over $a \in \mathbb{C}$. The proof of the first claim in (ii) is identical. For the second claim in (ii) let $a \in \mathbb{C}$ and Q be a cube. Consider the smallest ball $B \supset Q$ with the same center as Q. Then

$$\frac{1}{|B|}\int_B |f-a| \gtrsim_n \frac{1}{|Q|}\int_Q |f-a|.$$

Thus,

$$\sup_B \inf_{a\in\mathbb{C}} \frac{1}{|B|}\int_B |f-a| \gtrsim_n \inf_{a\in\mathbb{C}} \frac{1}{|Q|}\int_Q |f-a|,$$

for any cube Q. Taking also the supremum over cubes Q proves the one direction of the inequality. The proof of the opposite inequality is similar. □

Thus a function f in BMO has the property that for any cube Q there is a constant c_Q such that $\frac{1}{|Q|}\int |f - c_Q| \leq \|f\|_{\mathrm{BMO}}$. That is, the values of f oscillate around c_Q by at most $\|f\|_{\mathrm{BMO}}$ in average. Locally, and in the *mean*, the function f has bounded oscillation.

The space BMO contains L^∞ but also contains unbounded functions.

PROPOSITION 7.4. *We have the following statements*

(i) *For every $f \in L^\infty(\mathbb{R}^n)$ we have that*

$$\|f\|_{\mathrm{BMO}(\mathbb{R}^n)} \lesssim \|f\|_{L^\infty(\mathbb{R}^n)},$$

thus $L^\infty(\mathbb{R}^n) \subset \mathrm{BMO}(\mathbb{R}^n)$.

(ii) *The function* $f(x) = \log|x|$ *is in* $\mathrm{BMO}(\mathbb{R}^n)$. *Thus* $L^\infty(\mathbb{R}^n)$ *is a proper subset of* $\mathrm{BMO}(\mathbb{R}^n)$.

EXERCISE 7.5. Prove Proposition 7.4.

With this definition of BMO, one now has a lot more flexibility in defining Tf for bounded functions f.

REMARK 7.6. Let T be a CZO with kernel K and $f \in L^\infty(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. Observe that there are two possible definitions for $Tf(x)$. One coming from the a priori definition of T on $L^2(\mathbb{R}^n)$, let us still call it $Tf(x)$. Since $f \in L^\infty(\mathbb{R}^n)$ we can also define $Tf(x)$ by means of (7.2). Let us temporarily call this L^∞-definition $\tilde{T}f(x)$. We claim that $Tf(x)$ and $\tilde{T}f(x)$ coincide as BMO functions. To see this first consider f be a bounded function with compact support. In particular $f \in L^2(\mathbb{R}^n)$. Let Q be a large cube such that $x \in Q$ and $\mathrm{supp}(f) \subset Q$. We then see that

$$\tilde{T}f(x) = T(f\mathbf{1}_{Q^*}f)(x) + \int_{\mathbb{R}^n\setminus Q^*} [K(x,y) - K(c_Q,y)]f(y)dy = T(f\mathbf{1}_{Q^*}f)(x) = Tf(x).$$

Thus $\tilde{T}$ and T agree on the space of bounded functions with compact support which are dense in $L^2(\mathbb{R}^n)$. Since T is known to be bounded on $L^2(\mathbb{R}^n)$ the two operators have a unique bounded extension to $L^2(\mathbb{R}^n)$, and thus coincide on $L^2(\mathbb{R}^n)$. However, the formula used for $\tilde{T}$ only defines $\tilde{T}$ up to a constant so $Tf(x)$ and $\tilde{T}f(x)$ only agree modulo constants, that is, in the BMO sense. We will henceforth just write $Tf(x)$ to denote the action of the operator T on $L^\infty(\mathbb{R}^n)$

Our interest in the space BMO mainly lies in the fact that it serves as a substitute endpoint for the boundedness of CZO's, namely a CZO T is bounded from L^∞ to BMO, where T should be defined as in (7.2). Note here that even though (7.2) only defines T "up to constants", this is the only possible definition of a BMO function.

THEOREM 7.7. *Let T be a CZO. Then for every $f \in L^\infty(\mathbb{R}^n)$ we have that*

$$\|Tf\|_{\mathrm{BMO}(\mathbb{R}^n)} \lesssim_{n,\sigma} \|f\|_{L^\infty(\mathbb{R}^n)}.$$

PROOF. Let Q be a cube $\mathbb{R}^n$, centered at c_Q. We need to show that for some constant β_Q we have

$$\frac{1}{|Q|}\int_Q |Tf - \beta_Q| \lesssim_{n,\sigma} \|f\|_{L^\infty}.$$

Since T is of strong type $(2,2)$ we have

$$\|T(f\mathbf{1}_{Q^*})\|_{L^2(\mathbb{R}^n)} \lesssim_{n,\sigma} \|f\|_{L^2(Q^*)} \leq \|f\|_{L^\infty(\mathbb{R}^n)}|Q^*|^{\frac{1}{2}}.$$

Thus by Cauchy-Schwartz we have

$$\frac{1}{|Q|}\int_Q |T(f\mathbf{1}_{Q^*})| \leq \frac{1}{|Q|}\|T(f\mathbf{1}_{Q^*})\|_{L^2(\mathbb{R}^n)}|Q|^{\frac{1}{2}} \lesssim_{n,\sigma} \|f\|_{L^\infty}.$$

On the other hand for $x \in Q$ and $y \notin Q^*$ we have $|x - c_Q| \leq \frac{1}{2}\sqrt{n}\,\mathrm{side}(Q) \leq \frac{1}{2}|x-y|$ thus

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus Q^*}[K(x,y) - K(c_Q,y)]f(y)dy &\leq \|f\|_{L^\infty(\mathbb{R}^n)}\int_{\mathbb{R}^n\setminus Q^*}|K(x,y) - K(c_Q,y)|dy \\
&\leq \|f\|_{L^\infty(\mathbb{R}^n)}\int_{|y-c_Q|\geq\sqrt{n}\,\mathrm{side}(Q)} \frac{|x-c_Q|^\sigma}{|x-y|^{n+\sigma}}dy \\
&\lesssim_{n,\sigma} \|f\|_{L^\infty(\mathbb{R}^n)}\,\mathrm{side}(Q)\int_{|y-c_Q|\geq\sqrt{n}\,\mathrm{side}(Q)} \frac{1}{|x-y|^{n+\sigma}}dy \\
&\lesssim_{n,\sigma} \|f\|_{L^\infty(\mathbb{R}^n)}.
\end{aligned}$$

Remembering that (7.2) only defines $Tf(x)$ up to some arbitrary finite constant β_Q, we have

$$\begin{aligned}
\frac{1}{|Q|}\int_Q |Tf(x) - \beta_Q|dx &\leq \frac{1}{|Q|}\int_Q |T(f\mathbf{1}_{Q^*})(x)|dx \\
&\quad + \frac{1}{|Q|}\int_Q\left|\int_{\mathbb{R}^n\setminus Q^*}[K(x,y) - K(c_Q,y)]f(y)dy\right|dx \\
&\lesssim_{n,\sigma} \|f\|_{L^\infty(\mathbb{R}^n)}
\end{aligned}$$

By Proposition 7.3 this proves the theorem. □

EXERCISE 7.8. Let $f \in \mathrm{BMO}(\mathbb{R}^n)$ and consider a sequence of nested cubes $Q_0 \subsetneq Q_1 \subsetneq \cdots \subsetneq Q_N$ where each cube Q_{j+1} in this sequence satisfies $|Q_{j+1}| = 2^n|Q_j|$, $\quad 0 \leq j \leq N-1$. Show that

$$|f_{Q_1} - f_{Q_N}| \leq N2^n\|f\|_{\mathrm{BMO}(\mathbb{R}^n)}.$$

7.2. The John-Nirenberg Inequality

We will now see that although the space BMO contains unbounded functions like $\log|x|$, this is in a sense the maximum possible growth for a BMO function. Although such a claim is not precise in a pointwise sense, it can be rigorously proved in the sense of level sets. Indeed, assuming $\|f\|_{\mathrm{BMO}} = 1$ then

$$\frac{1}{|Q|}\int_Q |f - f_Q| \leq 1,$$

for all cubes Q. Using Chebyshev's inequality this implies

$$|\{x \in Q : |f(x) - f_Q| > \lambda\}| \leq \frac{|Q|}{\lambda}.$$

This estimate is interesting for λ large, and states that on any cube Q the function f exceeds its average by λ only on a small fraction $1/\lambda$ of the measure of the cube $|Q|$. This can be substantially improved.

THEOREM 7.9 (John-Nirenberg inequality). *Let $f \in \mathrm{BMO}(\mathbb{R}^n)$. Then for any Euclidean cube Q we have that*

$$|\{x \in Q : |f(x) - f_Q| > \lambda\}| \lesssim_n e^{-c_n\lambda/\|f\|_{\mathrm{BMO}}}|Q|,$$

for all $\lambda > 0$, where the constant $c_n > 0$ depends only on the dimension n.

REMARK 7.10. Obviously it doesn't make any difference to work with balls instead of cubes so the the previous theorem remains valid with balls B replacing cubes Q.

PROOF. For $\lambda > 0$ let us denote by $\psi(\lambda)$ the best constant in the inequality

$$|\{x \in Q : |f(x) - f_Q| > \lambda\}| \leq \psi(\lambda)|Q|,$$

valid for any cube Q and f with $\|f\|_{\mathrm{BMO}} = 1$. By Chebyshev's inequality combined with the trivial bound we get

$$\psi(\lambda) \leq \min(1, 1/\lambda),$$

which is of course quite far from the desired estimate

$$\psi(\lambda) \lesssim_n e^{-c_n\lambda}.$$

We will however bootstrap this first trivial estimate to obtain the desired bound by iterating a local Calderón-Zygmund decomposition as follows.

Let us fix a cube Q_o and let $\mathcal{D}_m$ denote the family of dyadic cubes inside Q_o of sidelength $2^{-m}\,\mathrm{side}(Q)$. The family of all dyadic cubes inside Q_o will be denoted by $\mathcal{D}_o$. For a level $\Lambda > 1$ to be chosen later let $\mathcal{B}'$ be the "bad" cubes in $\mathcal{D}_o$, that is the cubes $Q \in \mathcal{D}_o$ such that

$$\frac{1}{|Q|}\int_Q F(w)dw > \Lambda,$$

where $F(w) = |f(w) - f_{Q_o}|$. Let $\mathcal{B}$ be the family of maximal bad cubes. Since $\frac{1}{|Q_o|}\int_{Q_o} F(w)dw \leq 1 < \Lambda$ for the original cube Q_o, every bad cube is contained in a maximal bad cube. As in the global Calderón-Zygmund decomposition we conclude that

$$\Lambda \leq \frac{1}{|Q|}\int_Q F(w)dw \leq r_n\Lambda$$

for each cube $Q \in \mathcal{B}$ where the constant r_n depends only on the dimension n. We also conclude that

$$F(w) \leq \Lambda$$

if $w \notin \cup_{Q\in\mathcal{B}} Q$ by the dyadic maximal theorem. Remembering the initial normalization $\|f\|_{\mathrm{BMO}} = 1$ we get

$$\sum_{Q\in\mathcal{B}} |Q| \leq \frac{1}{\Lambda} \sum_{Q\in\mathcal{B}} \int_Q F(w)dw \leq \frac{1}{\Lambda}|Q_o|$$

and for $Q \in \mathcal{B}$

$$|f_Q - f_{Q_o}| = \Big|\frac{1}{|Q|}\int_Q (f - f_{Q_o})\Big| \leq \frac{1}{|Q|}\int_Q F(w)dw \leq r_n\Lambda.$$

Now consider $\lambda > r_n\Lambda$. We have

$$\begin{aligned}
|\{x \in Q_o : |(f - f_{Q_o})(x)| > \lambda\}| &\leq |\{x \in \cup_{Q\in\mathcal{B}} Q : |f(x) - f_{Q_o}| > \lambda\}| \\
&\leq |\{x \in \cup_{Q\in\mathcal{B}} Q : |f(x) - f_Q| > \lambda - |f_Q - f_{Q_o}|\}| \\
&\leq \sum_{Q\in\mathcal{B}} |\{x \in Q : F(x) > \lambda - r_n\Lambda\}| \leq \psi(\lambda - r_n\Lambda) \sum_{Q\in\mathcal{B}} |Q| \\
&\leq \psi(\lambda - r_n\Lambda)\frac{1}{\Lambda}|Q_o|.
\end{aligned}$$

However this means that

$$\psi(\lambda) \leq \frac{\psi(\lambda - r_n\Lambda)}{\Lambda}$$

whenever $\lambda > r_n\Lambda$. Suppose that $Nr_n\Lambda < \lambda \leq (N+1)r_n\Lambda$ for some $N \geq 1$. Since $\psi(\lambda)$ is non-increasing and we have the trivial estimate $\psi(\lambda) \leq 1$ we get

$$\psi(\lambda) \leq \psi(Nr_n\Lambda) \leq \frac{\psi(r_n\Lambda)}{\Lambda^{N-1}} \leq \frac{1}{r_n\Lambda^N} \lesssim_n e^{-N\ln\Lambda} \leq e^{-(\frac{\lambda}{r_n\Lambda}-1)\ln\Lambda}.$$

Choosing $\Lambda = e$ (say) we get $\psi(\lambda) \lesssim_n e^{-c_n\lambda}$ for $\lambda > r_n e$. On the other hand, for $\lambda < r_n e$ we have

$$\psi(\lambda) \leq 1 \lesssim_n e^{-c_n\lambda}$$

so the proof is complete. □

COROLLARY 7.11. *For $1 \leq p < +\infty$ we consider the L^p version of the BMO norm*

$$\begin{aligned}
\|f\|_{\mathrm{BMO,p}} &:= \sup_B \Big(\frac{1}{|B|}\int_B |f - f_B|^p\Big)^{\frac{1}{p}} \simeq_{p,n} \sup_B \inf_{a\in\mathbb{C}} \Big(\frac{1}{|B|}\int_B |f - a|^p\Big)^{\frac{1}{p}} \\
&\simeq_{n,p} \sup_Q \Big(\frac{1}{|Q|}\int_Q |f - f_Q|^p\Big)^{\frac{1}{p}} \simeq_{n,p} \sup_Q \inf_{a\in\mathbb{C}} \Big(\frac{1}{|Q|}\int_Q |f - a|^p\Big)^{\frac{1}{p}}.
\end{aligned}$$

Then

$$\|f\|_{\mathrm{BMO}} \simeq_{p,n} \|f\|_{\mathrm{BMO,p}}.$$

COROLLARY 7.12. *Let $f \in \mathrm{BMO}(\mathbb{R}^n)$. There exists a constant c_n depending only on the dimension such that*

$$\sup_Q \frac{1}{|Q|}\int_Q \exp\Big(\frac{c_n}{\|f\|_{\mathrm{BMO}(\mathbb{R}^n)}}|f - f_Q|\Big) \lesssim_n 1,$$

where the implicit constant depends only on the dimension and the supremum is taken over all cubes $Q \subseteq \mathbb{R}^n$.

EXERCISE 7.13. Use the John-Nirenberg inequality and the description of L^p norms in terms of level sets to prove Corollary 7.11

EXERCISE 7.14. Use the John-Nirenberg inequality to prove the exponential integrability of Corollary 7.12

7.2.1. On exponential Orlicz classes. Here is a small note on some special Orlicz-type spaces. Let $(X, \sigma, \mathbb{P})$ be a probability space and consider $\psi : \mathbb{R} \to [0, \infty)$ a symmetric convex function such that $\psi(0) = 0$. Typical examples we should keep in mind here are

$$\psi(s) = s^p, \quad 1 \leq p < \infty, \quad \psi(s) \begin{cases} s, & 0 \leq s \leq 1 \\ e^{s^\alpha - 1} & for \quad s > 1 \end{cases}, \quad \alpha > 0$$

and the probability space $(Q, \mathcal{B}(Q), \frac{dx}{|Q|})$ where Q is a cube in $\mathbb{R}^n$ and (Q) denotes the Borel subsets of Q.

The Orlicz space L_ψ is defined the set of locally integrable functions f such that

$$\|f\|_{L_\psi} := \inf\{\lambda > 0 : \mathbb{E}\psi(|f|/\lambda) \leq 1\} < +\infty.$$

For the first example of ψ above we recover the local (normalized) L^p spaces on the cube Q while for the second example we recover the exponential Orlicz classes which we will denote by $\exp(L^\alpha)$.

We have already seen that the John-Nirenberg estimate for $\mathrm{BMO}(\mathbb{R}^n)$ amounts to saying that given $f \in \mathrm{BMO}(\mathbb{R}^n)$ and any cube Q we have

$$\|f - f_Q\|_{\exp(L^1)(Q)} \lesssim_n \|f\|_{\mathrm{BMO}(\mathbb{R}^n)}$$

uniformly over cubes Q, and thus f is exponentially integrable over every cube.

The following proposition is many times useful.

PROPOSITION 7.15. *Let Q be a cube and consider the normalized* $\exp L^\alpha(Q)$ *class with respect to the measure $dx/|Q|$ as above. Then*

$$\|f\|_{\exp(L^\alpha)(Q)} \eqsim_{n,\alpha} \sup_{p>1} p^{-\frac{1}{\alpha}} \|f\|_{L^p(dx/|Q|)]}$$

PROOF. Let us take some $\lambda > \|f\|_{\exp(L^\alpha)(Q)}$. Then we look for a constant $c = c(p, \alpha) \geq 1$ such that

$$|x|^p \leq c \exp(|x|^\alpha)$$

for $|x| \geq 1$. Setting $\phi(x) := x^p \exp(-x^\alpha)$ for $x > 0$ and differentiating we see that ϕ has a single maximum at the point $x_o = (p/\alpha)^{\frac{1}{\alpha}}$. Thus we get

$$|x|^p \leq \phi(x_o) e^{|x|^\alpha}$$

With this in mind we now have for $p \geq 1$

$$\fint_Q |f(y)|^p dx \leq \phi(x_o)\lambda^p \fint_Q \exp(|f(y)|^\alpha/\lambda^\alpha) dy \leq \phi(x_o)\lambda^p.$$

Thus $\|f\|_{L^p(dx/|Q|)} \leq \phi(x_o)^{\frac{1}{p}} \lambda \lesssim_\alpha p^{\frac{1}{\alpha}} \lambda$. Letting $\lambda \to \|f\|_{\exp(L^\alpha)(Q)}$ gives

$$\sup_{p>0} p^{-\frac{1}{\alpha}} \|f\|_{L^p(dx/|Q|)} \lesssim_\alpha \|f\|_{\exp(L^\alpha)(Q)}.$$

To see the other direction let us call $\rho := \sup_{p\geq 1} p^{-\frac{1}{\alpha}}\|f\|_{L^p(dx/|Q|)]}$ we write for $\lambda > 0$

$$\fint_Q \exp(|f|^\alpha/\lambda^\alpha) - 1 = \sum_{k=1}^{\infty} \frac{1}{\lambda^{k\alpha}k!}\fint_Q |f|^{k\alpha} \lesssim \sum_{k\geq 1} \frac{\rho^{k\alpha}(k\alpha)^k}{k!\lambda^{k\alpha}} \leq \sum_{k\geq 1}\frac{\rho^{k\alpha}(k\alpha e)^k}{k^{k+1/2}\lambda^{k\alpha}}$$
$$= \sum_{k\geq 1}\frac{(\rho^\alpha \alpha e)^k}{k^{\frac{1}{2}}\lambda^{k\alpha}}$$

by Stirling's formula $k! \approx k^{k+1/2}e^{-k}$. Now we see that the series in the last display converges to a number smaller than 1 as long as $\lambda > \lambda_o \approx \rho$. This however means that

$$\|f\|_{\exp(L^\alpha)(Q)} \leq \lambda_o \approx \rho$$

and the proof is complete. □

7.3. Interpolation and BMO.

One of the motivations for considering the space $\mathrm{BMO}(\mathbb{R}^n)$ is that it can serve as a replacement of L^∞ as an endpoint for different interpolation arguments. The first such simple example shows that the space $\mathrm{BMO}(\mathbb{R}^n)$ can serve as a different endpoint in the log-convexity estimates for the L^p norms.

LEMMA 7.16. *Let* $0 < p < q < \infty$ *and* $f \in L^p(\mathbb{R}^n)\cap \mathrm{BMO}(\mathbb{R}^n)$. *Then* $f \in L^q(\mathbb{R}^n)$ *and*

$$\|f\|_{L^q(\mathbb{R}^n)} \lesssim_{p,q,d} \|f\|_{L^p(\mathbb{R}^n)}^{\frac{p}{q}}\|f\|_{\mathrm{BMO}(\mathbb{R}^n)}^{1-\frac{p}{q}}.$$

PROOF. Obviously it is enough to assume that $\|f\|_{\mathrm{BMO}} \neq 0$ otherwise there is nothing to prove. Also by homogeneity we can normalize so that $\|f\|_{\mathrm{BMO}} = 1$. Now form the Calderón-Zygmund decomposition of $|f|^p$ at level 1 and denote by $\mathcal{B}$ the family of bad cubes as usual. For each cube $Q\in\mathcal{B}$ we then have

$$\frac{1}{|Q|}\Big|\int_Q f\Big| \leq \Big(\frac{1}{|Q|}\int_Q |f|^p\Big)^{\frac{1}{p}} \lesssim_{n,p} 1.$$

Using the John-Nirenberg inequality and the previous estimate we conclude that

$$|\{x\in Q : |f(x)| > \lambda\}| \leq \Big|\Big\{x \in Q : |f(x) - f_Q| > \lambda - |f_Q|\Big\}\Big| \lesssim_n e^{c_n|f_Q|}e^{-c_n\lambda}|Q| \lesssim_{n,p} e^{-c_n\lambda},$$

for all the bad cubes $Q\in\mathcal{B}$, where $c_n > 0$ is a dimensional constant. Since we have that $|f(x)| \leq 1$ for $x \notin \cup_{Q\in\mathcal{B}}Q$ we get

$$|\{x\in\mathbb{R}^n : |f(x)| > \lambda\}| \lesssim_n e^{-c_n\lambda}\sum_{Q\in\mathcal{B}}|Q| \leq e^{-c_n\lambda}\|f\|_{L^p}^p, \tag{7.3}$$

for all $\lambda > 1$. On the other hand, since $f\in L^p$ we have

$$|\{x\in\mathbb{R}^n : |f(x)| > \lambda\}| \leq \frac{\|f\|_{L^p}^p}{\lambda^p}. \tag{7.4}$$

We conclude the proof by using the description of the L^p norm in terms of level sets and using (7.4) for $\lambda < 1$ and (7.3) for $\lambda > 1$. □

EXERCISE 7.17 (The sharp Maximal function). For $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ define *the sharp maximal function*

$$M^\sharp(f)(x) := \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |f(y) - f_Q| dy, \quad x \in \mathbb{R}^n.$$

Observe that $f \in \mathrm{BMO}(\mathbb{R}^n)$ if and only if $M^\sharp(f) \in L^\infty(\mathbb{R}^n)$ and, in particular,

$$\|f\|_{\mathrm{BMO}(\mathbb{R}^n)} = \|M^\sharp\|_{L^\infty(\mathbb{R}^n)}.$$

Show that for every $x \in \mathbb{R}^n$ we have

$$M^\sharp(f)(x) \lesssim_n M'_\square(f)(x).$$

Thus the maximal function $M'_\square$ controls $M^\sharp(f)$ in the pointwise sense. Of course the opposite can not be true as it would imply that $M'_\square$ is bounded whenever $f \in \mathrm{BMO}(\mathbb{R}^n)$. Since $M'_\square(f) \geq f$ almost everywhere, this is impossible as we know that $\mathrm{BMO}(\mathbb{R}^n)$ contains unbounded functions. However we can reverse the inequality in the L^p sense.

THEOREM 7.18. *Let $1 \leq p_o \leq p < \infty$ and assume that $f \in L^{p_o}(\mathbb{R}^n)$. Then*

$$\|M'_\square(f)\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} \|M^\sharp(f)\|_{L^p}(\mathbb{R}^n).$$

By Lemma 4.47 the L^p norm of $M'_\square$ is controlled by the L^p norm of the dyadic maximal function $M_{\mathcal{D}}$ so it will be enough to prove Theorem 7.18 for the latter operator. The proof relies on the technique of a *good-λ inequality*.

LEMMA 7.19. *Let $\lambda, \gamma > 0$. Then*

$$|\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > 2\lambda,\, M^\sharp f \leq \gamma\lambda\}| \leq 2^n \gamma |\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\}|.$$

PROOF. Let us consider the Calderón-Zygmund decomposition of f at level λ and write $\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\} = \cup_j Q_j$. Observe that the stopping cubes Q_j exist as we assume that $f \in L^{p_o}(\mathbb{R}^n)$ for some $p_o < \infty$. Furthermore, we have that

$$E_{\gamma,\lambda} := \{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > 2\lambda,\, M^\sharp f(x) \leq \gamma\lambda\} \subseteq \{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\} = \cup_j Q_j$$

and thus we can estimate $|E_{\gamma,\lambda}| \leq \sum_j |E_{\gamma,\lambda} \cap Q_j|$. Now for $x \in E_{\gamma,\lambda} \cap Q_j$ we have that $M_{\mathcal{D}} f(x) \geq 2\lambda$. Thus there exists a dyadic cube $Q \ni x$ with $\fint_Q |f| \geq 2\lambda$. However, both $Q_j, Q \ni x$ thus one is contained in the other. As both cubes give f-averages greater than λ and Q_j is maximal we conclude that $Q \subset Q_j$. This observation implies that for $x \in E_{\gamma,\lambda} \cap Q_j$

$$M_{\mathcal{D}}((f - f_{Q_j^{(1)}})\mathbf{1}_{Q_j})(x) \geq |M_{\mathcal{D}}(f\mathbf{1}_{Q_j})(x) - f_{Q_j^{(1)}}| \geq \fint_Q |f| - |f|_{Q_j^{(1)}} > \lambda.$$

Using the weak $(1,1)$ inequality for $M_{\mathcal{D}}$ we can now estimate

$$\begin{aligned} |E_{\gamma,\lambda} \cap Q_j| &\leq |\{M_{\mathcal{D}}((f - f_{Q_j^{(1)}})\mathbf{1}_{Q_j})(x) > \lambda,\, M^\sharp f(x) \leq \gamma\lambda\} \cap Q_j| \\ &\leq \frac{1}{\lambda} \int_{Q_j} |f - f_{Q_j^{(1)}}| \leq \frac{1}{\lambda} |Q_j^{(1)}| \inf_{y \in Q_j^{(1)}} M^\sharp(f)(y). \end{aligned}$$

Observe that if there exists some $y \in Q_j$ such that $M^\sharp(f)(y) \leq \gamma\lambda$ then we have proved the estimate $|E_{\gamma,\lambda} \cap Q_j| \leq \gamma |Q_j^{(1)}|$. However, if such y does not exist, then $E_{\gamma,\lambda} \cap Q_j = \emptyset$ and there is nothing to prove. We have thus showed that

$$|E_{\gamma,\lambda}| \leq \gamma \sum_j |Q_j^{(1)}| \leq 2^n \gamma |\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\}|$$

as desired. □

The good-λ inequality just proved easily implies Theorem 7.18.

EXERCISE 7.20. Use the good-λ inequality of Lemma 7.19 in order to prove Theorem 7.18

Hint: Use the description of the L^p norm of $M_{\mathcal{D}}$ on the left hand in terms of level sets of the dyadic maximal function. In turn, these can be split into two parts, one where $M^\sharp(f)$ is big and another where $M_{\mathcal{D}} f$ is big and $M^\sharp(f)$ is small. The first term gives the right hand side while the second can be absorbed in the left hand side by a use of the good-λ inequality. Some care has to be taken when using this argument as one needs some apriori assumption in order to make sure that the L^p norm of $M_{\mathcal{D}} f$ is finite.

Another instance, and maybe a more important one, where $\mathrm{BMO}(\mathbb{R}^n)$ can replace $L^\infty(\mathbb{R}^n)$ as an endpoint is when it comes to interpolation of operators. The precise statement is as follows.

THEOREM 7.21 (Interpolation between L^p and BMO). *Let T be a linear operator such that T is bounded from $L^{p_o}(\mathbb{R}^n)$ to itself and T is bounded from $L^\infty(\mathbb{R}^n)$ to $\mathrm{BMO}(\mathbb{R}^n)$. Then T is bounded from $L^p(\mathbb{R}^n)$ to $L^p(\mathbb{R}^n)$ for all $p_o < p < \infty$.*

PROOF. We consider the operator $T^\sharp f := (T(f))^\sharp$. By the assumption we have that

$$\|T^\sharp f\|_{L^\infty(\mathbb{R}^n)} = \|Tf\|_{\mathrm{BMO}(\mathbb{R}^n)} \lesssim_T \|f\|_{L^\infty(\mathbb{R}^n)}.$$

Again, using the assumption and the boundedness of $M'_\square$ we have

$$\|T^\sharp f\|_{L^{p_o}(\mathbb{R}^n)} \lesssim \|M'_\square(Tf)\|_{L^{p_o}(\mathbb{R}^n)} \lesssim_{T,n,p_o} \|f\|_{L^{p_o}(\mathbb{R}^n)}$$

Since $T^\sharp$ is sublinear and bounded on L^{p_o} and on L^∞, Marcinkiewicz interpolation implies that $T^\sharp$ is bounded on $L^p(\mathbb{R}^n)$ for all $p_o < p < \infty$. Now consider any $f \in L^{p_o}(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ for $p > p_o$. By the assumption we have that $T(f) \in L^{p_o}$ and thus Theorem 7.18 applies for $T(f)$ (in place of f) to show that

$$\|T(f)\|_{L^p(\mathbb{R}^n)} \leq \|M'_\square(T(f))\|_{L(\mathbb{R}^n)} \lesssim_{p,n} \|T^\sharp f\|_{L^p(\mathbb{R}^n)} \lesssim_{T,p,n} \|f\|_{L^p(\mathbb{R}^n)}$$

and we are done. □

REMARK 7.22. If T is a Calderón-Zygmund operator remember that showing that $T : \mathrm{BMO}(\mathbb{R}^n) \to L^\infty$ was substantially easier than showing the weak $(1,1)$ property of T. Another way to establish the L^p bounds for T is to interpolate between the $L^2(\mathbb{R}^n) - L^2(\mathbb{R}^n)$ and the $L^\infty(\mathbb{R}^n) - \mathrm{BMO}(\mathbb{R}^n)$ bound. However, substantial machinery had to be used in order to prove the interpolation result for $\mathrm{BMO}(\mathbb{R}^n)$.

7.4. Commutators of singular integrals

We assume throughout this section that T is a Calderón-Zygmund operator and b is a function. The *commutator* of T and b is defined as

$$[T,b](f) := T(bf) - bT(f)$$

with gauges the non-commutativity of the operator T and the operator of multiplication by the function b. A first easy observation is that $[T,b]$ is bounded on L^p for $1 < p < \infty$ whenever $b \in L^\infty$. However, the Theorem of Coifman, Rochberg, and Weiss shows that the boundedness of the commutator $[T,b]$ remains valid whenever $b \in \mathrm{BMO}(\mathbb{R}^n)$, and in fact, it characterizes $\mathrm{BMO}(\mathbb{R}^n)$ in some sense that will be made precise in the following. The main purpose of this section is to describe this characterization. We begin with the positive direction.

THEOREM 7.23. *Let T be a Calderón-Zygmund operator and $b \in \mathrm{BMO}(\mathbb{R}^n)$. Then*

$$\|[T,b]\|_{L^p(\mathbb{R}^n)} \lesssim_{T,n,p} \|b\|_{\mathrm{BMO}(\mathbb{R}^n)} \|f\|_{L^p(\mathbb{R}^n)}$$

for all $1 < p < \infty$.

PROOF. The proof relies on the following pointwise estimate, valid say for all $f \in \mathcal{S}(\mathbb{R}^n)$:

$$M^\sharp([T,b]f) \lesssim_{r,n} M_r(Tf) + M_r(f),$$

where for $1 < r < \infty$, we denote by M_r the maximal operator

$$M_r(f)(x) := \sup_{Q \ni x} \Big(\frac{1}{|Q|} \int_Q |f(y)|^r dy \Big)^{\frac{1}{r}}.$$

Note that for $p > r$ the maximal operator M_r is bounded on $L^p(\mathbb{R}^n)$. Thus the identity above for $r > 1$ sufficiently small, combined with the estimate of Theorem 7.18 will give the L^p bound for the commutator.

In order to prove the pointwise estimate let Q be some cube in $\mathbb{R}^n$ and $Q^* := c_n Q$ be a concentric cube enlarged by $c_n > 1$ which will be chosen sufficient large depending upon dimension only. We estimate the commutator as follows

$$\begin{aligned}
[T,b]f &= T(bf) - bT(f) = T((b - b_{Q^*})f) + (b_{Q^*} - b)T(f) \\
&= T((b - b_{Q^*})f\mathbf{1}_{Q^*}) + T((b - b_{Q^*})f\mathbf{1}_{Q^{*c}}) + (b_{Q^*} - b)T(f) \\
&=: I + II + III.
\end{aligned}$$

First we use the John-Nirenberg equivalence of $\mathrm{BMO}(\mathbb{R}^n)$ and $\mathrm{BMO,p}(\mathbb{R}^n)$ from Corollary 7.11 to estimate III:

$$\begin{aligned}
\fint_Q |III| &\le \fint_Q |b - b_{Q^*}|\,|T(f)| \le \Big(\fint_Q |f - f_{Q^*}|^{r'} \Big)^{\frac{1}{r'}} \Big(\fint_Q |T(f)|^r \Big)^{\frac{1}{r}} \\
&\lesssim_{n,r} \|b\|_{\mathrm{BMO}(\mathbb{R}^n)} \inf_{y \in Q} M_r(Tf)(y).
\end{aligned}$$

For I, $1<q<r$ and use again the John-Nirenberg estimate from Corollary 7.11, the boundedness of T on L^q, and Hölder's inequality to estimate

$$\begin{aligned}\fint_Q |I| &\leq \Big(\fint_Q |T((b-b_{Q^*})f\mathbf{1}_{Q^*})|^q\Big)^{\frac{1}{q}} \lesssim_{T,n,q} \Big(\fint_{Q^*} |(b-b_{Q^*})f|^q\Big)^{\frac{1}{q}}\\ &\leq_{q,r,n} \|b\|_{\mathrm{BMO}(\mathbb{R}^n)}\Big(\fint_{Q^*}|f|^r\Big)^{\frac{1}{r}} \leq \|b\|_{\mathrm{BMO}(\mathbb{R}^n)} \inf_{y\in Q^*} M_r f(y)\\ &\leq \|b\|_{\mathrm{BMO}(\mathbb{R}^n)} \inf_{y\in Q} M_r f(y).\end{aligned}$$

Finally let $c(Q) := T((b-b_{Q^*})f\mathbf{1}_{Q^{*\mathsf{c}}})(c_Q)$ where c_Q is the center of Q. We have

$$\begin{aligned}\fint_Q |II-c(Q)|dx &\leq \fint_Q\int_{Q^{*\mathsf{c}}} |[K(x,y)-K(c_Q,y)](b(y)-b_{Q^*})f(y)|dy\,dx\\ &\lesssim_T \fint_Q\int_{Q^{*\mathsf{c}}} \frac{|x-c_Q|^\sigma}{|x-y|^{n+\sigma}}|b(y)-b_{Q^*}||f(y)|dy\,dx\\ &\lesssim_n \mathrm{side}(Q)^\sigma \int_{Q^{*\mathsf{c}}} \frac{1}{|y-c_Q|^{n+\sigma}}|b(y)-b_{Q^*}||f(y)|dy\end{aligned}$$

provided that the enlargement $Q^* = c_n Q$ is sufficiently large, depending upon dimension. Now we write

$$Q^{*\mathsf{c}} = \cup_{\tau\geq 1} 2^\tau Q^* \setminus 2^{\tau-1}Q^* =: \cup_{\tau\geq 1} E_\tau =: \cup_{\tau\geq 1} Q_\tau \setminus Q_{\tau-1}$$

and note that for $y\in 2^\tau Q^*\setminus 2^{\tau-1}Q^*$ we have that $|y-c_Q| \eqsim_n 2^\tau \mathrm{side}(Q)$. With this remark in hand we proceed with the estimate

$$\begin{aligned}\fint_Q |II-c(Q)|dx &\lesssim_{n,T,\sigma} \mathrm{side}(Q)^\sigma \sum_{\tau\geq 1}\frac{1}{(2^\tau \mathrm{side}(Q))^{n+\sigma}}\int_{E_\tau}|b(y)-b_{Q^*}||f(y)|dy\\ &\lesssim_n \sum_{\tau\geq 1} 2^{-\tau\sigma}\fint_{Q_\tau}|b(y)-b_{Q^*}||f(y)|dy\\ &\lesssim_n \sum_{\tau\geq 1}\tau 2^{-\tau\sigma}\fint_{Q_\tau}|b-b_{Q_\tau}||f(y)|dy\end{aligned}$$

where the last inequality uses Exercise 7.8 to replace the average b_{Q^*} by the average b_{Q_τ}, paying with a factor τ in the series. Now a use of Hölder's inequality with exponents r,r' and the fact that the series sums to a constant depending only on σ allows us to conclude the bound

$$\fint_Q |II-c(Q)|dx \lesssim_{n,T,\sigma} \|b\|_{\mathrm{BMO}(\mathbb{R}^n)} \inf_{y\in Q} M_r f(y).$$

Observe that for any constant $c(Q)$ depending on Q and function g we have

$$\fint_Q |g-g_Q| \leq \fint_Q |g-c(Q)| + \fint_Q |g_Q-c(Q)| \leq 2\fint_Q |g-c(Q)|$$

and thus we have proved the claimed pointwise estimate, which in turn proves the theorem. $\square$

The Theorem of Coifman, Rochberg, and Weiss, tells us that the commutator of a Calderón-Zygmund operator T with the multiplication by b is bounded on (say) L^2 whenever $b\in \mathrm{BMO}(\mathbb{R}^n)$. Now assume that $[T,b]$ is bounded on L^2 for a sufficiently large family of CZO's. Can we say anything about b? The answer is again provided by Coifman, Rochberg, and Weiss. Their original theorem states that if $[R_j, b]$ is bounded on $L^2(\mathbb{R}^n)$ for all the Riesz transforms R_j then $b \in \mathrm{BMO}(\mathbb{R}^n)$. We illustrate the result in one dimension when there is only one Riesz transform, namely the Hilbert transform.

THEOREM 7.24. *Let H denote the Hilbert transform on $\mathbb{R}$. Then for all $1 < p < \infty$ we have*

$$\|b\|_{\mathrm{BMO}(\mathbb{R})} \eqsim_p \| [H, b] \|_{L^p(\mathbb{R}) \to L^p(\mathbb{R})}.$$

PROOF. We only need to prove the estimate $\|[H,b]\| \gtrsim \|b\|_{\mathrm{BMO}(\mathbb{R})}$ as the other direction is contained in Theorem 7.23. For an interval I in $\mathbb{R}$ we need to estimate the quantity

$$\fint_I |b - b_I|.$$

From now on we fix the interval I and call $\Gamma_I := \mathrm{sgn}(b - b_I)\mathbf{1}_I$. Then

$$|b(x) - b_I|\mathbf{1}_Q(x) = (b(x) - b_I)\Gamma_I(x) = \frac{1}{|I|}\int (b(x) - b(y))\mathbf{1}_I(y)\Gamma(x)dy.$$

Let $\epsilon > 0$. Calling c_I the center of I we can write the series of identities

$$\int_{|x-y|>\epsilon} (b(x) - b(y))\mathbf{1}_I(y)\Gamma_I(x)dy = \int_{|x-y|>\epsilon} \frac{b(x) - b(y)}{x - y}(x - c_I - (y - c_I))\mathbf{1}_I(y)\Gamma_I(x)dx$$

$$\leq \left|(x - c_I)\Gamma_I(x)\int_{|x-y|>\epsilon} \frac{b(x) - b(y)}{x - y}\mathbf{1}_I(y)dy\right| + \left|\Gamma_I(x)\int_{|x-y|>\epsilon} \frac{b(x) - b(y)}{x - y}\mathbf{1}_I(y)(y - c_I)dy\right|.$$

We now need a small technical observation. Given a function $g \in L^\infty(I)$ we have

$$\int_{|x-y|>\epsilon} \frac{b(x) - b(y)}{x - y}g(y)\mathbf{1}_I(y)dy = b(x)\int_{|x-y|>\epsilon} \frac{g(y)\mathbf{1}_I(y)}{x - y}dy - \int_{|x-y|>\epsilon} \frac{g(y)b(y)\mathbf{1}_I(y)}{x - y}dy.$$

Observe that the functions $g\mathbf{1}_I$ and $gb\mathbf{1}_I$ are in $L^2(\mathbb{R})$ and thus we can write

$$\lim_{\epsilon\to 0^+}\int_{|x-y|>\epsilon} \frac{b(x) - b(y)}{x - y}g(y)\mathbf{1}_I(y)dy = \pi\Big(b(x)H(g\mathbf{1}_I)(x) - H(gb\mathbf{1}_I)(x)\Big) = \pi[H, b](g\mathbf{1}_I)(x).$$

On the other hand, since $b \in L^1(I)$ dominated convergence implies that

$$\lim_{\epsilon\to 0^+}\int_{|x-y|>\epsilon} (b(x) - b(y))\mathbf{1}_I(y)\Gamma(x)dy = |b(x) - b_I|\mathbf{1}_Q(x).$$

With these observations, the estimate above implies that

$$|b(x) - b_I|\mathbf{1}_Q(x) \leq \frac{\pi}{2}\mathbf{1}_I(x)|[H, b](\mathbf{1}_I)(x)| + \frac{1}{|I|}\mathbf{1}_I(x)|[H, b](\mathbf{1}_I(\cdot - c_I))(x)|$$

and thus, using the boundedness of the commutator on L^p we have

$$\int_I |b - b_I| \lesssim \|[H, b]\|_{L^p\to L^p}\Big(|I|^{\frac{1}{p}}|I|^{\frac{1}{p'}} + \frac{1}{|I|}|I|^{\frac{1}{p}}\|(\cdot - c_I)\|_{L^{p'}(I)}\Big) \lesssim \|[H, b]\|_{L^p\to L^p}|I|$$

and thus

$$\|b\|_{\mathrm{BMO}(\mathbb{R}^n)} \lesssim \|[H, b]\|_{L^p\to L^p}$$

which concludes the proof of the theorem. $\square$

CHAPTER 8

Littlewood-Paley inequalities and multiplier operators

In this chapter we will study the *Littlewood-Paley decomposition* and the *Littlewood-Paley inequalities*. These consist of very basic tools in harmonic analysis which allow us to decompose a function, on the frequency side, to pieces that have almost disjoint frequency supports. These pieces, the Littlewood-Paley pieces of the function, are almost orthogonal to each other, each piece oscillating at a different frequency.

8.1. The Littlewood-Paley decomposition

We start our analysis with forming a smooth Littlewood-Paley decomposition as follows. Let ϕ be a smooth real radial function supported on the open ball $\{\xi \in \mathbb{R}^n : 0 < |\xi| < 2\}$ of the frequency plane, which is identically equal to 1 on $\{\xi \in \mathbb{R}^n : 0 \leq |\xi| \leq 1\}$ and satisfies $0 \leq \phi \leq 1$. We then form the function ψ as

$$\psi(\xi) := \phi(\xi) - \phi(2\xi), \quad \xi \in \mathbb{R}^n.$$

Observing that $\phi(2\xi) = \phi(\xi) = 1$ if $|\xi| < 1/2$ and also that $\phi(\xi) = \phi(2\xi) = 0$ if $|\xi| \geq 2$ we see that ψ is supported on the annulus $\{\xi \in \mathbb{R}^n : 1/2 \leq |\xi| < 2\}$. Now the sequence of functions $\{\mathrm{Dil}^{\infty}_{2^k}\psi\}_{k\in\mathbb{Z}} = \{\psi(\cdot/2^k)\}_{k\in\mathbb{Z}}$ forms a *partition of unity*:

$$\sum_{k\in\mathbb{Z}} \psi(\xi/2^k) = 1, \quad \xi \in \mathbb{R}^n \setminus \{0\}.$$

To see this first observe that each function $\psi(\xi/2^k)$ has support inside the annulus $\{2^{k-1} \leq |\xi| \leq 2^{k+1}\}$. Thus for each given $\xi \in \mathbb{R}^n \setminus \{0\}$ there are only finitely many non-zero terms in the previous sum. In particular if $2^\ell \leq |\xi_o| < 2^{\ell+1}$ then

$$\sum_{k\in\mathbb{Z}} \psi(\xi_o/2^k) = \psi(\xi_o/2^\ell) + \psi(\xi_o/2^{\ell+1}) = \phi(\xi_o/2^{\ell+1}) - \phi(\xi_o/2^{\ell-1}) = 1.$$

Note that we miss the origin in our decomposition of the frequency space as each piece $\psi(\xi/2^k)$ is supported away from 0. Some attention is needed concerning this point but usually it creates no real difficulty.

Thus we partition the unity in the form $1 = \sum \psi_k$ and each ψ_k is smooth and has frequency support on an annulus of the form $|\xi| \simeq 2^k$. Now for $k \in \mathbb{Z}$ let us define the multiplier operators

$$\widehat{\Delta_k(f)}(\xi) := \psi(\xi/2^k)\hat{f}(\xi),$$

and

$$\widehat{S_k(f)}(\xi) := \sum_{\ell\leq k} \widehat{\Delta_\ell(f)}(\xi) = \phi(\xi/2^k)\hat{f}(\xi),$$

initially defined for $f \in L^2(\mathbb{R}^n)$ or $f \in \mathcal{S}(\mathbb{R}^n)$. The frequency cut-off operator Δ_k is almost a projection to the corresponding frequency annulus $\{2^{k-1} \le |\xi| < 2^{k+1}\}$; it is not exactly a projection since the function $\psi(\xi/2^k)$ is smooth, introducing a small tail which is mostly harmless. Similarly, the operator S_k is almost a projection on the ball $|\xi| \lesssim 2^k$.

We have the following simple properties of the Littlewood-Paley decomposition:

PROPOSITION 8.1. *We have the following estimates.*

(i) *For every $f \in L^2(\mathbb{R}^n)$ we have $\Delta_k(f) = S_k(f) - S_{k-1}(f)$ that is $\Delta_k = S_k - S_{k-1}$ in $L^2(\mathbb{R}^n)$.*
(ii) *For every $f \in L^2(\mathbb{R}^n)$ we have that $\lim_{k\to-\infty} S_k f = 0$ and $\lim_{k\to+\infty} S_k f = f$ where the limits are taken in the $L^2(\mathbb{R}^n)$-sense.*
(iii) *For every $f \in L^2(\mathbb{R}^n)$ we have that*
$$\sum_{k\in\mathbb{Z}} \Delta_k f = f$$
in $L^2(\mathbb{R}^n)$.

REMARK 8.2. Property (iii) above holds in a more general sense and for a wider class of functions, for example L^p functions and more generally, locally integrable functions that have some decay at infinity. The decomposition fails however if f has no decay. Indeed, the function 1 satisfies $\Delta_k 1 = 0$ for all $k \in \mathbb{Z}$. Observe here that $\hat{1} = \delta_0$ and thus the function 1 has frequency support on $\{0\}$ which is the point missed in our partition of unity.

Thus, with a Littlewood-Paley decomposition we managed to write any L^2 function (and thus any Schwartz function) as a sum of pieces $\Delta_k f$, each piece being well localized in frequency inside the annulus $|\xi| \simeq 2^k$.

It is pretty obvious how the operators S_k, Δ_k act on the frequency variable so let us take a look on what the pieces $S_k f, \Delta_k f$ look in the *physical space*. From the general facts about the Fourier transform (see for example Exercise 3.7) we know already that $S_k f, \Delta_k f$ cannot have compact spatial support. Since
$$\widehat{S_k f}(\xi) = \phi(\xi/2^k)\hat{f}(\xi) = \mathrm{Dil}^{\infty}_{2^k}\phi \hat{f}(\xi) = \mathcal{F}(\mathrm{Dil}^1_{2^{-k}}\check{\phi} * f)(\xi),$$
and $\check{\phi} = \hat{\hat{\phi}} = \hat{\phi}$, we have
$$S_k(f)(x) = (\mathrm{Dil}^1_{2^{-k}}\hat{\phi} * f)(x) = \int_{\mathbb{R}^n} f(x-y)2^{kn}\hat{\phi}(2^k y)dy = \int_{\mathbb{R}^n} f(x-2^{-k}y)\hat{\phi}(y)dy.$$
Here we remember that $\int \hat{\phi} = \phi(0) = 1$. From the discussion that followed the definition of convolutions in § 2.1.1 we thus see that $S_k f(x)$ is an average of f around the point x at scale $\simeq 2^{-k}$. Remembering that $\widehat{S_k f}$ is supported on the ball $\{|\xi| \lesssim 2^k\}$ this is also consistent with the uncertainty principle which also implies that the function $S_k(f)$ is essentially constant at scales $\lesssim 2^{-k}$. Now since a piece $\Delta_k f$ has frequency support contained in $\{|\xi| \le 2^{k+1}\}$ and $\phi(\xi/2^{k+1})$ is identically 1 on the ball $|\xi| \le 2^{k+1}$ we get that
$$\Delta_k(f) = S_{k+1}\Delta_k(f) = \int_{\mathbb{R}^n} \Delta_k f(x - 2^{-(k+1)}y)\hat{\phi}(y)dy.$$

Thus $\Delta_k(f)$ is essentially constant on scales $\lesssim 2^{-(k+1)}$. On the other hand, since $\Delta_k f$ has frequency support on the annulus $\{2^{k-1} \leq |\xi| \leq 2^{k+1}\}$ we have that

$$S_{k-2}\Delta_k f = 0.$$

As before we can rewrite this as

$$\int_{\mathbb{R}^n} \Delta_k f(x - 2^{-k+2}y)\hat{\phi}(y)dy = 0.$$

The previous identity roughly says that the function $\Delta_k(f)(x)$ has zero mean on every ball around x of radius $\gtrsim 2^{-k+2}$.

REMARK 8.3. We have mentioned in passing that the operators Δ_k can be seen as smooth approximations of the exact projections operators

$$\widehat{P_k(f)}(\xi) := \mathbf{1}_{\{2^k \leq |\xi| \leq 2^{k+1}\}}(\xi)\hat{f}(\xi).$$

Similarly, S_k can be viewed as a smooth approximation of the frequency projection

$$\widehat{\Sigma_k(f)}(\xi) := \mathbf{1}_{\{|\xi| \leq 2^k\}}(\xi)\hat{f}(\xi).$$

There are however important differences between the rough and smooth versions of these projections. For example, since ϕ is a Schwartz function the function $\hat{\phi}$ is also Schwartz and Young's inequality shows that

$$\|S_k(f)\|_{L^p} = \|\mathrm{Dil}^1_{2^{-k}}\hat{\phi} * f\|_{L^p} \leq \|\hat{\phi}\|_{L^1}\|f\|_{L^p},$$

thus S_k is bounded on L^p. Now, consider the rough version Σ_k given as

$$\Sigma_k(f)(x) = (\mathrm{Dil}^1_{2^{-k}}\widehat{\mathbf{1}}_{B(0,1)} * f)(x).$$

Of course Σ_k is still bounded on L^2 because of Plancherel's theorem. However, the function $\widehat{\mathbf{1}}_{B(0,1)}$ is no longer in L^1 and Young's inequality cannot be used. In fact Σ_k is not bounded on L^p whenever $n \geq 2$ and $p \neq 2$. This is a deep result of Charles Fefferman.

8.2. Littlewood-Paley Projections and derivatives

Recall the basic relation describing the interaction of derivatives with the Fourier transform:

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

In particular

$$|\widehat{\nabla f}|^2 = \sum_{j=1}^n \left|\widehat{\frac{\partial f}{\partial x_j}}\right|^2 = 4\pi^2|\xi|^2|\hat{f}|^2$$

If f has support on some annulus $|\xi| \simeq 2^k$ we immediately get

$$\|\nabla f\|_{L^2(\mathbb{R}^n)} \simeq_n 2^k\|f\|_{L^2(\mathbb{R}^n)},$$

and thus for any function $f \in L^2(\mathbb{R}^n)$ that

$$\|\nabla(\Delta_k f)\|_{L^2(\mathbb{R}^n)} \simeq_n 2^k\|\Delta_k f\|_{L^2(\mathbb{R}^n)},$$

In fact the same approximate identity extends to all L^p spaces for $1 \leq p \leq \infty$.

PROPOSITION 8.4. *Let g have Fourier support inside $\{2^{k-1} \leq |\xi| \leq 2^{k+1}\}$. For all $1 \leq p \leq \infty$ we have that*

$$\|\nabla g\|_{L^p} \simeq_n 2^k \|g\|_{L^p}$$

In particular

$$\|\nabla(\Delta_k f)\|_p \simeq_n 2^k \|\Delta_k f\|_p.$$

PROOF. Let us first prove that $\|\nabla g\|_p \lesssim_{n,p} 2^k \|g\|_p$ which is the "easy" direction. Since $\operatorname{supp}(\hat{g}) \subset \{2^{k-1} \leq |\xi| \leq 2^{k+1}\}$ we have

$$\begin{aligned} g(x) = S_{k+1} g(x) = \mathrm{Dil}^1_{2^{-(k+1)}} \hat{\phi} * g(x) = 2^{(k+1)n} \int_{\mathbb{R}^n} \hat{\phi}(2^{k+1} y) g(x-y) dy \\ = \int_{\mathbb{R}^n} g(x - 2^{-(k+1)} y) \hat{\phi}(y) dy. \end{aligned}$$

Thus we can write

$$\begin{aligned} \nabla_x g(x) = \int_{\mathbb{R}^n} \nabla_x g(x - 2^{-(k+1)} y) \hat{\phi}(y) dy = -2^{k+1} \int_{\mathbb{R}^n} \nabla_y g(x - 2^{-(k+1)} y) \hat{\phi}(y) dy \\ = 2^{k+1} \int_{\mathbb{R}^n} g(x - 2^{-(k+1)} y) \nabla_y \hat{\phi}(y) dy \end{aligned}$$

where the last equality follows by integration by parts. Now by Young's inequality we can conclude that for any $1 \leq p \leq +\infty$ we have

$$\|\nabla g\|_{L^p} \leq 2^{k+1} \|g\|_{L^p} \|\nabla \hat{\phi}\|_{L^1} \lesssim 2^k \|g\|_{L^p}$$

since for each j the function $\partial_{x_j} \hat{\phi}$ is a Schwartz function so that $\|\nabla \hat{\phi}\|_{L^1} \lesssim_n 1$. The precise value of the implied constant depends on the exact choice of ϕ but this is of small importance.

To prove the opposite inequality, $\|\nabla g\|_{L^p} \gtrsim 2^k \|g\|_{L^p}$ we will essentially proceed to "invert" the operator ∇. Let ρ be a Littlewood-Paley cutoff function which is identically 1 on $\{\frac{1}{2} \leq |\xi| \leq 2\}$, is compactly supported inside $\{\frac{1}{4} \leq |\xi| \leq 4\}$ (say), and satisfies $0 \leq \rho \leq 1$. We can the write

$$\rho(\xi/2^k) \widehat{\partial_{x_j} g}(\xi) = \rho(\xi/2^k)(2\pi i \xi_j) \hat{g}(\xi) = (2\pi i \xi_j) \hat{g}(\xi)$$

and thus

$$\xi_j \rho(\xi/2^k) \widehat{\partial_{x_j} g}(\xi) = 2\pi i \xi_j^2 \hat{g}(\xi).$$

Summing in $j \in \{1, 2, \ldots, n\}$ we get

$$\hat{g}(\xi) = \sum_{j=1}^n \rho(\xi/2^k) \frac{\xi_j}{2\pi i |\xi|^2} \widehat{\partial_{x_j} g}(\xi),$$

Inverting the Fourier transform we conclude that

$$g(x) = 2^{-k} \sum_{j=1}^n K_{k,j} * \partial_{x_j} g(x)$$

where

$$K_{k,j}(x) := 2^k \int_{\mathbb{R}^n} \rho(\xi/2^k) \frac{\xi_j}{2\pi i |\xi|^2} e^{2\pi i x \cdot \xi} d\xi = 2^{kn} \int_{\mathbb{R}^n} \rho(\xi) \frac{\xi_j}{2\pi i |\xi|^2} e^{2\pi i 2^k x \cdot \xi} d\xi.$$

Remembering that $\rho(\xi) \equiv 0$ whenever $|\xi| < \frac{1}{4}$ we can thus estimate

$$|K_{k,j}(x)| \leq \frac{4}{2\pi} \cdot 2^{kn} \|\rho\|_{L^1} \lesssim 2^{kn}.$$

By repeated integration by parts we can also estimate

$$|K_{k,j}(x)| \lesssim_N 2^{kn}(1 + 2^k|x|)^{-N} \tag{8.1}$$

for any positive integer N. The previous estimate allows us to estimate

$$|\mathrm{Dil}^1_{2^k} K_{k,j}(x)| \lesssim_N (1 + |x|)^{-N}$$

for every N which means that $K_{k,j}$ can be written in the form

$$K_{k,j} = \mathrm{Dil}^1_{2^{-k}} \mathrm{Dil}^1_{2^k} K_{k,j}$$

which means that K_k, j is an approximation to the identity at scales 2^k. In particular $\|K_{k,j}\|_1 \lesssim 1$ and thus

$$\|g\|_{L^p} = 2^{-k} \sum_{j=1}^n \|K_{k,j}\|_{L^1} \|\partial_{x_j} g(x)\|_{L^p} \lesssim_n 2^{-k} \|\nabla g\|_{L^p}$$

which proves the desired estimate. □

EXERCISE 8.5. Prove estimate (8.1) above.

8.3. The Littlewood-Paley inequalities

The Littlewood-Paley inequalities quantify the heuristic principle that the pieces $\Delta_k(f)$, having well separated frequency supports, behave independently of each other, meaning that

$$\Big|\sum_k \Delta_k(f)\Big| \simeq \Big(\sum_k |\Delta_k(f)|^2\Big)^{\frac{1}{2}},$$

in some appropriate sense (for example in L^p). In L^2 this is already an easy consequence of Plancherel's theorem. Indeed, note that

$$\Big\|\Big(\sum_k |\Delta_k(f)|^2\Big)^{\frac{1}{2}}\Big\|^2 = \int_{\mathbb{R}^n} \sum_{k\in\mathbb{Z}} |\psi(\xi/2^k)|^2 |\hat{f}(\xi)|^2 d\xi.$$

Like before observe that for every $\xi \in \mathbb{R}^n$ there are only two terms $\psi(\xi/2^\ell), \psi(\xi/2^{\ell+1})$ which don't vanish, and these add up to 1. We have

$$1 = (\psi(\xi/2^\ell) + \psi(\xi/2^{\ell+1}))^2 = |\psi(\xi/2^\ell)|^2 + |\psi(\xi/2^{\ell+1})|^2 + 2\psi(\xi/2^\ell)\psi(\xi/2^{\ell+1}),$$

and thus

$$\sum_{k\in\mathbb{Z}} |\psi(\xi/2^k)|^2 = |\psi(\xi/2^\ell)|^2 + |\psi(\xi/2^{\ell+1})|^2 \simeq 1.$$

We can equivalently write this identity in the form

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |\Delta_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^2} \simeq \|f\|_{L^2}.$$

The following theorem is the central result of this chapter and provides an extension of this approximate identity to all L^p spaces for $1 < p < \infty$.

THEOREM 8.6. *Define the Littlewood-Paley square function as*

$$S(f)(x) := \Big(\sum_{k\in\mathbb{Z}} |\Delta_k(f)(x)|^2\Big)^{\frac{1}{2}}.$$

Then for all $1 < p < \infty$ *we have*

$$\|S(f)\|_{L^p(\mathbb{R}^n)} \simeq_{n,p} \|f\|_{L^p(\mathbb{R}^n)}.$$

PROOF. Consider the vector valued singular integral operator

$$\vec{S}(f)(x) := \{\Delta_k f(x)\}_{k\in Z},$$

and observe that

$$S(f)(x) = \|\vec{S}(f)(x)\|_{\ell^2(\mathbb{Z})}.$$

Thus the statement of the Theorem is equivalent to

$$\|\vec{S}(f)\|_{L^p(\mathbb{R}^n;\ell^2(\mathbb{Z}))} \simeq_{n,p} \|f\|_{L^p(\mathbb{R}^n)}. \tag{8.2}$$

Observe that $\vec{S}$ is a bounded linear operator from $L^2(\mathbb{R}^n;\mathbb{C})$ to $L^2(\mathbb{R}^n;\ell^2(\mathbb{Z}))$. Indeed the strong $(2,2)$ type of $\vec{S}$ follows from the remarks before the theorem. Furthermore, defining

$$\vec{K}(x,y) := \{2^{nk}\hat{\psi}(2^k(x-y))\}_{k\in\mathbb{Z}}, \quad (x,y)\in\mathbb{R}^n\times\mathbb{R}^n\setminus\Delta,$$

we can verify that $\vec{K}$ is a singular kernel associated to the vector valued singular integral $\vec{S}$:

LEMMA 8.7. *The kernel* K *defined above is a singular kernel from* $\mathbb{C}$ *to* $\ell^2(\mathbb{Z})$.

Postponing the proof of this lemma for now, we use the vector valued version of the Calderón-Zygmund theorem to show that $\vec{S}$ is bounded from $L^p(\mathbb{R}^n)$ to $L^p(\mathbb{R}^n;\ell^2(\mathbb{Z}))$:

$$\|\vec{S}(f)\|_{L^p(\mathbb{R}^n,\ell^2(\mathbb{Z}))} \lesssim_{n,p,\psi} \|f\|_{L^p(\mathbb{R}^n)},$$

which is one of the estimates in (8.2). To prove the lower estimate we argue as follows. Let $\vec{g} = \{g_j\}_{j\in\mathbb{Z}} : \mathbb{R}^n \to \ell^2(\mathbb{Z})$. Then

$$\begin{aligned}\int_{\mathbb{R}^n}\langle\vec{S}(f)(x),\vec{g}(x)\rangle dx &= \int_{\mathbb{R}^n}\sum_{k\in\mathbb{Z}}\Delta_k(f)(x)\overline{g_k(x)}dx = \int_{\mathbb{R}^n}\sum_{k\in\mathbb{Z}}\psi_k(\xi/2^k)\hat{f}(\xi)\overline{\widehat{g_k}(\xi)}d\xi\\ &= \int_{\mathbb{R}^n} f(x)\sum_{k\in\mathbb{Z}}\overline{\Delta_k(g_k)}(x)dx =: \int_{\mathbb{R}^n} f(x)\overline{\vec{S}^*(\vec{g})(x)}dx.\end{aligned}$$

By vector valued duality and the estimate $\|\vec{S}(f)\|_{L^p(\mathbb{R}^n,\ell^2(\mathbb{Z}))} \lesssim_{n,p,\psi} \|f\|_{L^p(\mathbb{R}^n)}$ we conclude that the adjoint operator $\vec{S}^*$ satisfies

$$\Big\|\sum_{k\in\mathbb{Z}}\Delta_k(g_k)\Big\|_{L^p(\mathbb{R}^n)} = \|\vec{S}^*(\vec{g})\|_{L^p(\mathbb{R}^n)} \lesssim_{n,p,\psi} \|\vec{g}\|_{L^p(\mathbb{R}^n;\ell^2(\mathbb{Z}))}, \quad 1<p<\infty.$$

Now we repeat the Littlewood-Paley decomposition but starting with the function

$$\tilde{\psi}(\xi) = \phi(\xi/4) - \phi(4\xi),$$

and setting

$$\widehat{\tilde{\Delta}_k(f)}(\xi) := \tilde{\psi}(\xi/2^k)\hat{f}(\xi) = (\phi(\xi/(42^k)) - \phi(4\xi/2^k))\hat{f}(\xi)$$

or equivalently

$$\tilde{\Delta}_k(f) = S_{k+2}(f) - S_{k-2}(f).$$

Observe that in this case we also have $\sum_{k\in\mathbb{Z}} \tilde{\psi}(\xi/2^k) \simeq 1$ for all $\xi \in \mathbb{R}^n \setminus \{0\}$. Using exactly the same arguments as before we can show that we have the same estimates for these modified Littlewood-Paley projections, namely

$$\Big\|\Big(\sum_{k\in\mathbb{Z}} |\tilde{\Delta}_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} \|f\|_{L^p(\mathbb{R}^n)} \tag{8.3}$$

and

$$\Big\|\sum_{k\in\mathbb{Z}} \tilde{\Delta}_k(g_k)\Big\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} \|\vec{g}\|_{L^p(\mathbb{R}^n,\ell^2(\mathbb{Z}))}. \tag{8.4}$$

Observe that for $2^{k-1} \leq |\xi| \leq 2^{k+1}$ we have that $|\xi|/(42^k) \leq 2^{k+1}/(42^k) = 1/2$ and $|4\xi|/2^k \geq 42^{k+1}/2^k = 8$ thus for any function h with $\operatorname{supp}(h) \subset \{2^{k-1} < |\xi| \leq 2^{k+1}\}$ we have that $\tilde{\Delta}_k h = h$. In particular observe that $\tilde{\Delta}_k\Delta_k(f) = \Delta_k(f)$ since we already have that $\operatorname{supp}(\Delta_k(f)) \subset \{2^{k-1} \leq |\xi| \leq 2^{k+1}\}$. Applying (8.4) with $\vec{g} := \{\Delta_k(f)\}_{k\in\mathbb{Z}}$ we get

$$\Big\|\sum_{k\in\mathbb{Z}} \tilde{\Delta}_k(\Delta_k(f))\Big\|_{L^p(\mathbb{R}^n)} = \Big\|\sum_{k\in\mathbb{Z}} \Delta_k(f)\Big\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} \Big\|\Big(\sum_{k\in\mathbb{Z}} |\Delta_k f|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R}^n)}.$$

However on the left hand side we have the pointwise identity $\sum_k \Delta_k(f)(x) = f(x)$ which shows that

$$\|f\|_{L^p(\mathbb{R}^n)} \lesssim_{p,n} \|\vec{S}(f)\|_{L^p(\mathbb{R}^n,\ell^2(\mathbb{Z}))},$$

as we wanted. □

We now go back to the proof of Lemma 8.7.

PROOF OF LEMMA 8.7. Remember that the kernel $\vec{K}$ is given as

$$\vec{K}(x,y) = \{2^{nk}\hat{\psi}(2^k(x-y))\}_{k\in\mathbb{Z}}.$$

Let $\psi_k(\xi) := \psi(\xi/2^k)$ so that

$$2^{nk}\hat{\psi}(2^k x) = \widehat{\psi_k}(x)$$

First of all we prove the estimates

$$|\widehat{\psi_k}(x)| \lesssim \frac{1}{|x|^n}\min((2^k|x|)^n, (2^k|x|)^{-2}), \tag{8.5}$$

and

$$|\nabla\widehat{\psi_k}(x)| \lesssim \frac{1}{|x|^{n+1}}\min((2^k|x|)^{n+1}, (2^k|x|)^{-1}), \tag{8.6}$$

For (8.5) we write

$$|\widehat{\psi_k}(x)| = 2^{nk}\Big|\int_{\mathbb{R}^n} \psi(\xi)e^{-2\pi i 2^k x\cdot\xi}d\xi\Big|.$$

On the one hand we have that

$$|\psi_k(x)| \leq 2^{nk}\int_{\mathbb{R}^n} |\psi(\xi)|d\xi \lesssim_\psi \frac{1}{|x|^n}(2^k|x|)^n.$$

Furthermore, for any positive integer N we have

$$|\psi_k(x)| = 2^{nk}\Big|\int_{\mathbb{R}^n} \psi(\xi)(\frac{x}{2\pi i 2^k|x|^2}\cdot\nabla_\xi)^N e^{2\pi i 2^k x\cdot\xi}dx\Big| \lesssim_{\psi,N} 2^{nk}\frac{1}{(2^k|x|)^N},$$

by integrating by parts N times and passing the derivatives to ψ. Applying this estimate for $N = n+2$ gives the second estimate in (8.5).

The proof of (8.6) is very similar by observing that

$$\nabla\widehat{\psi_k}(x) = 2^{nk}\int_{\mathbb{R}^n}\psi(\xi)\nabla_x e^{-2\pi i 2^k x\cdot\xi}d\xi = 2^{nk}\int_{\mathbb{R}^n}\psi(\xi)(-2\pi i 2^k\xi)e^{-2\pi i 2^k x\cdot\xi}d\xi.$$

Now the same analysis as in (8.5) applies (with an extra 2^k factor) and gives (8.6). Estimates (8.5) and (8.6) now imply the size and regularity conditions for the singular kernel $\vec{K}$ in $\ell^2(\mathbb{Z})$. □

8.3.1. A rough version for n-dimensional dyadic intervals. So far we carried out the Littlewood-Paley decomposition based on a smooth partition of unity. The use of smooth functions to form the Littlewood-Paley decomposition has many advantages since then the projections Δ_j are bounded multiplier operators. On the other hand Remark 8.3 shows that in dimensions $n > 1$ the multiplier associated with a Euclidean ball is not bounded on L^p. This means that the Littlewood-Paley inequalities based on the projections

$$\widehat{P_k(f)} = \mathbf{1}_{\{2^k<|\xi|\leq 2^{k+1}\}}\hat{f}, \tag{8.7}$$

will fail in any dimension $n \geq 2$.

The previous discussion leaves the one-dimensional case open. In fact we will see now that one can form the Littlewood-Paley decomposition in one dimension based on the rough partition of unity

$$1 = \sum_{k\in\mathbb{Z}}\mathbf{1}_{\{2^k<|\xi|\leq 2^{k+1}\}}(\xi), \quad \xi\in\mathbb{R}^n\setminus\{0\},$$

and still have the Littlewood-Paley inequalities. So let us define P_k to be the exact frequency projection as in (8.7). We have the following.

THEOREM 8.8. *Let $f \in L^p(\mathbb{R})$, $1 < p < \infty$. Then we have the one dimensional Littlewood-Paley inequalities for the rough projections P_k:*

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}|P_k(f)|^2\Big)^{\frac12}\Big\|_{L^p(\mathbb{R})} \simeq_p \|f\|_{L^p(\mathbb{R})}.$$

PROOF. Writing $P_k(f)$ in the form

$$\widehat{P_k f} = \mathbf{1}_{[-2^{k+1},2^k)}\hat{f} + \mathbf{1}_{(2^k,2^{k+1}]}\hat{f},$$

we have the following representation in terms of the Hilbert transform

$$P_k(f) = \frac{i}{2}\Big(\mathrm{Mod}_{2^k}H\mathrm{Mod}_{-2^k}f - \mathrm{Mod}_{2^{k+1}}H\mathrm{Mod}_{-2^{k+1}}f\Big). \tag{8.8}$$

For $\vec{g} = \{g_k\}_{k\in\mathbb{Z}} \in L^p(\mathbb{R};\ell^2(\mathbb{Z}))$ let us define the vector valued analogue

$$\vec{P}(\vec{g}) := \{P_k(g_k)\}_{k\in\mathbb{Z}}.$$

Using the fact that H is a CZO and the representation (8.8) of P_k in terms of H we can see that $\vec{P}$ is a vector valued Calderón-Zygmund operator, thus $\vec{P}$ is bounded from $L^p(\mathbb{R};\ell^2(\mathbb{Z}))$ to $L^p(\mathbb{R};\ell^2(\mathbb{Z}))$. Applying this property to the vector-valued function

$$\vec{g} := \{\tilde{\Delta}_k f\}_{k\in\mathbb{Z}}$$

we get

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}|P_k\tilde{\Delta}_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R})}\lesssim_p\Big\|\Big(\sum_{k\in\mathbb{Z}}|\tilde{\Delta}_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R})},$$

where the $\tilde{\Delta}_k$'s are as in the proof of Theorem 8.6. Now we remember that that $\widehat{\tilde{\Delta}_k(f)}(\xi)=\hat{f}(\xi)$ whenever $2^{k-1}\leq|\xi|\leq 2^{k+1}$. Thus we have the identity $P_k\tilde{\Delta}_k(f)=\tilde{\Delta}_kP_k(f)=P_k(f)$. The previous estimate implies that

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}|P_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R})}\lesssim_p\Big\|\Big(\sum_{k\in\mathbb{Z}}|\tilde{\Delta}_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R})}.$$

By (8.3) in the proof of Theorem 8.6 we get one of the inequalities in the statement of the theorem:

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}|P_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R})}\lesssim_p\|f\|_{L^p(\mathbb{R})}.$$

To prove the opposite inequality, we write the dual estimate that was obtained in proof of Theorem 8.6:

$$\Big\|\sum_{k\in\mathbb{Z}}\tilde{\Delta}_k(\vec{g})\Big\|_{L^p(\mathbb{R})}\lesssim_p\|\vec{g}\|_{L^p(\mathbb{R};\ell^2(\mathbb{Z}))}.$$

Now take $\vec{g}:=\{P_kf\}_{k\in\mathbb{Z}}$ and use the observation $\tilde{\Delta}_kP_kf=P_k(f)$ to write

$$\|f\|_{L^p(\mathbb{R})}=\Big\|\sum P_kf\Big\|_{L^p(\mathbb{R})}\lesssim_p\|\vec{g}\|_{L^p(\mathbb{R};\ell^2(\mathbb{Z}))}=\Big\|\Big(\sum_{k\in\mathbb{Z}}|P_k(f)|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R})},$$

which gives the other inequality in the theorem. □

EXERCISE 8.9. Let T be a scalar-valued Calderón-Zygmund Operator and let $1<r,p<+\infty$. If $\vec{f}\in L^p(\mathbb{R}^n;\ell^r(\mathbb{Z}))$ then show that

$$\Big\|\Big(\sum_{k\in\mathbb{Z}}|T(f_k)|^r\Big)^{\frac{1}{r}}\Big\|_{L^p(\mathbb{R}^n)}\lesssim_{p,n,r,T}\Big\|\Big(\sum_{k\in\mathbb{Z}}|f_j|^r\Big)^{\frac{1}{r}}\Big\|_{L^p(\mathbb{R}^n)}.$$

Hint: Consider the vector valued operator

$$\vec{T}(\vec{f}):=\{Tf_k\}_{k\in\mathbb{Z}}.$$

The problem reduces to showing that $\vec{T}$ is a vector valued CZO from $L^p(\mathbb{R}^n;\ell^r(\mathbb{Z}))$ to $L^p(\mathbb{R}^n;\ell^r(\mathbb{Z}))$. Observe that $\vec{T}$ is associated with the kernel

$$\vec{K}(x,y):=K(x,y)\mathrm{id}_{\ell^r},$$

where id_{ℓ^r} is the identity from $\ell^r(\mathbb{Z})$ to $\ell^r(\mathbb{Z})$ and K is the (scalar) kernel associated with T. You can assume the Banach space version of the vector valued Calderón-Zygmund theorem.

EXERCISE 8.10. Let $\{I_k\}_{k\in\Lambda}$ be a sequence of bounded or unbounded intervals on the real line, where Λ is a finite or countably infinite index set. Define the frequency projections

$$\widehat{P_{I_k}f}:=\mathbf{1}_{I_k}\hat{f}.$$

Show that

$$\Big\|\Big(\sum_{k\in\Lambda}|P_{I_k}f|^r\Big)^{\frac{1}{r}}\Big\|_{L^p(\mathbb{R})}\lesssim_{p,r}\Big\|\Big(\sum_{k\in\Lambda}|f|^r\Big)^{\frac{1}{r}}\Big\|_{L^p(\mathbb{R})}.$$

Hint: Like in the proof of Theorem 8.8 use the representation of the projections P_{I_k} in terms of the Hilbert transform and Exercise 8.9.

We have already remarked (see Remark 8.3) that Theorem 8.8 does not generalize to annuli in the n-dimensional Euclidean space if we insist on using the rough projections $\mathbf{1}_{\{2^k<|\xi|\leq 2^{k+1}\}}\hat{f}$. However, there is a generalization of the "rough" Littlewood-Paley theorem to dimensions $n>1$. This is based on a decomposition of the frequency space $\mathbb{R}^n$ into a union of disjoint dyadic "intervals", that is, n-dimensional rectangles with sides parallel to the coordinate axes where every side of the rectangle is an interval of the form $(2^k,2^{k+1}]$ or $[-2^{k+1},-2^k]$. This allows for "tensoring" Theorem 8.8 to several dimensions without great difficulty. This is done as follows. For $k=(k_1,\ldots,k_n)\in\mathbb{Z}^n$ we set

$$P^{(k)}:=P_{k_1}P_{k_2}\cdots P_{k_n}$$

where each P_{k_j} is the one-dimensional projection previously defined acting only on the j-th variable. For $f:\mathbb{R}^n\to\mathbb{C}$ we have

$$\widehat{P_{k_j}f}(\xi)=\mathbf{1}_{\{2^{k_j}<|\xi_j|\leq 2^{k_j+1}\}}(\xi_j)\hat{f}(\xi),\quad \xi=(\xi_1,\ldots\xi_j,\ldots,x_n)\in\mathbb{R}^n.$$

The corresponding square function is defined as

$$S_\square(f)(\xi):=\Big(\sum_{k=(k_1,\ldots,k_n)\in\mathbb{Z}^n}|P^{(k)}f(\xi)|^2\Big)^{\frac{1}{2}}$$

This leads to:

THEOREM 8.11. *For $1<p<\infty$ we have*

$$\|S_\square(f)\|_{L^p(\mathbb{R}^n)}\simeq_{p,n}\|f\|_{L^p(\mathbb{R}^n)}.$$

We omit the proof of this theorem as it is mostly technical, based on induction and starting from the one dimensional version of the theorem already proved. You can find the proof for example in [D] or [S].

8.4. Two theorems on multipliers

We now go back to multiplier operators and reconsider them from the point of view of Calderón-Zygmund theory. We have already seen that a multiplier operator is the linear operator T_m with $\widehat{T_mf}=m(\xi)\hat{f}(\xi)$ for some $m\in L^\infty(\mathbb{R}^n)$. This definition automatically implies that T_m is bounded on L^2 with norm $\|T_m\|_{L^2\to L^2}=\|m\|_{L^\infty}$. Alternatively, the discussion from § 3.7 reveals that these are all the bounded linear operators on L^2 that commute with translations and can be realized in the form

$$T_mf(x)=(K*f)(x),\quad f\in\mathcal{S}(\mathbb{R}^n),$$

where $K\in\mathcal{S}'(\mathbb{R}^n)$ is the unique tempered distribution such that $\hat{K}=m$.

If the operator T_m extends to a bounded linear operator on $L^p(\mathbb{R}^n)$ we say that m is an L^p-multiplier and write $m\in\mathcal{M}^p$. We set

$$\|m\|_{\mathcal{M}^p}:=\|T_m\|_{L^p\to L^p}.$$

The previous remarks then show that $\|m\|_{\mathcal{M}^2}=\|m\|_{L^\infty}$. It turns out that the space $(\mathcal{M}^p,\|\cdot\|_{\mathcal{M}^p})$ is a Banach space but we will not dwell on this issue here. We also have the following easy proposition:

PROPOSITION 8.12. *We have the following statements:*

(i) *Let $1 \le p \le \infty$ and p' be the conjugate exponent of p. Then*

$$m \in \mathcal{M}^p \Leftrightarrow m \in \mathcal{M}^{p'}$$

and in this case we have that

$$\|m\|_{\mathcal{M}^p} = \|m\|_{\mathcal{M}^{p'}}.$$

(ii) *For all $1 \le p \le \infty$ we have*

$$\|m\|_{L^\infty(\mathbb{R}^n)} \le \|m\|_{\mathcal{M}^p}.$$

PROOF. The claim in (i) is a consequence of the following obvious identity; for $f, g \in \mathcal{S}(\mathbb{R}^n)$ we have

$$\int_{\mathbb{R}^n} T_m f(x)\overline{g(x)} = \int_{\mathbb{R}^n} m(\xi)\hat{f}(\xi)\overline{\hat{g}(\xi)} = \int_{\mathbb{R}^n} f(x) T_{\overline{m}} g(x) dx,$$

That is, $T_{\overline{m}}$ is the adjoint of T_m. Thus

$$\|m\|_{\mathcal{M}^p} = \|T_m\|_{L^p \to L^p} = \|T_{\overline{m}}\|_{L^{p'} \to L^{p'}} = \|\overline{m}\|_{\mathcal{M}^{p'}} = \|m\|_{\mathcal{M}^{p'}},$$

since m and $\overline{m}$ have the same norm. To prove the second assertion assume that $\|m\|_{\mathcal{M}^p} < +\infty$ otherwise there is nothing to prove. By (i), the linear operator T_m is of strong type (p, p) and (p', p') with the same operator norm. By the Riesz-Thorin interpolation theorem we get that

$$\|m\|_{L^\infty(\mathbb{R}^n)} = \|T_m\|_{L^2 \to L^2} \le \|m\|_{\mathcal{M}^p}^{\theta} \|m\|_{\mathcal{M}^{p'}}^{1-\theta} = \|m\|_{\mathcal{M}^p},$$

which proves (ii). □

REMARK 8.13. Observation (ii) above shows that L^p multipliers are necessarily bounded functions. The opposite however is not true. Another easy consequence of the discussion above is the following. We always have

$$T_m f(x) = (K * f)(x) = (\check{m} * f)(x),$$

where $K = \check{m} \in \mathcal{S}'(\mathbb{R}^n)$ as observed above. The problem with this representation is that we don't know whether $K = \check{m}$ is actually a function that can give meaning to the formula

$$T_m f(x) = \int_{\mathbb{R}^n} K(x - y) f(y) dy.$$

If however it happens that $K = \check{m} \in L^1(\mathbb{R}^n)$ then Young's inequality readily applies to show that

$$\|T_m f\|_{L^p(\mathbb{R}^n)} \le \|K\|_{L^1(\mathbb{R}^n)} \|f\|_{L^p(\mathbb{R}^n)},$$

so that

$$\|m\|_{\mathcal{M}^p} = \|T_m\|_{L^p \to L^p} \le \|\check{m}\|_{L^1(\mathbb{R}^n)}.$$

The main problem in the theory of multipliers is to get away from the case $p = 2$ and place suitable conditions on m so that we can conclude that $m \in \mathcal{M}^p$. The previous generalities easily imply that if $m \in \mathcal{S}(\mathbb{R}^n)$ then $m \in \mathcal{M}^p$ since $\check{m} \in L^1(\mathbb{R}^n)$ in this case. A similar result with weaker hypothesis requires the following definition.

DEFINITION 8.14. We define the Sobolev space $W^{s,2}$ to be the space of tempered distributions f such that $\hat{f}$ agrees with a function almost everywhere that satisfies

$$\|f\|_{W^{s,2}}^2 := \int_{\mathbb{R}^n} |\hat{f}(\xi)(1+4\pi^2|\xi|^2)^{\frac{s}{2}}|^2 d\xi < +\infty.$$

We then have the following simple multiplier theorem.

PROPOSITION 8.15. *Let $m : \mathbb{R}^n \to \mathbb{C}$ and suppose that $m \in W^{s,2}$ for some $s > n/2$. Then $m \in \mathcal{M}^p$ and $\|m\|_{\mathcal{M}^p} \lesssim_{n,s} \|m\|_{W^{s,2}}$.*

PROOF. We will show that $\check{m} \in L^1(\mathbb{R}^n)$. We have

$$\begin{aligned}\int_{\mathbb{R}^n} |\check{m}(x)|dx &= \int_{\mathbb{R}^n} |\hat{\check{m}}(x)|(1+4\pi^2|x|^2)^{\frac{s}{2}}(1+4\pi^2|x|^2)^{-\frac{s}{2}}dx \\ &\leq \|m\|_{W^{s,2}}\Big(\int_{\mathbb{R}^n}(1+4\pi^2|x|^2)^{-s}dx\Big)^{\frac{1}{2}} \lesssim_{n,s} \|m\|_{W^{s,2}}\end{aligned}$$

since $2s > n$. Thus we have

$$\|T_m f\|_{L^p(\mathbb{R}^n)} = \|\check{m} * f\|_{L^p(\mathbb{R}^n)} \leq \|\check{m}\|_{L^1(\mathbb{R}^n)}\|f\|_{L^p(\mathbb{R}^n)} \lesssim_s \|m\|_{W^{s,2}}\|f\|_{L^p(\mathbb{R}^n)}$$

which proves the desired estimate. □

REMARK 8.16. Observe that for any tempered distribution f we have that

$$\widehat{(-\Delta f)}(\xi) = 4\pi^2|\xi|^2\hat{f}(\xi).$$

If k is an even integer we can write

$$\mathcal{F}((I-\Delta)^{\frac{k}{2}}f)(\xi) = (1+4\pi^2|\xi|^2)^{\frac{k}{2}}\hat{f}(\xi).$$

Thus, at least when k is an even integer, the Sobolev space $W^{k,2}$ is the space of tempered distributions such that

$$(I-\Delta)^{\frac{k}{2}}f \in L^2(\mathbb{R}^n),$$

where $(I-\Delta)^{\frac{k}{2}}$ makes sense as a partial differentiable operator since $k/2$ is an integer. Similarly one can define the Sobolev spaces $W^{k,p}$ to be the space of tempered distributions f such that

$$(I-\Delta)^{\frac{k}{2}}f \in L^p(\mathbb{R}^n).$$

In fact one can take one step further and define the space $W^{s,p}$ for any real number s and $1 < p < +\infty$. In the case $p = 2$ this presents no difficulty since one has a direct interpretation of $(I-\Delta)^{\frac{s}{2}}$ as a Fourier integral operator. In particular, $(I-\Delta)^{\frac{s}{2}}$ is the multiplier operator defined as

$$(I-\Delta)^{\frac{s}{2}}f := \mathcal{F}^{-1}\Big((1+2\pi|\xi|^2)^{\frac{s}{2}}\hat{f}\Big).$$

With this discussion we have

$$W^{s,p} := \{f \in L^p(\mathbb{R}^n) : \|(I-\Delta)^{\frac{s}{2}}f\|_{L^p} < +\infty\}.$$

The general flavor of the previous results is that if a function m has no local singularities and, together with its derivatives, decays fast enough at infinity, then m is an $\mathcal{M}^p$ multiplier for all $1 \leq p \leq \infty$. Besides a (controllable) singularity at infinity, one can also allow for a singularity at the origin.

We present two instances of this principle, usually referred to as the Hörmander multiplier theorem. We start with an "easy" version where the function m is bounded, to assure the $(2,2)$ hypothesis is satisfied, C^∞ away from the origin and its derivatives decay at least as fast as their order.

THEOREM 8.17 (Hörmander-Mikhlin multiplier theorem version I). *Let $m : \mathbb{R}^n \to \mathbb{C}$ be a bounded function which belongs to the class $C^\infty(\mathbb{R}^n \setminus \{0\})$ and satisfies*

$$|\partial_\xi^\alpha m(\xi)| \lesssim_{n,\alpha} |\xi|^{-|\alpha|}, \quad \xi \in \mathbb{R}^n \setminus \{0\},$$

for all multi-indices α. Then $K = \check{m}$ agrees with a C^∞ function away from the origin and satisfies

$$|\partial^\alpha K(x)| \lesssim_\alpha |x|^{-n-|\alpha|}, \quad x \in \mathbb{R}^n \setminus \{0\},$$

for all multi-indices α. In particular, m is an $\mathcal{M}^p$ multiplier for all $1 < p < \infty$ with $\|m\|_{\mathcal{M}^p} \lesssim_{p,n} 1$.

PROOF. Using the Littlewood-Paley decomposition we can write

$$m(\xi) = \sum_{j\in\mathbb{Z}} \psi(\xi/2^k) m(\xi) =: \sum_{j\in\mathbb{Z}} m_j(\xi),$$

whenever $\xi \neq 0$. Each piece m_j is supported on the annulus $2^{j-1} \leq |\xi| \leq 2^{j+1}$ and is a C^∞-function as a product of smooth functions so it makes sense to define

$$K_j(x) = \int_{\mathbb{R}^n} m_j(\xi) e^{2\pi i x\cdot\xi} d\xi = \check{m}_j(x).$$

Furthermore, from our hypotheses on m we can get some good estimates on each K_j together with its derivatives. Indeed since $\|m_j\|_{L^\infty} \leq \|m\|_{L^\infty} \lesssim_n 1$ by our hypothesis (with the zero multi-index α) we have

$$|K_j(x)| \leq \int_{|\xi|\simeq 2^j} |m_j(\xi)| d\xi \lesssim_n 2^{jn}.$$

Likewise

$$|\partial^\alpha K_j(x)| \leq \int_{|\xi|\simeq 2^j} |(2\pi i\xi)^\alpha m_j(\xi)| d\xi \lesssim_{n,\alpha} \int_{|\xi|\simeq 2^j} d\xi \leq 2^{j(n+|\alpha|)}.$$

On the other hand for every multi-index α we have

$$|\partial^\alpha K_j(x)| = \Big| \int_{|\xi|\simeq 2^j} (2\pi i\xi)^\alpha m_j(\xi) \Big(\frac{x\cdot\nabla_\xi}{2\pi i|x|^2}\Big)^M e^{2\pi i x\cdot\xi} d\xi \Big|,$$

for every non-negative integer M. Integrating by parts M times to pass the derivatives to the term $(2\pi i\xi)^\alpha m_j(\xi)$, using Leibniz's rule and the hypothesis on the derivatives $\partial^\alpha m_j$ we get the estimate

$$|\partial^\alpha K_j(x)| \lesssim_{n,\alpha,M} |x|^{-M} 2^{j(n+|\alpha|-M)},$$

for all multi-indices α and non-negative integers M. We summarize these estimates in the form

$$|\partial^\alpha K_j(x)| \lesssim_{n,\alpha,M} \min(2^{j(n+|\alpha|)}, |x|^{-M} 2^{j(n+|\alpha|-M)}) \tag{8.9}$$

for all multi-indices α and non-negative integers M. Using (8.9) for $M = 0$ we have

$$\sum_{2^j \leq |x|^{-1}} |\partial^\alpha K_j(x)| \leq \sum_{2^j \leq |x|^{-1}} 2^{j(n+|\alpha|)} \lesssim_{n,\alpha} |x|^{-(n+\alpha)}$$

On the other hand, using (8.9) for $M > n + |\alpha|$ we get

$$\sum_{2^j > |x|^{-1}} |\partial^\alpha K_j(x)| \leq |x|^{-M} \sum_{2^j > |x|^{-1}} 2^{j(n+|\alpha|-M)} \simeq_{n,\alpha} |x|^{-M} |x|^{-(n+|\alpha|-M)} = |x|^{-(n+|\alpha|)}.$$

Now since the series $\sum_j \partial^\alpha K_j(x)$ converges absolutely whenever $x \neq 0$ we conclude that for every K the series $\sum_j K_j$ converges locally in C^∞_{loc}, away from a neighborhood of 0, to some function $\tilde{K} \in C^\infty(\mathbb{R}^n \setminus \{0\})$ that satisfies

$$|\partial^\alpha \tilde{K}(x)| \lesssim_{n,\alpha} \frac{1}{|x|^{n+|\alpha|}} \tag{8.10}$$

for every multi-index α. Since $\sum_j m_j = \sum_j \widehat{K_j}$ converges to $m = \hat{K}$ in $L^2(\mathbb{R}^n)$ we conclude that $K(x) = \tilde{K}(x)$ when $x \neq 0$. In particular,

$$T_m f = (K * f)(x) = \int \tilde{K}(x-y) f(y) dy,$$

whenever $f \in L^2(\mathbb{R}^n)$ has compact support and $x \notin \text{supp}(f)$ since then $x - y \neq 0$. However, $\tilde{K}$ satisfies

$$|\tilde{K}(x-y)| \lesssim_n |x-y|^{-n}, \quad x \neq y,$$

by taking the zero multi-index in (8.10) and furthermore

$$|\nabla_y K(x-y)| \lesssim_n |x-y|^{-(n+1)}, \quad |\nabla_x K(x-y)| \lesssim_n |x-y|^{-(n+1)}, \quad x \neq y,$$

by considering multi-indices α with $|\alpha| = 1$ in (8.10). These estimates are enough to assure that $\tilde{K}$ and thus K is a singular kernel so T_m is a CZO associated with K. However this means that $m \in \mathcal{M}^p$ with $\|m\|_{\mathcal{M}^p} \lesssim_{n,p} 1$ and we are done. $\square$

REMARK 8.18. The hypothesis of the previous theorem is not optimal as one can get away with less derivatives of m. However it already applies to many practical cases. For example for any multi-index β of order $|\beta| = 2$, consider the operator T_m with symbol

$$m_\beta(\xi) = \frac{\xi^\beta}{|\xi|^2}.$$

Observe that m_β falls under the scope of Theorem 8.17 since for $\xi \neq 0$ we have

$$|\partial^\alpha m_\beta(\xi)| \lesssim \frac{1}{|\xi|^{|\alpha|}},$$

for all multi-indices α. So $m_\beta \in \mathcal{M}^p$ for all $1 < p < \infty$. Now observe that for $f \in \mathcal{S}(\mathbb{R}^n)$ (say) we have

$$\widehat{(\partial^\beta f)}(\xi) = (2\pi i \xi)^\beta \hat{f}(\xi) = \frac{(2\pi i \xi)^\beta}{-4\pi^2 |\xi|^2} (-4\pi^2 |\xi|^2) \hat{f}(\xi) = m_\beta(\xi) \widehat{\Delta f}(\xi).$$

which shows in particular that

$$\|\partial^\beta f\|_{L^p} \lesssim \|\Delta f\|_{L^p}$$

for all multi-indices of order 2, whenever $\Delta f \in L^p(\mathbb{R}^n)$. Thus all partial derivatives of order 2 are controlled by the Laplacian in L^p.

Now consider the space $W^{2,p}(\mathbb{R}^n)$ to be the space of L^p functions f such that all the partial derivatives of order up to 2 are in L^p and equip this space with the norm

$$\|f\|_{W^{2,p}} := \sum_{0\leq|\alpha|\leq 2} \|\partial^\alpha f\|_{L^p(\mathbb{R}^n)}.$$

By the remarks above this norm is equivalent to

$$\|f\|_{W^{2,p}} \simeq_{n,p} \|(I-\Delta)f\|_{L^p(\mathbb{R}^n)}.$$

Similar conclusions hold for any $s \geq 0$ and the space $W^{s,p}$. Thus the two definitions of the Sobolev space $W^{k,p}$, the one given here and then one given in Remark 8.13 coincide whenever $s \geq 0$:

$$\|f\|_{W^{s,p}} = \sum_{0\leq|\alpha|\leq k} \|\partial^\alpha f\|_{L^p(\mathbb{R}^n)} \simeq_{n,p} \|(I-\Delta)^{\frac{s}{2}} f\|_{L^p(\mathbb{R}^n)}, \quad s \geq 0.$$

We now give a sharper form of the multiplier theorem which requires control only on $\sim n/2$ derivatives of m.

THEOREM 8.19 (Hörmander-Mikhlin multiplier theorem version II). *Let m be a bounded function on $\mathbb{R}^n$.*

(i) *Let k be the smallest integer $> n/2$ and suppose that the multiplier $m : \mathbb{R}^n \to \mathbb{C}$ is of class $C^k(\mathbb{R}^n \setminus \{0\})$ with*

$$|\partial^\alpha m(\xi)| \lesssim_\alpha |\xi|^{-|\alpha|},$$

for all multi-indices α with $|\alpha| \leq k$. Then $\check{m}$ agrees with a function $K(x)$ away from the origin which is locally integrable away from the origin and satisfies

$$\int_{|x|>2|y|} |K(x-y) - K(x)| dx \lesssim_n 1,$$

for all $y \neq 0$.

(ii) *Under the assumptions of (i) we have that $m \in \mathcal{M}^p$ for all $1 < p < \infty$ and $\|m\|_{\mathcal{M}^p} \lesssim_{n,p} 1$.*

PROOF. As in the proof of Theorem 8.17 it will be enough to control the pieces K_j. For this, let β be a multi-index. We have

$$\int_{\mathbb{R}^n} |(-2\pi i x)^\beta K_j(x)|^2 dx = \int_{\mathbb{R}^n} |\partial_\xi^\beta m_j(\xi)|^2 d\xi.$$

For $M \leq k$ this implies that

$$\int_{\mathbb{R}^n} (|x|^M)^2 |K_j(x)|^2 dx = \int_{\mathbb{R}^n} (x_1^2 + \cdots + x_n^2)^M |K_j(x)|^2 dx \lesssim_{n,M} 2^{nj} 2^{-2Mj}.$$

Now for any $R > 0$ we have

$$\int_{|x|\leq R} |K_j(x)| dx \leq \Big(\int_{|x|\leq R} |K_j(x)|^2 dx \Big)^{\frac{1}{2}} R^{\frac{n}{2}} \lesssim_{n,M} 2^{\frac{nj}{2}} R^{\frac{n}{2}}.$$

On the other hand

$$\int_{|x|>R} |K_j(x)| dx = \int_{|x|>R} |K_j(x)||x|^k |x|^{-k} dx \leq 2^{\frac{nj}{2}} 2^{-kj} R^{n/2-k}, \tag{8.11}$$

where $n/2 - k < 0$. Choosing $R = 2^{-j}$ these estimates imply that

$$\int_{\mathbb{R}^n} |K_j(x)|dx \lesssim_n 1 + 2^{j(n-2k)} \lesssim_n 1.$$

We will now prove a similar estimate for the derivatives of K_j using a very similar approach. Indeed, we start from the identity

$$\int_{\mathbb{R}^n} |(-2\pi i x)^\beta \partial^\alpha K_j(x)dx|^2 dx = \int_{\mathbb{R}^n} |\partial_\xi^\beta[(2\pi i\xi)^\alpha m_j(\xi)]|^2 d\xi.$$

Now for $M \leq k$ and using the Leibniz rule we get

$$\int_{\mathbb{R}^n} (|x|^M)^2 |\partial^\alpha K_j(x)|^2 dx \lesssim_{M,\alpha} 2^{nj} 2^{-2Mj} 2^{2j|\alpha|}.$$

Thus we have

$$\int_{|x|\leq R} |\partial^\alpha K_j(x)|dx \leq \Big(\int_{\mathbb{R}^n} |\partial^\alpha K_j(x)|^2\Big)^{\frac{1}{2}} R^{\frac{n}{2}} \lesssim_{\alpha,n} 2^{\frac{nj}{2}} 2^{j|\alpha|} R^{\frac{n}{2}}.$$

Also

$$\int_{|x|>R} |\partial^\alpha K_j(x)|dx = \int_{|x|>R} |x|^{-k}|x|^k |\partial^\alpha K_j(x)|dx \lesssim_{n,\alpha,k} 2^{\frac{nj}{2}} 2^{j|\alpha|} 2^{-kj} R^{\frac{n}{2}-k}.$$

Choosing $R = 2^{-j}$ and combining the last two estimates we conclude

$$\int_{\mathbb{R}^n} |\partial^\alpha K_j(x)|dx \lesssim_{n,\alpha} 2^{j|\alpha|}.$$

This estimate for $|\alpha| = 1$ together with the mean value theorem implies that

$$\int_{\mathbb{R}^n} |K_j(x+h) - K_j(x)|dx \lesssim_n 2^j|h|.$$

We now have for all $y \neq 0$

$$\sum_{2^j \leq |y|^{-1}} \int_{|x|\geq 2|y|} |K_j(x-y) - K(x)|dx \lesssim_n \sum_{2^j\leq|y|^{-1}} 2^j|y| \lesssim_{n,k} 1.$$

On the other hand

$$\begin{aligned}\sum_{2^j > |y|^{-1}} \int_{|x|\geq 2|y|} |K_j(x-y) - K(x)|dx &\lesssim \sum_{2^j>|y|^{-1}} \int_{|x|\geq|y|} |K_j(x)|dx \\ &\lesssim_n \sum_{2^j>|y|^{-1}} 2^{\frac{nj}{2}} 2^{-kj} |y|^{\frac{n}{2}-k} \lesssim_{n,k} 1,\end{aligned}$$

by (8.11). Using now that $\sum_j K_j(x)$ converges in $L^1(V)$ to some locally integrable function for every compact set V that doesn't contain 0 we conclude that K coincides with a locally integrable function away from 0 and satisfies

$$\int_{|x|\geq 2|y|} |K(x-y) - K(x)|dx \lesssim_n 1, \tag{8.12}$$

for $y \neq 0$. Since $K = \check{m}$ away from the origin we have that T_m satisfies

$$T_m f(x) = \int K(x-y) f(y) dy$$

whenever f in $L^2(\mathbb{R}^n)$ and has compact support and $x \notin \operatorname{supp}(f)$. Furthermore, by the assumption $m \in L^\infty(\mathbb{R}^n)$ we automatically get that T_m is bounded on

$L^2(\mathbb{R}^n)$. Here condition (8.12) is enough to substitute the conditions given in the definition of a singular kernel and show that T_m is a CZO with K playing the role of the kernel. Indeed, the $(2,2)$ type of T_m can be used to treat the bad part in the Calderón-Zygmund decomposition of a function f. On the other hand, if b_Q is a bad piece supported on a dyadic cube Q with center w_Q and $Q^* = (1+2\sqrt{n})Q$ is the cube with the same center and $(1+2\sqrt{n})$ times the side-length of Q, we have

$$\int_{\mathbb{R}^n\setminus Q^*}\left|\int_Q K(x-y)b_Q(y)dy\right|dx \leq \int_{\mathbb{R}^n\setminus Q^*}\int_Q |K(x-y)-K(x-w_Q)|\,|b_Q(y)|\,dy\,dx$$
$$\leq \int_Q |b_Q(y)|\int_{\mathbb{R}^n\setminus Q^*}|K(x-y)-K(x-w_Q)|\,dx\,dy.$$

Now if $y\in Q$ and $x\notin Q^*$ we have that $|x-w_Q|\geq \sqrt{n}\,\mathrm{side}(Q)\geq 2|y-w_Q|$. Thus for $y\in Q$ we have from (8.12) that

$$\int_{\mathbb{R}^n\setminus Q^*}|K(x-y)-K(x-w_Q)|dx = \int_{\mathbb{R}^n\setminus Q^*}|K(x-w_Q-(y-w_Q))-K(x-w_Q)|dx \lesssim_n 1,$$

so that

$$\int_{\mathbb{R}^n\setminus Q^*}|\int_Q K(x-y)b_Q(y)dy|dx \lesssim_n \|b_Q\|_{L^1(Q)}.$$

This treats the bad part of the Calderón-Zygmund decomposition of f so we conclude the proof that T_m is of weak type $(1,1)$ as in the general case of a CZO. Interpolating between this bound and the strong $(2,2)$ bound we get that $m\in\mathcal{M}^p$ for $1<p<2$. By Proposition 8.12 or using the symmetry of $K(x-y)$ in x and y, we also get the range $2<p'<\infty$ with $\|m\|_{\mathcal{M}^p}=\|m\|_{\mathcal{M}^{p'}}=\|T_m\|_{L^p\to L^p}$. □

EXERCISE 8.20. The purpose of this exercise is to clear out some of the calculations in the proofs of the two versions of Hörmander's theorem.

(i) Prove the identity

$$\left(\frac{x\cdot\nabla_\xi}{2\pi i|x|^2})\right)^N e^{2\pi i x\cdot\xi} = e^{2\pi i x\cdot\xi},$$

for any positive integer N. Here the meaning of the symbol $x\cdot\nabla_\xi$ is

$$x\cdot\nabla_\xi := x_1\partial_{\xi_1}+\cdots+x_m\partial_{\xi_m}.$$

(ii) Let $\alpha=(\alpha_1,\ldots,\alpha_n)$ and $\beta=(\beta_1,\ldots,\beta_n)$ be two multi-indices in $\mathbb{N}^n$. We write $\beta\leq\alpha$ if $\beta_j\leq\alpha_j$ for all $j\in\{1,2,\ldots,n\}$. With this notation the *Leibniz rule* says that for any multi-index α and functions f,g we have

$$\partial^\alpha(fg)=\sum_{\beta\leq\alpha}\binom{\alpha}{\beta}(\partial^{\alpha-\beta}f)(\partial^\beta g).$$

Here the *generalized binomial coefficients* $\binom{\alpha}{\beta}$ are defined as

$$\binom{\alpha}{\beta}:=\binom{\alpha_1}{\beta_1}\binom{\alpha_2}{\beta_2}\cdots\binom{\alpha_n}{\beta_n}.$$

Alternatively we use the notation

$$\alpha! := \alpha_1!\cdots\alpha_n! \quad \text{so that} \quad \binom{\alpha}{\beta}=\frac{\alpha!}{\beta!(\alpha-\beta)!},\quad \beta\leq\alpha.$$

For any two multi-indices $\alpha, \beta \in \mathbb{N}_o^n$ show that

$$\partial_x^\alpha (x^\beta) = \begin{cases} \frac{\beta!}{(\beta-\alpha)!} x^{\beta-\alpha}, & \alpha \le \beta, \\ 0, & \text{otherwise.} \end{cases}$$

(iii) Let $m : \mathbb{R}^n \to \mathbb{C}$ satisfy the estimate

$$|\partial^\alpha m(\xi)| \lesssim_{n,\alpha} |\xi|^{-|\alpha|},$$

and let $\psi_j(\xi) = \psi(\xi/2^j)$, $j \in \mathbb{Z}$, be as in the Littlewood-Paley decomposition. Show that $m_j = m\psi_j$ satisfies the same estimates, that is,

$$|\partial^\alpha m_j(\xi)| \lesssim_{n,\alpha} |\xi|^{-|\alpha|},$$

with different implied constants of course. Remember that ψ_j and thus m_j is supported on $|\xi| \simeq 2^j$.

(iv) Let $m : \mathbb{R}^n \to \mathbb{C}$ satisfy the estimate

$$|\partial^\alpha m(\xi)| \lesssim_{n,\alpha} |\xi|^{-|\alpha|},$$

and $\psi_j = \psi(\xi/2^j)$, $j \in \mathbb{Z}$, be as in the Littlewood-Paley decomposition. Set $m_j = m\psi_j$. Show that for every multi-index γ of order $|\gamma| = M$ we have

$$|\partial^\gamma((2\pi i\xi)^\alpha m_j(\xi))| \lesssim_{n,\alpha,M} |\xi|^{|\alpha|-M}.$$

(v) Let h be a smooth function which is supported on $A_k := \{2^{k-1} \le |\xi| \le 2^{k+1}\}$. Show that

$$\int_{A_k} h(\xi)\Big(\frac{x\cdot\nabla_\xi}{2\pi i|x|^2}\Big)e^{2\pi i x\cdot\xi}d\xi = \int_{A_k}\Big[\Big(-\frac{x\cdot\nabla_\xi}{2\pi i|x|^2}\Big)h(\xi)\Big]e^{2\pi i x\cdot\xi}d\xi,$$

and by iterating that

$$\int_{A_k} h(\xi)\Big(\frac{x\cdot\nabla_\xi}{2\pi i|x|^2}\Big)^N e^{2\pi i x\cdot\xi}d\xi = \int_{A_k}\Big[\Big(-\frac{x\cdot\nabla_\xi}{2\pi i|x|^2}\Big)^N h(\xi)\Big]e^{2\pi i x\cdot\xi}d\xi,$$

for all positive integers N.

EXERCISE 8.21. Let $K \in L^2(\mathbb{R}^n)$ be such that $m := \hat{K} \in L^\infty(\mathbb{R}^n)$. Furthermore suppose that K satisfies the mean regularity condition

$$\int_{|x|>2|y|} |K(x-y) - K(x)|dx \lesssim_n 1, \quad y \ne 0.$$

Show that $m \in \mathcal{M}^p(\mathbb{R}^n)$.

Hint: Briefly describe the key elements of the proof showing that $T_m f = K * f$ is of weak type $(1,1)$. Argue why this implies that $m \in \mathcal{M}^p$ for $1 < p < 2$. You get the complementary interval $2 < p < \infty$ for free (why?).

8.5. The Sobolev spaces $W^{s,p}(\mathbb{R}^n)$

Let $s \ge 0$ be a non-negative integer and $1 \le p \le \infty$. The Sobolev space $W^{s,p}(\mathbb{R}^n)$ is defined to be the space of all functions $f \in L^p(\mathbb{R}^n)$ whose distributional (or weak) derivatives of orders $\le k$ belong to $L^p(\mathbb{R}^n)$. We can equip this space with the norm

$$\|f\|_{W^{s,p}(\mathbb{R}^n)} := \sum_{0\le|\alpha|\le k} \|\partial^\alpha f\|_{L^p(\mathbb{R}^n)}.$$

For $s = 0$ we have of course $W^{0,p}(\mathbb{R}^n) = L^p(\mathbb{R}^n)$ and for $s > 0$ the space $W^{s,p}$ is a linear subspace of $L^p(\mathbb{R}^n)$. In fact the space $W^{s,p}$, equipped for example with the norm given above, is a Banach space for all $1 \leq p \leq \infty$ and non-negative integers s.

An equivalent way to define and gauge Sobolev functions is the following. For $j \geq 0$ a non-negative integer let ∇^j denote the vector of all the distributional partial derivatives of f of order j. Thus $\nabla^j f$ is a vector of n^j partial derivatives of f of order j. Then let us set

$$|\nabla^j f| := \Big(\sum_{|\alpha|=j} |\partial^\alpha f|^2\Big)^{\frac{1}{2}}.$$

With these definitions it is elementary to see that we also have

$$\|f\|_{W^{s,p}(\mathbb{R}^n)} \simeq_{n,p,s} \sum_{j=0}^{s} \|\nabla^j f\|_{L^p(\mathbb{R}^n)}.$$

Using the Fourier transform and the way it interacts with (weak) derivatives we get another expression which is comparable to the $W^{s,p}$-norm in the range $1 < p < +\infty$. This is the analogue of Proposition 8.4 for the differential operator ∇^s instead of $\nabla = \nabla^1$.

PROPOSITION 8.22. *For $s \geq 0$ a non-negative integer and $1 < p < \infty$ we have*

$$\|\nabla^s f\|_{L^p(\mathbb{R}^n)} \simeq_{n,p,s} \Big\|\Big(\sum_{k\in\mathbb{Z}} |2^{ks}\Delta_k f|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R}^n)}.$$

We conclude that

$$\|f\|_{W^{s,p}(\mathbb{R}^n)} \simeq_{n,p,s} \Big\|\Big(\sum_{k\in\mathbb{Z}} |(1+2^k)^s\Delta_k f|^2\Big)^{\frac{1}{2}}\Big\|_{L^p(\mathbb{R}^n)}.$$

PROOF. Observe that the case $s = 0$ of the proposition is Theorem 8.6 while the case $s = 1$ is compatible with Proposition 8.4.

Observe that $\nabla^s f = (\partial^\alpha f)_{|\alpha|=s}$ is a vector with n^s components, each component being of the form $\partial^\alpha f$ with $|\alpha| = s$. As in the proof of Proposition 8.4 we have

$$\Delta_k f(x) = S_{k+1}\Delta_k f(x) = \int_{\mathbb{R}^n} \Delta_k f(x - 2^{-(k+1)y})\hat{\phi}(y)dy.$$

Thus we have the vector identity

$$\begin{aligned}
\nabla_x^s \Delta_k f(x) &= \int_{\mathbb{R}^n} \nabla_x^s \Delta_k f(x - 2^{-(k+1)}y)\hat{\phi}(y)dy \\
&= -2^{s(k+1)} \int_{\mathbb{R}^n} \nabla_y^s \Delta_k f(x - 2^{-(k+1)}y)\hat{\phi}(y)dy \\
&= 2^{s(k+1)} \int_{\mathbb{R}^n} \Delta_k f(x - 2^{-(k+1)}y)\nabla_y^s\hat{\phi}(y)dy \\
&\simeq_s (\mathrm{Dil}^1_{2^{-k}} \nabla_y^s\hat{\phi} * 2^{sk}\Delta_k f)(x).
\end{aligned}$$

Now the operation

$$h \mapsto \Delta_k^* h := \mathrm{Dil}^1_{2^{-k}} \nabla_y^s\hat{\phi} * h$$

is a (modified) Littlewood-Paley projection at frequency $|\xi| \simeq 2^k$. Thus we have the identity

$$\Delta_k \nabla_x^s f \simeq_s 2^{sk} \Delta_k^* \Delta_k f$$

We then have

$$\|\nabla^s f\|_{L^p(\mathbb{R}^n)} \leq \Big\| \sum_k \Delta_k (\nabla^s f) \Big\|_{L^p(\mathbb{R}^n)} \lesssim_s \Big\| \sum_k \Delta_k^* 2^{sk} \Delta_k f \Big\|_{L^p(\mathbb{R}^n)}.$$

Using estimate (8.4) for the operators Δ_k^* we get one side of the desired estimate

$$\|\nabla^s f\|_{L^p(\mathbb{R}^n)} \lesssim_{s,p,n} \Big\| \Big(\sum_k |2^{sk} \Delta_k f|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p(\mathbb{R}^n)}.$$

The converse estimate is slightly more involved, but again follows the ideas in the proof of Proposition 8.4. Choosing a function $\rho \in \mathcal{S}(\mathbb{R}^n)$ which is identically 1 on $\{\frac{1}{2} \leq |\xi| \leq 2\}$ and vanishes outside $\{\frac{1}{4} \leq |\xi| \leq 4\}$ we can write the identity

$$\rho(\xi/2^k)\widehat{\Delta_k \partial_x^\alpha f}(\xi) = \rho(\xi/2^k)\psi(\xi/2^k)\widehat{\partial_x^\alpha f}(\xi) = (2\pi i \xi)^\alpha \widehat{\Delta_k f}(\xi)$$

for every multi-index α with $|\alpha| = s$. The generalized binomial theorem now states that for $\xi = (\xi_1, \ldots, \xi_n)$ we have

$$|\xi|^2 = (\xi_1^2 + \cdots + \xi_n^2)^s = \sum_{|\alpha|=s} \frac{s!}{\alpha_1! \cdots \alpha_n!} \xi^\alpha =: \sum_{|\alpha|=s} c_{s,\alpha} \xi^\alpha.$$

Thus we have

$$\widehat{\Delta_k f}(\xi) = \frac{1}{(2\pi i)^s} \sum_{|\alpha|=s} c_{\alpha,s} \rho(\xi/2^k) \frac{\xi^\alpha}{|\xi|^{2s}} \widehat{\Delta_k \partial_x^\alpha f}(\xi).$$

Inverting the Fourier transform we get

$$\Delta_k f(x) = 2^{-ks} \sum_{|\alpha|=s} K_{k,\alpha,s} * \Delta_k \partial_x^\alpha f(x)$$

where

$$K_{k,\alpha,s}(x) = \frac{c_{\alpha,s}}{(2\pi i)^s} 2^{ks} \int_{\mathbb{R}^n} \rho(\xi/2^k) \frac{\xi^\alpha}{|\xi|^{2s}} e^{2\pi i x \cdot \xi} d\xi.$$

Again we see that for each α with $|\alpha| = s$ the operation

$$h \mapsto \Delta_{k,\alpha} h := K_{k,\alpha,s} * h$$

is a (modified) Littlewood-Paley at frequencies $|\xi| \simeq 2^k$. Thus we have proved the identity

$$2^{ks} \Delta_k f = \sum_{|\alpha|=s} \Delta_{k,\alpha} \Delta_k \partial_x^\alpha f = \sum_{|\alpha|=s} \Delta_{k,\alpha}^* \partial_x^\alpha f$$

for some other Littlewood-Paley projection Δ_k^* at frequency $|\xi| \simeq 2^k$. Thus

$$\begin{aligned} \Big\| \Big(\sum_k 2^{ks} |\Delta_k f|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p(\mathbb{R}^n)} &\lesssim \sum_{|\alpha|=s} \Big\| \Big(\sum_k |\Delta_{k,\alpha}^* \partial_x^\alpha f|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p(\mathbb{R}^n)} \\ &\simeq_{n,p} \sum_{|\alpha|=s} \|\partial_x^\alpha f\|_{L^p(\mathbb{R}^n)} \simeq \|\nabla^s f\|_{L^p(\mathbb{R}^n)}. \end{aligned}$$

These two estimates complete the proof. □

Although the original definition of Sobolev spaces was given in terms of derivatives, thus s was a non-negative integer, the equivalence in Proposition 8.22 makes sense for any $s \in \mathbb{R}$. Thus one can use Proposition 8.22 in order to define Sobolev spaces for any $s \in \mathbb{R}$. This defined the *fractional Sobolev spaces* $W^{s,p}(\mathbb{R}^n)$ for any $s \in \mathbb{R}$ and $1 < p < +\infty$. For $s \geq 0$ these spaces can be thought of as spaces of functions $f \in L^p(\mathbb{R}^n)$ whose distributional *fractional* derivatives of order s belong to L^p.

But what is a fractional derivative? Using the Fourier transform we have the familiar identity:

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

and thus, applying for $|\alpha| = 1$ we have for all non-negative integers s that

$$\widehat{\nabla f}(\xi) = 2\pi i\xi \hat{f}(\xi).$$

We then define the differential operator $|\nabla|$ of order 1 as

$$\widehat{|\nabla| f}(\xi) := 2\pi|\xi|\hat{f}(\xi).$$

Now for any $s \in \mathbb{R}$ we can define the differential operator of *fractional order* s, $|\nabla f|^s$, as

$$\widehat{|\nabla f|^s}(\xi) := (2\pi|\xi|)^s \hat{f}(\xi).$$

When $-n < s < 0$ the operator $|\nabla|^s$ should be understood more as a *fractional integration operator*. By a bit of Fourier analysis we can see that in this case $|\nabla|^s$ has an integral representation as

$$|\nabla|^{-s} f(x) = I_s f(x) := c_{n,s} \int \frac{f(y)}{|x-y|^{n-s}} dy, \quad 0 < s < n.$$

The kernel of the convolution operator above is called a *Riesz potential* (not to be confused with the *Riesz transforms* which are Calderón-Zygmund operators). By the Hardy-Littlewood-Sobolev theorem of Exercise 4.50 we have the estimate

$$\||\nabla|^{-s} f\|_{L^p(\mathbb{R}^n)} \lesssim_{n,p,q,s} \|f\|_{L^q(\mathbb{R}^n)}, \quad 0 < s < n, \quad 1 < p < q < \infty, \quad \frac{1}{q} = \frac{1}{p} - \frac{s}{n}.$$

PROPOSITION 8.23 (Fractional Sobolev spaces). *Let $s \in \mathbb{R}$ and $1 < p < +\infty$. We have the equivalence*

$$\| |\nabla|^s f\|_{L^p(\mathbb{R}^n)} \simeq_{n,p,s} \Big\| \Big(\sum_{k \in \mathbb{Z}} |2^{sk} \Delta_k f|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p(\mathbb{R}^n)}.$$

From this we can conclude that for $s \geq 0$

$$\|f\|_{W^{s,p}(\mathbb{R}^n)} \simeq_{n,p,s} \|f\|_{L^p(\mathbb{R}^n)} + \| |\nabla|^s f\|_{L^p(\mathbb{R}^n)}.$$

For $s < 0$, an equivalent norm for the Sobolev space $W^{s,p}(\mathbb{R}^n)$ can be given as

$$\|f\|_{W^{s,p}(\mathbb{R}^n)} \simeq_{n,p,s} \inf_{f=g+h} \Big(\|g\|_{L^p(\mathbb{R}^n)} + \| |\nabla|^s h\|_{L^p(\mathbb{R}^n)} \Big).$$

A more robust description of the Sobolev space $W^{s,p}(\mathbb{R}^n)$ can be given in terms of “functions of the Laplacian”. To make this precise we remember the identity

$$\widehat{-\Delta f}(\xi) = 4\pi^2|\xi|^2 \hat{f}(\xi)$$

from which we immediately see that $\mathcal{F}((I-\Delta)f)(\xi) = (1+4\pi^2|\xi|^2)\hat{f}(\xi)$. Now one can easily define the first order differential operator $(I-\Delta)^{\frac{1}{2}}$ by means of

$$\mathcal{F}((I-\Delta)^{\frac{1}{2}}f)(\xi) := (1+4\pi^2|\xi|^2)^{\frac{1}{2}}\hat{f}(\xi).$$

The corresponding differential operator of fractional order $s \in \mathbb{R}$ is then

$$\mathcal{F}((I-\Delta)^{\frac{s}{2}}f)(\xi) := (1+4\pi^2|\xi|^2)^{\frac{s}{2}}\hat{f}(\xi).$$

For $s \geq 0$ the operator $(I-\Delta)^{\frac{s}{2}}$ corresponds, to derivatives of order s, plus the identity. For high frequencies $|\xi| \gtrsim 1$ the operator $(I-\Delta)^{\frac{s}{2}}$ behaves like $|\nabla|^s$. However, for small frequencies, $|\xi| \lesssim 1$, the operator $(I-\Delta)^{\frac{s}{2}}$ behaves like the identity unlike the operator $|\nabla|^s$ which damps low frequencies whenever $s \geq 0$.

For $s < 0$ the operator $(I-\Delta)^{\frac{s}{2}}$ is better behaved than $|\nabla|^s$. Indeed, for $s < 0$ the operators $(I-\Delta)^{\frac{s}{2}}$ are known as *smoothing operators* and their kernels as *Bessel potentials*. We also have the notation

$$\mathcal{I}_s f(x) := (I-\Delta)^{-\frac{s}{2}}f(x), \quad s \geq 0.$$

The Sobolev space $W^{s,p}(\mathbb{R}^n)$ can now be described in terms of $(I-\Delta)^{\frac{s}{2}}$:

PROPOSITION 8.24. *For $s \in \mathbb{R}$ and $1 < p < \infty$ we have the equivalent description of $W^{s,p}(\mathbb{R}^n)$ as*

$$\|f\|_{W^{s,p}(\mathbb{R}^n)} \simeq_{n,s,p} \|(I-\Delta)^{\frac{s}{2}}f\|_{L^p(\mathbb{R}^n)}.$$

For $s \geq 0$ we can formally write $(I-\Delta)^{\frac{s}{2}}f =: g \Rightarrow f = \mathcal{I}_s g$; we see that $g \in L^p(\mathbb{R}^n)$ whenever $f \in W^{s,p}(\mathbb{R}^n)$ and $\|f\|_{W^{s,p}(\mathbb{R}^n)} = \|g\|_{L^p(\mathbb{R}^n)}$. Thus for $s \geq 0$ we can symbolically write

$$W^{s,p}(\mathbb{R}^n) \subseteq \mathcal{I}_s(L^p(\mathbb{R}^n)).$$

In fact the smoothing operators $\mathcal{I}_s$ have an integral representation of the form $I_s f(x) = G_s * f(x)$ where G_s are the *Bessel potentials* which satisfy $G_s \geq 0$ and $\|G_s\|_{L^1(\mathbb{R}^n)} = 1$. For every $f \in L^p(\mathbb{R}^n)$ we have

$$\|\mathcal{I}_s f\|_{L^p(\mathbb{R}^n)} \leq \|f\|_{L^p(\mathbb{R}^n)}, \quad 1 \leq p \leq \infty.$$

Thus if $f \in L^p(\mathbb{R}^n)$ then $\mathcal{I}_s f \in L^p(\mathbb{R}^n)$ and obviously $(I-\Delta)^{\frac{s}{2}}\mathcal{I}_s f = f$, $s \geq 0$. We conclude that we also have the inclusion

$$\mathcal{I}_s(L^p(\mathbb{R}^n)) \subseteq W^{s,p}(\mathbb{R}^n), \quad s \geq 0.$$

Combining the previous observations we see that

$$W^{s,p}(\mathbb{R}^n) = \mathcal{I}_s(L^p(\mathbb{R}^n)), \quad s \geq 0, \quad 1 < p < \infty,$$

that is $W^{s,p}(\mathbb{R}^n)$ for $s \geq 0$ is the image of $L^p(\mathbb{R}^n)$ under the operator $\mathcal{I}_s$. This point of view also allows us to define the fractional Sobolev spaces $W^{s,p}$ for $s \geq 0$ and $p = 1, \infty$.