Fourier Analysis

Javier Duoandikoetxea

Translated and revised by

David Cruz-Uribe, SFO

Graduate Studies
in Mathematics

Volume 29

American Mathematical Society
Providence, Rhode Island

Contents

Preface

Fourier Analysis is a large branch of mathematics whose point of departure is the study of Fourier series and integrals. However, it encompasses a variety of perspectives and techniques, and so many different introductions with that title are possible. The goal of this book is to study the real variable methods introduced into Fourier analysis by A. P. Calderón and A. Zygmund in the 1950's.

We begin in Chapter 1 with a review of Fourier series and integrals, and then in Chapters 2 and 3 we introduce two operators which are basic to the field: the Hardy-Littlewood maximal function and the Hilbert transform. Even though they appeared before the techniques of Calderón and Zygmund, we treat these operators from their point of view. The goal of these techniques is to enable the study of analogs of the Hilbert transform in higher dimensions; these are of great interest in applications. Such operators are known as singular integrals and are discussed in Chapters 4 and 5 along with their modern generalizations. We next consider two of the many contributions to the field which appeared in the 1970's. In Chapter 6 we study the relationship between H^1, BMO and singular integrals, and in Chapter 7 we present the elementary theory of weighted norm inequalities. In Chapter 8 we discuss Littlewood-Paley theory; its origins date back to the 1930's, but it has had extensive later development which includes a number of applications. Those presented in this chapter are useful in the study of Fourier multipliers, which also uses the theory of weighted inequalities. We end the book with an important result of the 80's, the so-called $T1$ theorem, which has been of crucial importance to the field.

At the end of each chapter there is a section in which we try to give some idea of further results which are not discussed in the text, and give references for the interested reader. A number of books and all the articles cited appear only in these notes; the bibliography at the end of the text is reserved for books which treat in depth the ideas we have presented.

The material in this book comes from a graduate course taught at the Universidad Autónoma de Madrid during the academic year 1988-89. Part of it is based on notes I took as a student in a course taught by José Luis Rubio de Francia at the same university in the fall of 1985. It seemed to have been his intention to write up his course, but he was prevented from doing so by his untimely death. Therefore, I have taken the liberty of using his ideas, which I learned both in his class and in many pleasant conversations in the hallway and at the blackboard, to write this book. Although it is dedicated to his memory, I almost regard it as a joint work. Also, I would like to thank my friends at the Universidad Autónoma de Madrid who encouraged me to teach this course and to write this book.

The book was first published in Spanish in the *Colección de Estudios* of the Universidad Autónoma de Madrid (1991), and then was republished with only some minor typographical corrections in a joint edition of Addison-Wesley/Universidad Autónoma de Madrid (1995). From the very beginning some colleagues suggested that there would be interest in an English translation which I never did. But when Professor David Cruz-Uribe offered to translate the book I immediately accepted. I realized at once that the text could not remain the same because some of the many developments of the last decade had to be included in the informative sections closing each chapter together with a few topics omitted from the first edition. As a consequence, although only minor changes have been introduced to the core of the book, the sections named "Notes and further results" have been considerably expanded to incorporate new topics, results and references.

The task of updating the book would have not been accomplished as it has been without the invaluable contribution of Professor Cruz-Uribe. Apart from reading the text, suggesting changes and clarifying obscure points, he did a great work on expanding the above mentioned notes, finding references and proposing new results to be included. The improvements of this book with respect to the original have certainly been the fruit of our joint work, and I am very grateful to him for sharing with me his knowledge of the subject much beyond the duties of a mere translator.

Javier Duoandikoetxea

Bilbao, June 2000

Chapter 1

Fourier Series and Integrals

1. Fourier coefficients and series

The problem of representing a function f, defined on (an interval of) $\mathbb{R}$, by a trigonometric series of the form

$$f(x) = \sum_{k=0}^{\infty} a_k \cos(kx) + b_k \sin(kx) \tag{1.1}$$

arises naturally when using the method of separation of variables to solve partial differential equations. This is how J. Fourier arrived at the problem, and he devoted the better part of his *Théorie Analytique de la Chaleur* (1822, results first presented to the Institute de France in 1807) to it. Even earlier, in the middle of the 18th century, Daniel Bernoulli had stated it while trying to solve the problem of a vibrating string, and the formula for the coefficients appeared in an article by L. Euler in 1777.

The right-hand side of (1.1) is a periodic function with period 2π, so f must also have this property. Therefore it will suffice to consider f on an interval of length 2π. Using Euler's identity, $e^{ikx} = \cos(kx) + i\sin(kx)$, we can replace the functions $\sin(kx)$ and $\cos(kx)$ in (1.1) by $\{e^{ikx} : k \in \mathbb{Z}\}$; we will do so from now on. Moreover, we will consider functions with period 1 instead of 2π, so we will modify the system of functions to $\{e^{2\pi ikx} : k \in \mathbb{Z}\}$. Our problem is thus transformed into studying the representation of f by

$$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx}. \tag{1.2}$$

If we assume, for example, that the series converges uniformly, then by multiplying by $e^{-2\pi imx}$ and integrating term-by-term on $(0,1)$ we get

$$c_m = \int_0^1 f(x)e^{-2\pi imx}\,dx$$

because of the orthogonality relationship

$$\int_0^1 e^{2\pi ikx}e^{-2\pi imx}\,dx = \begin{cases} 0 & \text{if } k \neq m \\ 1 & \text{if } k = m. \end{cases} \tag{1.3}$$

Denote the additive group of the reals modulo 1 (that is $\mathbb{R}/\mathbb{Z}$) by $\mathbb{T}$, the one-dimensional torus. This can also be identified with the unit circle, S^1. Saying that a function is defined on $\mathbb{T}$ is equivalent to saying that it is defined on $\mathbb{R}$ and has period 1. To each function $f \in L^1(\mathbb{T})$ we associate the sequence $\{\hat{f}(k)\}$ of Fourier coefficients of f, defined by

$$\hat{f}(k) = \int_0^1 f(x)e^{-2\pi ikx}\,dx. \tag{1.4}$$

The trigonometric series with these coefficients,

$$\sum_{k=-\infty}^{\infty} \hat{f}(k)e^{2\pi ikx}, \tag{1.5}$$

is called the Fourier series of f.

Our problem now consists in determining when and in what sense the series (1.5) represents the function f.

2. Criteria for pointwise convergence

Denote the N-th symmetric partial sum of the series (1.5) by $S_N f(x)$; that is,

$$S_N f(x) = \sum_{k=-N}^{N} \hat{f}(k)e^{2\pi ikx}.$$

Note that this is also the N-th partial sum of the series when it is written in the form of (1.1).

Our first approach to the problem of representing f by its Fourier series is to determine whether $\lim S_N f(x)$ exists for each x, and if so, whether it is equal to $f(x)$. The first positive result is due to P. G. L. Dirichlet (1829), who proved the following convergence criterion: if f is bounded, piecewise continuous, and has a finite number of maxima and minima, then $\lim S_N f(x)$ exists and is equal to $\frac{1}{2}[f(x+) + f(x-)]$. Jordan's criterion, which we prove below, includes this result as a special case.

In order to study $S_N f(x)$ we need a more manageable expression. Dirichlet wrote the partial sums as follows:

$$\begin{aligned} S_N f(x) &= \sum_{k=-N}^{N} \int_0^1 f(t) e^{-2\pi i k t}\, dt \cdot e^{2\pi i k x} \\ &= \int_0^1 f(t) D_N(x-t)\, dt \\ &= \int_0^1 f(x-t) D_N(t)\, dt, \end{aligned}$$

where D_N is the Dirichlet kernel,

$$D_N(t) = \sum_{k=-N}^{N} e^{2\pi i k t}.$$

If we sum this geometric series we get

$$D_N(t) = \frac{\sin(\pi(2N+1)t)}{\sin(\pi t)}. \tag{1.6}$$

This satisfies

$$\int_0^1 D_N(t)\, dt = 1 \quad \text{and} \quad |D_N(t)| \le \frac{1}{\sin(\pi\delta)}, \quad \delta \le |t| \le 1/2.$$

We will prove two criteria for pointwise convergence.

Theorem 1.1 (Dini's Criterion)**.** *If for some x there exists $\delta > 0$ such that*

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

then

$$\lim_{N\to\infty} S_N f(x) = f(x).$$

Theorem 1.2 (Jordan's Criterion)**.** *If f is a function of bounded variation in a neighborhood of x, then*

$$\lim_{N\to\infty} S_N f(x) = \frac{1}{2}[f(x+) + f(x-)].$$

At first it may seem surprising that these results are local, since if we modify the function slightly, the Fourier coefficients of f change. Nevertheless, the convergence of a Fourier series is effectively a local property, and if the modifications are made outside of a neighborhood of x, then the behavior of the series at x does not change. This is made precise by the following result.

Theorem 1.3 (Riemann Localization Principle). *If f is zero in a neighborhood of x, then*

$$\lim_{N\to\infty} S_N f(x) = 0.$$

An equivalent formulation of this result is to say that if two functions agree in a neighborhood of x, then their Fourier series behave in the same way at x.

From the definition of Fourier coefficients (1.4) it follows immediately that

$$|\hat{f}(k)| \le \|f\|_1,$$

but a sharper estimate is true which we will use to prove the preceding results.

Lemma 1.4 (Riemann-Lebesgue). *If $f \in L^1(\mathbb{T})$ then*

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

Proof. Since $e^{2\pi ix}$ has period 1,

$$\begin{aligned}\hat{f}(k) &= \int_0^1 f(x)e^{-2\pi ikx}\,dx\\ &= -\int_0^1 f(x)e^{-2\pi ik(x+1/2k)}\,dx\\ &= -\int_0^1 f(x-1/2k)e^{-2\pi ikx}\,dx.\end{aligned}$$

Hence,

$$\hat{f}(k) = \frac{1}{2}\int_0^1 [f(x) - f(x-1/2k)]e^{-2\pi ikx}\,dx.$$

If f is continuous, it follows immediately that

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

For arbitrary $f \in L^1(\mathbb{T})$, given $\epsilon > 0$, choose g continuous such that $\|f-g\|_1 < \epsilon/2$ and choose k sufficiently large that $|\hat{g}(k)| < \epsilon/2$. Then

$$|\hat{f}(k)| \le |(f-g)\widehat{\ }(k)| + |\hat{g}(k)| \le \|f-g\|_1 + |\hat{g}(k)| < \epsilon.$$

□

Proof of Theorem 1.3. Suppose that $f(t) = 0$ on $(x - \delta, x + \delta)$. Then

$$\begin{aligned} S_N f(x) &= \int_{\delta \le |t| < 1/2} f(x-t) \frac{\sin(\pi(2N+1)t)}{\sin(\pi t)}\, dt \\ &= (ge^{\pi i \cdot})\hat{}(N) + (ge^{-\pi i \cdot})\hat{}(-N), \end{aligned}$$

where

$$g(t) = \frac{f(x-t)}{2i \sin(\pi t)} \chi_{\{\delta \le |t| < 1/2\}}(t)$$

is integrable. By the Riemann-Lebesgue lemma we conclude that

$$\lim_{N \to \infty} S_N f(x) = 0.$$

□

Proof of Theorem 1.1. Since the integral of D_N equals 1,

$$\begin{aligned} S_N f(x) - f(x) &= \int_{-1/2}^{1/2} [f(x-t) - f(x)] \frac{\sin(\pi(2N+1)t)}{\sin(\pi t)}\, dt \\ &= \int_{|t|<\delta} + \int_{\delta \le |t| < 1/2}. \end{aligned}$$

By the Riemann-Lebesgue lemma both of these integrals tend to 0. The second if we argue as in the previous proof, the first since by hypothesis the function

$$\frac{f(x-t) - f(x)}{\sin(\pi t)} \chi_{\{|t|<\delta\}}(t)$$

is integrable. (Recall that if $|t| < \delta$, $\sin(\pi t)$ and πt are equivalent.) □

Proof of Theorem 1.2. Since every function of bounded variation is the difference of two monotonic functions, we may assume that f is monotonic in a neighborhood of x. Since

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

it will be enough to show that for g monotonic

$$\lim_{N \to \infty} \int_0^{1/2} g(t) D_N(t)\, dt = \frac{1}{2} g(0+).$$

Further, we may assume that $g(0+) = 0$ and that g is increasing to the right of 0. Given $\epsilon > 0$, choose $\delta > 0$ such that $g(t) < \epsilon$ if $0 < t < \delta$. Then

$$\int_0^{1/2} g(t) D_N(t)\, dt = \int_0^{\delta} + \int_{\delta}^{1/2}.$$

Again by the Riemann-Lebesgue lemma, the second integral tends to 0. We apply the second mean value theorem for integrals[1] to the first integral. Then for some ν, $0 < \nu < \delta$,

$$\int_0^\delta g(t) D_N(t)\, dt = g(\delta-) \int_\nu^\delta D_N(t)\, dt.$$

Furthermore,

$$\begin{aligned}\left|\int_\nu^\delta D_N(t)\, dt\right| &\le \left|\int_\nu^\delta \sin(\pi(2N+1)t)\left(\frac{1}{\sin(\pi t)} - \frac{1}{\pi t}\right) dt\right| \\ &\quad + \left|\int_\nu^\delta \frac{\sin(\pi(2N+1)t)}{\pi t}\, dt\right| \\ &\le \int_\nu^\delta \left|\frac{1}{\sin(\pi t)} - \frac{1}{\pi t}\right| dt + 2 \sup_{M>0} \left|\int_0^M \frac{\sin(\pi t)}{t}\, dt\right| \\ &\le C.\end{aligned}$$

Hence,

$$\left|\int_0^\delta g(t) D_N(t)\, dt\right| \le C\epsilon.$$

□

3. Fourier series of continuous functions

If f satisfies a Lipschitz-type condition in a neighborhood of x, that is, $|f(x+t) - f(x)| \le C|t|^a$ for some $a > 0$, $|t| < \delta$, then Dini's criterion applies to it. However, continuous functions need not satisfy this condition or any other convergence criterion we have seen. This must be the case because of the following result due to P. du Bois-Reymond (1873).

Theorem 1.5. *There exists a continuous function whose Fourier series diverges at a point.*

Du Bois-Reymond constructed a function with this property, but we will show that one exists by applying the uniform boundedness principle, also known as the Banach-Steinhaus theorem.

Lemma 1.6 (Uniform Boundedness Principle). *Let X be a Banach space, Y a normed vector space, and let $\{T_a\}_{a\in A}$ be a family of bounded linear*

[1] If ϕ is continuous and h monotonic on $[a, b]$, then there exists c, $a < c < b$, such that

$$\int_a^b h\phi = h(b-) \int_c^b \phi + h(a+) \int_a^c \phi.$$

operators from X to Y. Then either

$$\sup_a \|T_a\| < \infty$$

or there exists $x \in X$ such that

$$\sup_a \|T_a x\|_Y = \infty.$$

(Recall that the operator norm of T_a is $\|T_a\| = \sup\{\|T_a x\|_Y : \|x\|_X \leq 1\}$.) A proof of this result can be found, for example, in Rudin [**14**, Chapter 5].

Now let $X = C(\mathbb{T})$ with the norm $\|\cdot\|_\infty$ and let $Y = \mathbb{C}$. Define $T_N : X \to Y$ by

$$T_N f = S_N f(0) = \int_{-1/2}^{1/2} f(t) D_N(t)\, dt.$$

Define the Lebesgue numbers L_N by

$$L_N = \int_{-1/2}^{1/2} |D_N(t)|\, dt;$$

it is immediate that $|T_N f| \leq L_N \|f\|_\infty$. $D_N(t)$ has a finite number of zeros so $\operatorname{sgn} D_N(t)$ has a finite number of jump discontinuities. Therefore, by modifying it on a small neighborhood of each discontinuity, we can form a continuous function f such that $\|f\|_\infty = 1$ and $|T_N f| \geq L_N - \epsilon$. Hence, $\|T_N\| = L_N$. Thus if we can prove that $L_N \to \infty$ as $N \to \infty$, then by the uniform boundedness principle there exists a continuous function f such that

$$\limsup_{N\to\infty} |S_N f(0)| = \infty;$$

that is, the Fourier series of f diverges at 0.

Lemma 1.7. $L_N = \dfrac{4}{\pi^2} \log N + O(1)$.

Proof.

$$\begin{aligned} L_N &= 2\int_0^{1/2} \left| \frac{\sin(\pi(2N+1)t)}{\pi t} \right| dt + O(1) \\ &= 2\int_0^{N+1/2} \left| \frac{\sin(\pi t)}{\pi t} \right| dt + O(1) \\ &= 2\sum_{k=0}^{N-1} \int_k^{k+1} \left| \frac{\sin(\pi t)}{\pi t} \right| dt + O(1) \\ &= \frac{2}{\pi} \sum_{k=0}^{N-1} \int_0^1 \frac{|\sin(\pi t)|}{t+k}\, dt + O(1) \end{aligned}$$

$$= \frac{2}{\pi}\int_0^1 |\sin(\pi t)| \sum_{k=1}^{N-1} \frac{1}{t+k}\,dt + O(1)$$
$$= \frac{4}{\pi^2}\log N + O(1).$$

□

4. Convergence in norm

The development of measure theory and L^p spaces led to a new approach to the problem of convergence. We can now ask:

(1) Does $\lim_{N\to\infty} \|S_N f - f\|_p = 0$ for $f \in L^p(\mathbb{T})$?

(2) Does $\lim_{N\to\infty} S_N f(x) = f(x)$ almost everywhere if $f \in L^p(\mathbb{T})$?

We can restate the first question by means the following lemma.

Lemma 1.8. *$S_N f$ converges to f in L^p norm, $1 \le p < \infty$, if and only if there exists C_p independent of N such that*

$$\|S_N f\|_p \le C_p \|f\|_p. \tag{1.7}$$

Proof. The necessity of (1.7) follows from the uniform boundedness principle.

To see that it is sufficient, first note that if g is a trigonometric polynomial, then $S_N g = g$ for $N \ge \deg g$. Therefore, since the trigonometric polynomials are dense in L^p (see Corollary 1.11), if $f \in L^p$ we can find a trigonometric polynomial g such that $\|f - g\|_p < \epsilon$, and so for N sufficiently large

$$\|S_N f - f\|_p \le \|S_N(f-g)\|_p + \|S_N g - g\|_p + \|f-g\|_p \le (C_p + 1)\epsilon.$$

□

If $1 < p < \infty$, then inequality (1.7) holds, as we will show in Chapter 3. When $p = 1$, the L^1 operator norm of S_N is again L_N, and so by Lemma 1.7 the answer to the first question is no.

When $p = 2$, the functions $\{e^{2\pi i k x}\}$ form an orthonormal system (by (1.3)) which is complete (i.e. an orthonormal basis) by the density of the trigonometric polynomials in L^2. Therefore, we can apply the theory of Hilbert spaces to get the following.

Theorem 1.9. *The mapping* $f \mapsto \{\hat{f}(k)\}$ *is an isometry from* L^2 *to* ℓ^2*, that is,*

$$\|f\|_2^2 = \sum_{k=-\infty}^{\infty} |\hat{f}(k)|^2.$$

Convergence in norm in L^2 follows from this immediately.

The second question is much more difficult. A. Kolmogorov (1926) gave an example of an integrable function whose Fourier series diverges at every point, so the answer is no if $p = 1$. If $f \in L^p$, $1 < p < \infty$, then the Fourier series of f converges almost everywhere. This was shown by L. Carleson (1965, $p = 2$) and R. Hunt (1967, $p > 1$). Until the result by Carleson, the answer was unknown even for f continuous.

5. Summability methods

In order to recover a function f from its Fourier coefficients it would be convenient to find some other method than taking the limit of the partial sums of its Fourier series since, as we have seen, this approach does not always work well.

One such method, Cesàro summability, consists in taking the limit of the arithmetic means of the partial sums. As is well known, if $\lim a_k$ exists then

$$\lim_{k\to\infty} \frac{a_1 + \cdots + a_k}{k}$$

also exists and has the same value.

Define

$$\begin{aligned}\sigma_N f(x) &= \frac{1}{N+1}\sum_{k=0}^{N} S_k f(x)\\ &= \int_0^1 f(t)\frac{1}{N+1}\sum_{k=0}^{N} D_k(x-t)\,dt\\ &= \int_0^1 f(t)F_N(x-t)\,dt,\end{aligned}$$

where F_N is the Fejér kernel,

$$F_N(t) = \frac{1}{N+1}\sum_{k=0}^{N} D_k(t) = \frac{1}{N+1}\left(\frac{\sin(\pi(N+1)t)}{\sin(\pi t)}\right)^2.$$

F_N has the following properties:

$$F_N(t) \geq 0,$$

$$(1.8)\qquad \|F_N\|_1 = \int_0^1 F_N(t)\,dt = 1,$$

$$\lim_{N\to\infty} \int_{\delta<|t|<1/2} F_N(t)\,dt = 0 \quad \text{if } \delta > 0.$$

Because F_N is positive, its L^1 norm coincides with its integral and is 1. This is not the case for the Dirichlet kernel: its integral equals 1 because of cancellation between its positive and negative parts while its L^1 norm tends to infinity with N.

Theorem 1.10. *If $f \in L^p$, $1 \le p < \infty$, or if f is continuous and $p = \infty$, then*

$$\lim_{N\to\infty} \|\sigma_N f - f\|_p = 0.$$

Proof. Since $\int F_N = 1$, by Minkowski's inequality we have that

$$\begin{aligned} \|\sigma_N f - f\|_p &= \int_{-1/2}^{1/2} \|f(\cdot - t) - f(\cdot)\|_p F_N(t)\,dt \\ &\le \int_{|t|<\delta} \|f(\cdot - t) - f(\cdot)\|_p F_N(t)\,dt + 2\|f\|_p \int_{\delta<|t|<1/2} F_N(t)\,dt. \end{aligned}$$

Since for $1 \le p < \infty$,

$$\lim_{t\to 0} \|f(\cdot - t) - f(\cdot)\|_p = 0,$$

and the same limit holds if $p = \infty$ and f is continuous, the first term can be made as small as desired by choosing a suitable δ. And for fixed δ, by (1.8) the second term tends to 0. □

Corollary 1.11.

(1) *The trigonometric polynomials are dense in L^p, $1 \le p < \infty$.*

(2) *If f is integrable and $\hat{f}(k) = 0$ for all k, then f is identically zero.*

A second summability method is gotten by treating a Fourier series as the formal limit on the unit circle (in the complex plane) of

$$(1.9)\qquad u(z) = \sum_{k=0}^{\infty} \hat{f}(k) z^k + \sum_{k=-\infty}^{-1} \hat{f}(k) \bar{z}^{|k|}, \quad z = re^{2\pi i\theta}.$$

Since $\{\hat{f}(k)\}$ is a bounded sequence, this function is well defined on $|z| < 1$. It can be rewritten as

$$u(re^{2\pi i\theta}) = \sum_{k=-\infty}^{\infty} \hat{f}(k) r^{|k|} e^{2\pi i k\theta} = \int_{-1/2}^{1/2} f(t) P_r(\theta - t)\,dt,$$

where

$$P_r(t) = \sum_{k=-\infty}^{\infty} r^{|k|} e^{2\pi i k t} = \frac{1 - r^2}{1 - 2r\cos(2\pi t) + r^2}$$

is the Poisson kernel. The Poisson kernel has properties analogous to those of the Fejér kernel:

$$P_r(t) \geq 0,$$

$$\int_0^1 P_r(t)\,dt = 1, \tag{1.10}$$

$$\lim_{r\to 1^-} \int_{\delta<|t|<1/2} P_r(t)\,dt = 0 \quad \text{if } \delta > 0.$$

Therefore, we can prove a result analogous to Theorem 1.10.

Theorem 1.12. *If $f \in L^p$, $1 \leq p < \infty$, or if f is continuous and $p = \infty$, then*

$$\lim_{r\to 1^-} \|P_r * f - f\|_p = 0.$$

Since the function u is harmonic on $|z| < 1$, it is the solution to the Dirichlet problem:

$$\Delta u = 0 \quad \text{if } |z| < 1,$$
$$u = f \quad \text{if } |z| = 1,$$

where the boundary condition is interpreted in terms of Theorem 1.12.

In Chapter 2 we will study the almost everywhere convergence of $\sigma_N f(x)$ and $P_r * f(x)$.

6. The Fourier transform of L^1 functions

Given a function $f \in L^1(\mathbb{R}^n)$, define its Fourier transform by

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i x\cdot\xi}\,dx, \tag{1.11}$$

where $x \cdot \xi = x_1\xi_1 + x_2\xi_2 + \cdots + x_n\xi_n$. The following is a list of properties of the Fourier transform:

(1.12) $(\alpha f + \beta g)\widehat{\ } = \alpha\hat{f} + \beta\hat{g}$ (linearity);

(1.13) $\|\hat{f}\|_\infty \leq \|f\|_1$ and $\hat{f}$ is continuous;

(1.14) $\lim_{|\xi|\to\infty} \hat{f}(\xi) = 0$ (Riemann-Lebesgue);

(1.15) $(f * g)\widehat{\ } = \hat{f}\hat{g}$;

(1.16) $(\tau_h f)\widehat{\ }(\xi) = \hat{f}(\xi)e^{2\pi i h\cdot\xi}$, where $\tau_h f(x) = f(x+h)$;

$(fe^{2\pi i h\cdot x})\widehat{\ }(\xi) = \hat{f}(\xi - h)$;

(1.17) if $\rho \in O_n$ (an orthogonal transformation), then

$(f(\rho\cdot))\widehat{\ }(\xi) = \hat{f}(\rho\xi)$;

(1.18) if $g(x) = \lambda^{-n} f(\lambda^{-1}x)$, then $\hat{g}(\xi) = \hat{f}(\lambda\xi)$;

(1.19) $\left(\dfrac{\partial f}{\partial x_j}\right)^{\widehat{\ }}(\xi) = 2\pi i \xi_j \hat{f}(\xi)$;

(1.20) $(-2\pi i x_j f)\widehat{\ }(\xi) = \dfrac{\partial \hat{f}}{\partial \xi_j}(\xi)$.

The continuity of $\hat{f}$ follows from the dominated convergence theorem; (1.14) can be proved like Lemma 1.4; the rest follow from a change of variables, Fubini's theorem and integration by parts. In (1.19) we assume that $\partial f/\partial x_j \in L^1$ and in (1.20) that $x_j f \in L^1$.

Unlike on the torus, $L^1(\mathbb{R}^n)$ does not contain $L^p(\mathbb{R}^n)$, $p > 1$, so (1.11) does not define the Fourier transform of functions in those spaces. For the same reason, the formula which should allow us to recover f from $\hat{f}$,

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

may not make sense since (1.13) and (1.14) are all that we know about $\hat{f}$, and they do not imply that $\hat{f}$ is integrable. (In fact, $\hat{f}$ is generally not integrable.)

7. The Schwartz class and tempered distributions

A function f is in the Schwartz class, $\mathcal{S}(\mathbb{R}^n)$, if it is infinitely differentiable and if all of its derivatives decrease rapidly at infinity; that is, if for all $\alpha, \beta \in \mathbb{N}^n$,

$$\sup_x |x^\alpha D^\beta f(x)| = p_{\alpha,\beta}(f) < \infty.$$

Functions in C_c^∞ are in $\mathcal{S}$, but so are functions like $e^{-|x|^2}$ which do not have compact support. The collection $\{p_{\alpha,\beta}\}$ is a countable family of seminorms on $\mathcal{S}$, and we can use it to define a topology on $\mathcal{S}$: a sequence $\{\phi_k\}$ converges to 0 if and only if for all $\alpha, \beta \in \mathbb{N}^n$,

$$\lim_{k\to\infty} p_{\alpha,\beta}(\phi_k) = 0.$$

With this topology $\mathcal{S}$ is a Fréchet space (complete and metrizable) and is dense in $L^p(\mathbb{R}^n)$, $1 \le p < \infty$. In particular, $\mathcal{S} \subset L^1$ and (1.11) defines the Fourier transform of a function in $\mathcal{S}$.

The space of bounded linear functionals on $\mathcal{S}$, $\mathcal{S}'$, is called the space of tempered distributions. A linear map T from $\mathcal{S}$ to $\mathbb{C}$ is in $\mathcal{S}'$ if

$$\lim_{k\to\infty} T(\phi_k) = 0 \quad \text{whenever} \quad \lim_{k\to\infty} \phi_k = 0 \quad \text{in } \mathcal{S}.$$

Theorem 1.13. *The Fourier transform is a continuous map from $\mathcal{S}$ to $\mathcal{S}$ such that*

$$\int_{\mathbb{R}^n} f\hat{g} = \int_{\mathbb{R}^n} \hat{f} g \tag{1.21}$$

and

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi. \tag{1.22}$$

Equality (1.22) is referred to as the inversion formula.

To prove Theorem 1.13 we need to compute the Fourier transform of a particular function.

Lemma 1.14. *If $f(x) = e^{-\pi|x|^2}$ then $\hat{f}(\xi) = e^{-\pi|\xi|^2}$.*

Proof. We could prove this result directly by integrating in $\mathbb{C}$, but we will give a different proof here. It is enough to prove this in one dimension, since in $\mathbb{R}^n$ $\hat{f}$ is the product of n identical integrals.

The function $f(x) = e^{-\pi x^2}$ is the solution of the differential equation

$$u' + 2\pi x u = 0,$$
$$u(0) = 1.$$

By (1.19) and (1.20) we see that $\hat{u}$ satisfies the same differential equation with the initial value

$$\hat{u}(0) = \int_{\mathbb{R}} u(x)\, dx = \int_{\mathbb{R}} e^{-\pi x^2}\, dx = 1.$$

Therefore, by uniqueness, $\hat{f} = f$. □

Proof of Theorem 1.13. By (1.19) and (1.20) we have

$$\xi^\alpha D^\beta \hat{f}(\xi) = C(D^\alpha x^\beta f)\widehat{\ }(\xi),$$

so

$$|\xi^\alpha D^\beta \hat{f}(\xi)| \le C\|D^\alpha x^\beta f\|_1.$$

The L^1 norm can be bounded by a finite linear combination of seminorms of f, which implies that the Fourier transform is a continuous map from $\mathcal{S}$ to itself.

Equality (1.21) is an immediate consequence of Fubini's theorem since $f(x)g(y)$ is integrable on $\mathbb{R}^n \times \mathbb{R}^n$.

From (1.18) and (1.21) we get

$$\int f(x)\hat{g}(\lambda x)\,dx = \int \hat{f}(x)\lambda^{-n}g(\lambda^{-1}x)\,dx.$$

If we make the change of variables $\lambda x = y$ in the first integral, this becomes

$$\int f(\lambda^{-1}x)\hat{g}(x)\,dx = \int \hat{f}(x)g(\lambda^{-1}x)\,dx;$$

if we then take the limit as $\lambda \to \infty$, we get

$$f(0)\int \hat{g}(x)\,dx = g(0)\int \hat{f}(x)\,dx.$$

Let $g(x) = e^{-\pi|x|^2}$; then by Lemma 1.14,

$$f(0) = \int \hat{f}(\xi)\,d\xi,$$

which is (1.22) for $x = 0$. If we replace f by $\tau_x f$, then by (1.16),

$$f(x) = (\tau_x f)(0) = \int (\tau_x f)\widehat{\ }(\xi)\,d\xi = \int \hat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi.$$

□

If we let $\tilde{f}(x) = f(-x)$, we get the following corollary.

Corollary 1.15. *For $f \in \mathcal{S}$, $(\hat{f})\widehat{\ } = \tilde{f}$, and so the Fourier transform has period 4 (i.e. if we apply it four times, we get the identity operator).*

Definition 1.16. The Fourier transform of $T \in \mathcal{S}'$ is the tempered distribution $\hat{T}$ given by

$$\hat{T}(f) = T(\hat{f}), \qquad f \in \mathcal{S}.$$

By Theorem 1.13, $\hat{T}$ is a tempered distribution, and in particular, if T is an integrable function, then $\hat{T}$ coincides with the Fourier transform defined by equation (1.11). Likewise, if μ is a finite Borel measure (i.e. a bounded linear functional on $C_0(\mathbb{R}^n)$, the space of continuous functions which vanish at infinity), then $\hat{\mu}$ is the bounded continuous function given by

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

For δ, the Dirac measure at the origin, this gives us $\hat{\delta} = 1$.

Theorem 1.17. *The Fourier transform is a bounded linear bijection from $\mathcal{S}'$ to $\mathcal{S}'$ whose inverse is also bounded.*

Proof. If $T_n \to T$ in $\mathcal{S}'$, then for any $f \in \mathcal{S}$,

$$\hat{T}_n(f) = T_n(\hat{f}) \to T(\hat{f}) = \hat{T}(f).$$

Furthermore, the Fourier transform has period 4, so its inverse is equivalent to applying it 3 times; therefore, its inverse is also continuous. □

If we define $\tilde{T}$ by $\tilde{T}(f) = T(\tilde{f})$, then it follows from Corollary 1.15 that $(\tilde{\hat{T}})\hat{} = T$. And if $\hat{T} \in L^1$ then by the inversion formula we get that

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

in particular, T is a bounded, continuous function.

8. The Fourier transform on L^p, $1 < p \leq 2$

If $f \in L^p$, $1 \leq p \leq \infty$, then f can be identified with a tempered distribution: for $\phi \in \mathcal{S}$ define

$$T_f(\phi) = \int_{\mathbb{R}^n} f\phi.$$

Clearly this integral is finite. To see that T_f is continuous, suppose that $\phi_k \to 0$ in $\mathcal{S}$ as $k \to \infty$. Then by Hölder's inequality,

$$|T_f(\phi_k)| \leq \|f\|_p \|\phi_k\|_{p'}.$$

Then $\|\phi_k\|_{p'}$ is dominated by the L^∞ norm of functions of the form $x^a\phi_k$, and so by a finite linear combination of seminorms of ϕ_k; hence, the left-hand side tends to 0 as $k \to \infty$.

Moreover, when $1 \leq p \leq 2$ we have that $\hat{f}$ is a function.

Theorem 1.18. *The Fourier transform is an isometry on L^2; that is, $\hat{f} \in L^2$ and $\|\hat{f}\|_2 = \|f\|_2$. Furthermore,*

$$\hat{f}(\xi) = \lim_{R\to\infty} \int_{|x|<R} f(x) e^{-2\pi i x\cdot\xi}\, dx$$

and

$$f(x) = \lim_{R\to\infty} \int_{|\xi|<R} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi,$$

where the limits are in L^2.

The identity $\|\hat{f}\|_2 = \|f\|_2$ is referred to as the Plancherel theorem.

Proof. Given $f, h \in \mathcal{S}$, let $g = \bar{\hat{h}}$, so that $\hat{g} = \bar{h}$. Then by (1.21) we have that

$$\int_{\mathbb{R}^n} f\bar{h} = \int_{\mathbb{R}^n} \hat{f}\bar{\hat{h}}. \tag{1.23}$$

If we let $h = f$ then we get $\|f\|_2 = \|\hat{f}\|_2$ for $f \in \mathcal{S}$. Since $\mathcal{S}$ is dense in L^2, the Fourier transform extends to all f in L^2 with equality of norms.

Finally, the continuity of the Fourier transform implies the given formulas for f and $\hat{f}$ as limits in L^2, since $f\chi_{B(0,R)}$ and $\hat{f}\chi_{B(0,R)}$ converge to f and $\hat{f}$ in L^2. □

If $f \in L^p$, $1 < p < 2$, then it can be decomposed as $f = f_1 + f_2$, where $f_1 \in L^1$ and $f_2 \in L^2$. (For example, let $f_1 = f\chi_{\{x:|f(x)|>1\}}$ and $f_2 = f - f_1$.) Therefore, $\hat{f} = \hat{f}_1 + \hat{f}_2 \in L^\infty + L^2$. However, by applying an interpolation theorem we can get a sharper result.

Theorem 1.19 (Riesz-Thorin Interpolation)**.** *Let* $1 \le p_0, p_1, q_0, q_1 \le \infty$, *and for* $0 < \theta < 1$ *define* p *and* q *by*

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

If T *is a linear operator from* $L^{p_0} + L^{p_1}$ *to* $L^{q_0} + L^{q_1}$ *such that*

$$\|Tf\|_{q_0} \le M_0\|f\|_{p_0} \quad \text{for } f \in L^{p_0}$$

and

$$\|Tf\|_{q_1} \le M_1\|f\|_{p_1} \quad \text{for } f \in L^{p_1},$$

then

$$\|Tf\|_q \le M_0^{1-\theta}M_1^{\theta}\|f\|_p \quad \text{for } f \in L^p.$$

The proof of this result uses the so-called "three-lines" theorem for analytic functions; it can be found, for example, in Stein and Weiss [**18**, Chapter 5] or Katznelson [**10**, Chapter 4].

Corollary 1.20 (Hausdorff-Young Inequality)**.** *If* $f \in L^p$, $1 \le p \le 2$, *then* $\hat{f} \in L^{p'}$ *and*

$$\|\hat{f}\|_{p'} \le \|f\|_p.$$

Proof. Apply Theorem 1.19 using inequality (1.13), $\|\hat{f}\|_\infty \le \|f\|_1$, and the Plancherel theorem, $\|\hat{f}\|_2 = \|f\|_2$. □

We digress to give another corollary of Riesz-Thorin interpolation which is not directly related to the Fourier transform but which will be useful in later chapters.

Corollary 1.21 (Young's Inequality). *If $f \in L^p$ and $g \in L^q$, then $f * g \in L^r$, where $1/r + 1 = 1/p + 1/q$, and*

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

Proof. If we fix $f \in L^p$ we immediately get the inequalities

$$\|f * g\|_p \leq \|f\|_p \|g\|_1$$

and

$$\|f * g\|_\infty \leq \|f\|_p \|g\|_{p'}.$$

The desired result follows by Riesz-Thorin interpolation. □

9. The convergence and summability of Fourier integrals

The problem of recovering a function from its Fourier transform is similar to the same problem for Fourier series. We need to determine if and when

$$\lim_{R\to\infty} \int_{B_R} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi = f(x),$$

where $B_R = \{Rx : x \in B\}$, B is an open convex neighborhood of the origin, and the limit is understood either as in L^p or as pointwise almost everywhere. If we define the partial sum operator S_R by

$$(S_R f)\widehat{\ } = \chi_{B_R} \hat{f},$$

then this problem is equivalent to determining if

$$\lim_{R\to\infty} S_R f = f.$$

Analogous to Lemma 1.8, a necessary and sufficient condition for convergence in norm is that

$$\|S_R f\|_p \leq C_p \|f\|_p,$$

where C_p is independent of R. When $n = 1$ this is the case; we will prove this in Chapter 3. We will also prove several partial results when $n > 1$, but in general there is no convergence in norm when $p \neq 2$. We will discuss this in Chapter 8.

In the case $n = 1$, if $B = (-1, 1)$ then

$$S_R f(x) = D_R * f(x),$$

where D_R is the Dirichlet kernel,

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

This is clearly not integrable, but it is in $L^q(\mathbb{R})$ for any $q > 1$, so $D_R * f$ is well defined if $f \in L^p$, $1 < p < \infty$.

Almost everywhere convergence depends on the bound

$$\| \sup_R |S_R f| \|_p \leq C_p \|f\|_p.$$

This holds if $1 < p < \infty$ (the Carleson-Hunt theorem) but we cannot prove it here.

For the Fourier transform, the method of Cesàro summability consists in taking integral averages of the partial sum operators,

$$\sigma_R f(x) = \frac{1}{R}\int_0^R S_t f(x)\, dt,$$

and determining if $\lim \sigma_R f(x) = f(x)$. When $n = 1$ and $B = (-1, 1)$,

$$\sigma_R f(x) = F_R * f(x),$$

where F_R is the Fejér kernel,

$$F_R(x) = \frac{1}{R}\int_0^R D_t(x)\, dt = \frac{\sin^2(\pi R x)}{R(\pi x)^2}. \tag{1.24}$$

Unlike the Dirichlet kernel, the Fejér kernel is integrable. Since it has properties analogous to (1.8), one can prove that in L^p, $1 \leq p < \infty$,

$$\lim_{R\to\infty} \sigma_R f = f.$$

The proof is similar to that of Theorem 1.10. In Chapter 2 we will prove two general results from which we can deduce convergence in L^p and pointwise almost everywhere for this and the following summability methods.

The method of Abel-Poisson summability consists in introducing the factor $e^{-2\pi t|\xi|}$ into the inversion formula. Then for any $t > 0$ the integral converges, and we take the limit as t tends to 0. If we instead introduce the factor $e^{-\pi t^2|\xi|^2}$, we get the method of Gauss-Weierstrass summability. More precisely, we define the functions

$$u(x,t) = \int_{\mathbb{R}^n} e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi, \tag{1.25}$$

$$w(x,t) = \int_{\mathbb{R}^n} e^{-\pi t^2|\xi|^2} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi, \tag{1.26}$$

and then try to determine if

$$\lim_{t\to 0^+} u(x,t) = f(x), \tag{1.27}$$

$$\lim_{t\to 0^+} w(x,t) = f(x) \tag{1.28}$$

in L^p or pointwise almost everywhere.

One can show that $u(x,t)$ is harmonic in the half-space $\mathbb{R}^{n+1}_+ = \mathbb{R}^n \times (0,\infty)$. When $n=1$ we have an equivalent formula analogous to (1.9):

$$u(z) = \int_0^\infty \hat{f}(\xi)e^{2\pi i z\xi}\,d\xi + \int_{-\infty}^0 \hat{f}(\xi)e^{2\pi i\bar{z}\xi}\,d\xi, \quad z = x+it, \tag{1.29}$$

which immediately implies that u is harmonic. The limit (1.27) can be interpreted as the boundary condition of the Dirichlet problem,

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

It follows from (1.25) that

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

where $\hat{P}_t(\xi) = e^{-2\pi t|\xi|}$. One can prove by a simple calculation if $n=1$, and a more difficult one when $n>1$ (see Stein and Weiss [**18**, p. 6]), that

$$P_t(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}} \frac{t}{(t^2+|x|^2)^{\frac{n+1}{2}}}. \tag{1.30}$$

This is called the Poisson kernel.

In the case of Gauss-Weierstrass summability, one can show that the function $\tilde{w}(x,t) = w(x,\sqrt{4\pi t})$ is the solution of the heat equation

$$\frac{\partial\tilde{w}}{\partial t} - \Delta\tilde{w} = 0 \quad \text{on } \mathbb{R}^{n+1}_+,$$
$$\tilde{w}(x,0) = f(x) \quad x \in \mathbb{R}^n,$$

and (1.28) can be interpreted as the initial condition for the problem. We also have the formula

$$w(x,t) = W_t * f(x),$$

where W_t is the Gauss-Weierstrass kernel,

$$W_t(x) = t^{-n}e^{-\pi|x|^2/t^2}. \tag{1.31}$$

This formula can be proved using Lemma 1.14 and (1.18).

10. Notes and further results

10.1. References.

The classic reference on trigonometric series is the book by Zygmund [**21**], which will also be a useful reference for results in the next few chapters. However, this work can be difficult to consult at times. Another comprehensive reference on trigonometric series is the book by Bary [**1**].

There are excellent discussions of Fourier series and integrals in Katznelson [**10**] and Dym and McKean [**4**]. The book by R. E. Edwards [**5**] is an exhaustive study of Fourier series from a more modern perspective. The article by Weiss [**20**] and the book by Körner [**12**] are also recommended. An excellent historical account by J. P. Kahane on Fourier series and their influence on the development of mathematical concepts is found in the first half of [**9**]. The book *Fourier Analysis and Boundary Value Problems*, by E. González-Velasco (Academic Press, New York, 1995), contains many applications of Fourier's method of separation of variables to partial differential equations and also contains historical information. (Also see by the same author, *Connections in mathematical analysis: the case of Fourier series*, Amer. Math. Monthly **99** (1992), 427–441.) The book by O. G. Jørsboe and L. Melbro (*The Carleson-Hunt Theorem on Fourier Series*, Lecture Notes in Math. **911**, Springer-Verlag, Berlin, 1982) is devoted to the proof of this theorem. The original references for this are the articles by L. Carleson (*On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157) and R. Hunt (*On the convergence of Fourier series*, Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967), pp. 235–255, Southern Illinois Univ. Press, Carbondale, 1968). Kolmogorov's example of an L^1 function whose Fourier series diverges everywhere appeared in *Une série de Fourier-Lebesgue divergente partout* (C. R. Acad. Sci. Paris **183** (1926), 1327–1328).

10.2. Multiple Fourier series.

Let $\mathbb{T}^n$ be the n-dimensional torus (which we can identify with the quotient group $\mathbb{R}^n/\mathbb{Z}^n$). A function defined on $\mathbb{T}^n$ is equivalent to a function on $\mathbb{R}^n$ which has period 1 in each variable. If $f \in L^1(\mathbb{R}^n)$ then we can define its Fourier coefficients by

$$\hat{f}(\nu) = \int f(x) e^{-2\pi i x\cdot\nu}\, dx, \quad \nu \in \mathbb{Z}^n,$$

and construct the Fourier series of f with these coefficients,

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

One can prove several results similar to those for Fourier series in one variable, but one needs increasingly restrictive regularity conditions as n increases. See Stein and Weiss [**18**, Chapter 7].

10.3. The Poisson summation formula.

Let f be a function such that for some $\delta > 0$,

$$|f(x)| \le A(1+|x|)^{-n-\delta} \quad \text{and} \quad |\hat{f}(\xi)| \le A(1+|\xi|)^{-n-\delta}.$$

(In particular, f and $\hat{f}$ are both continuous.) Then

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

This equality (or more precisely, the case when $x = 0$) is known as the Poisson summation formula and is nothing more than the inversion formula. The left-hand side defines a function on $\mathbb{T}^n$ whose Fourier coefficients are precisely $\hat{f}(\nu)$.

10.4. Gibbs phenomenon.

Let $f(x) = \operatorname{sgn}(x)$ on $(-1/2, 1/2)$. By Dirichlet's criterion, for example, we know that $S_N f(x)$ converges to $f(x)$ for all x. To the right of 0 the partial sums oscillate around 1 but, contrary to what one might expect, the amount by which they overstep 1 does not tend to 0 as N increases. One can show that

$$\lim_{N\to\infty} \sup_x S_N f(x) = \frac{2}{\pi} \int_0^\pi \frac{\sin(y)}{y}\, dy \approx 1.17898\ldots\ .$$

This phenomenon occurs whenever a function has a jump discontinuity. It is named after J. Gibbs, who announced it in Nature **59** (1899), although it had already been discovered by H. Wilbraham in 1848. See Dym and McKean [**4**, Chapter 1] and the paper by E. Hewitt and R. E. Hewitt (*The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis*, Arch. Hist. Exact Sci. **21** (1979/80), 129–160).

Gibbs phenomenon is eliminated by replacing pointwise convergence by Cesàro summability. For if $m \le f(x) \le M$, then by the first two properties of Féjer kernels in (1.8), $m \le \sigma_N f(x) \le M$. In fact, it can be shown that if $m \le f(x) \le M$ on an interval (a, b), then for any $\epsilon > 0$, $m - \epsilon \le \sigma_N f(x) \le M + \epsilon$ on $(a + \epsilon, b - \epsilon)$ for N sufficiently large.

10.5. The Hausdorff-Young inequality.

Corollary 1.20 was gotten by an immediate application of Riesz-Thorin interpolation. But in fact a stronger inequality is true: if $1 \le p \le 2$ then

$$\|\hat{f}\|_{p'} \le \left(\frac{p^{1/p}}{(p')^{1/p'}}\right)^{n/2} \|f\|_p.$$

This inequality is sharp since equality holds for $f(x) = e^{-\pi|x|^2}$. This result was proved by W. Beckner (*Inequalities in Fourier analysis*, Ann. of Math. **102** (1975), 159–182); the special case when p is even was proved earlier by K. I. Babenko (*An inequality in the theory of Fourier integrals* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **25** (1961), 531–542).

In the same article, Beckner also proved a sharp version of Young's inequality (Corollary 1.21).

10.6. Eigenfunctions for the Fourier transform in $L^2(\mathbb{R})$.

Since the Fourier transform has period 4, if f is a function such that $\hat{f} = \lambda f$, we must have that $\lambda^4 = 1$. Hence, $\lambda = \pm 1, \pm i$ are the only possible eigenvalues of the Fourier transform. Lemma 1.14 shows that $\exp(-\pi x^2)$ is an eigenfunction associated with the eigenvalue 1. The Hermite functions give the remaining eigenfunctions: for $n \geq 0$,

$$h_n(x) = \frac{(-1)^n}{n!} \exp(\pi x^2) \frac{d^n}{dx^n} \exp(-\pi x^2)$$

satisfies $\hat{h}_n = (-i)^n h_n$. If we normalize these functions,

$$e_n = \frac{h_n}{\|h_n\|_2} = [(4\pi)^{-n}\sqrt{2}n!]^{1/2} h_n,$$

we get an orthonormal basis of $L^2(\mathbb{R})$ such that

$$\hat{f} = \sum_{n=0}^{\infty} (-i)^n \langle f, e_n \rangle e_n.$$

Thus $L^2(\mathbb{R})$ decomposes into the direct sum $H_0 \oplus H_1 \oplus H_2 \oplus H_3$, where on the subspace H_j, $0 \leq j \leq 3$, the Fourier transform acts by multiplying functions by i^j.

This approach to defining the Fourier transform in $L^2(\mathbb{R})$ is due to N. Wiener and can be found in his book (*The Fourier Integral and Certain of its Applications*, original edition, 1933; Cambridge Univ. Press, Cambridge, 1988). Also see Dym and McKean [**4**, Chapter 2].

In higher dimensions, the eigenfunctions of the Fourier transform are products of Hermite functions, one in each coordinate variable. Also see Chapter 4, Section 7.2.

10.7. Interpolation of analytic families of operators.

The Riesz-Thorin interpolation theorem has a powerful generalization due to E. M. Stein. (See Stein and Weiss [**18**, Chapter 5].) Let $S = \{z \in \mathbb{C} : 0 \leq \operatorname{Re} z \leq 1\}$ and let $\{T_z\}_{z\in S}$ be a family of operators. This family is said to be admissible if given two functions $f, g \in L^1(\mathbb{R}^n)$, the mapping

$$z \mapsto \int_{\mathbb{R}^n} T_z(f) g \, dx$$

is analytic on the interior of S and continuous on the boundary, and if there exists a constant $a < \pi$ such that

$$e^{-a|\operatorname{Im} z|} \log \left| \int_{\mathbb{R}^n} T_z(f) g \, dx \right|$$

is uniformly bounded for all $z \in S$.

Theorem 1.22. *Let $\{T_z\}$ be an admissible family of operators, and suppose that for $1 \le p_0, p_1, q_0, q_1 \le \infty$ and $y \in \mathbb{R}$,*

$$\|T_{iy}f\|_{q_0} \le M_0(y)\|f\|_{p_0} \quad \textit{and} \quad \|T_{1+iy}f\|_{q_1} \le M_1(y)\|f\|_{p_1},$$

where for some $b < \pi$

$$\sup_{y\in\mathbb{R}} e^{-b|y|} \log M_j(y) < \infty, \quad j = 1, 2.$$

Then for $0 < \theta < 1$, $\mathrm{Re}\, z = \theta$ and p, q defined as in Theorem 1.19, there exists a constant M_θ such that

$$\|T_z f\|_q \le M_\theta \|f\|_p.$$

10.8. Fourier transforms of finite measures.

As we noted above, if μ is a finite Borel measure then $\hat{\mu}$ is a bounded, continuous function. The collection of all such functions obtained in this way is characterized by the following result.

Theorem 1.23. *If h is a bounded, continuous function, then the following are equivalent:*

(1) *$h = \hat{\mu}$ for some positive, finite Borel measure μ;*

(2) *h is positive definite: given any $f \in L^1(\mathbb{R}^n)$,*

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} h(x-y)f(x)\bar{f}(y)\,dxdy \ge 0.$$

This theorem is due to S. Bochner (*Lectures on Fourier Integrals*, Princeton Univ. Press, Princeton, 1959; translated from *Vorlesungen über Fouriersche Integrale*, Akad. Verlag, Leipzig, 1932). Also see Katznelson [**10**, Chapter 6].

Chapter 2

The Hardy-Littlewood Maximal Function

1. Approximations of the identity

Let ϕ be an integrable function on $\mathbb{R}^n$ such that $\int \phi = 1$, and for $t > 0$ define $\phi_t(x) = t^{-n}\phi(t^{-1}x)$. As $t \to 0$, ϕ_t converges in $\mathcal{S}'$ to δ, the Dirac measure at the origin: if $g \in \mathcal{S}$ then

$$\phi_t(g) = \int_{\mathbb{R}^n} t^{-n}\phi(t^{-1}x)g(x)\,dx = \int_{\mathbb{R}^n} \phi(x)g(tx)\,dx,$$

and so by the dominated convergence theorem,

$$\lim_{t\to 0} \phi_t(g) = g(0) = \delta(g).$$

Since $\delta * g = g$, for $g \in \mathcal{S}$ we have the pointwise limit

$$\lim_{t\to 0} \phi_t * g(x) = g(x).$$

Because of this we say that $\{\phi_t : t > 0\}$ is an approximation of the identity.

The summability methods in the previous chapter can be thought of as approximations of the identity. For Cesàro summability, $\phi = F_1$ and $F_R = \phi_{1/R}$. (See (1.24).) For Abel-Poisson summability, $\phi = P_1$ (see (1.30)) and for Gauss-Weierstrass summability, $\phi = W_1$ (see (1.31)). We see from the following result that these summability methods converge in L^p norm.

Theorem 2.1. *Let $\{\phi_t : t > 0\}$ be an approximation of the identity. Then*

$$\lim_{t\to 0} \|\phi_t * f - f\|_p = 0$$

if $f \in L^p$, $1 \le p < \infty$, and uniformly (i.e. when $p = \infty$) if $f \in C_0(\mathbb{R}^n)$.

Proof. Because ϕ has integral 1,

$$\phi_t * f(x) - f(x) = \int_{\mathbb{R}^n} \phi(y)[f(x-ty) - f(x)]\,dy.$$

Given $\epsilon > 0$, choose $\delta > 0$ such that if $|h| < \delta$,

$$\|f(\cdot + h) - f(\cdot)\|_p < \frac{\epsilon}{2\|\phi\|_1}.$$

(Note that δ depends on f.) For fixed δ, if t is sufficiently small then

$$\int_{|y|\geq\delta/t} |\phi(y)|\,dy \leq \frac{\epsilon}{4\|f\|_p}.$$

Therefore, by Minkowski's inequality

$$\begin{aligned}\|\phi_t * f - f\|_p &\leq \int_{|y|<\delta/t} |\phi(y)|\|f(\cdot + ty) - f(\cdot)\|_p\,dy \\ &\qquad + 2\|f\|_p \int_{|y|\geq\delta/t} |\phi(y)|\,dy \\ &< \epsilon.\end{aligned}$$

□

As a consequence of this theorem, we know that there exists a sequence $\{t_k\}$, depending on f, such that $t_k \to 0$ and

$$\lim_{k\to\infty} \phi_{t_k} * f(x) = f(x) \quad \text{a.e.}$$

Hence, if $\lim \phi_t * f(x)$ exists then it must equal $f(x)$ almost everywhere. In Section 4 we will study the existence of this limit.

2. Weak-type inequalities and almost everywhere convergence

Let (X,μ) and (Y,ν) be measure spaces, and let T be an operator from $L^p(X,\mu)$ into the space of measurable functions from Y to $\mathbb{C}$. We say that T is weak (p,q), $q < \infty$, if

$$\nu(\{y \in Y : |Tf(y)| > \lambda\}) \leq \left(\frac{C\|f\|_p}{\lambda}\right)^q,$$

and we say that it is weak (p,∞) if it is a bounded operator from $L^p(X,\mu)$ to $L^\infty(Y,\nu)$. We say that T is strong (p,q) if it is bounded from $L^p(X,\mu)$ to $L^q(Y,\nu)$.

If T is strong (p,q) then it is weak (p,q): if we let $E_\lambda = \{y \in Y : |Tf(y)| > \lambda\}$, then

$$\nu(E_\lambda) = \int_{E_\lambda} d\nu \leq \int_{E_\lambda} \left|\frac{Tf(x)}{\lambda}\right|^q d\nu \leq \frac{\|Tf\|_q^q}{\lambda^q} \leq \left(\frac{C\|f\|_p}{\lambda}\right)^q.$$

When $(X,\mu)=(Y,\nu)$ and T is the identity, the weak (p,p) inequality is the classical Chebyshev inequality.

The relationship between weak (p,q) inequalities and almost everywhere convergence is given by the following result. In it we assume that $(X,\mu) = (Y,\nu)$.

Theorem 2.2. *Let $\{T_t\}$ be a family of linear operators on $L^p(X,\mu)$ and define*

$$T^*f(x) = \sup_t |T_t f(x)|.$$

If T^* is weak (p,q) then the set

$$\{f \in L^p(X,\mu) : \lim_{t\to t_0} T_t f(x) = f(x) \ a.e.\}$$

is closed in $L^p(X,\mu)$.

T^* is called the maximal operator associated with the family $\{T_t\}$.

Proof. Let $\{f_n\}$ be a sequence of functions which converges to f in $L^p(X,\mu)$ norm and such that $T_t f_n(x)$ converges to $f_n(x)$ almost everywhere. Then

$$\begin{aligned}
&\mu(\{x\in X : \limsup_{t\to t_0} |T_t f(x) - f(x)| > \lambda\}) \\
&\quad \le \mu(\{x\in X : \limsup_{t\to t_0} |T_t(f-f_n)(x) - (f-f_n)(x)| > \lambda\}) \\
&\quad \le \mu(\{x\in X : T^*(f-f_n)(x) > \lambda/2\}) \\
&\qquad + \mu(\{x\in X : |(f-f_n)(x)| > \lambda/2\}) \\
&\quad \le \left(\frac{2C}{\lambda}\|f-f_n\|_p\right)^q + \left(\frac{2}{\lambda}\|f-f_n\|_p\right)^p,
\end{aligned}$$

and the last term tends to 0 as $n\to\infty$. Therefore,

$$\begin{aligned}
&\mu(\{x\in X : \limsup_{t\to t_0} |T_t f(x) - f(x)| > 0\}) \\
&\quad \le \sum_{k=1}^{\infty} \mu(\{x\in X : \limsup_{t\to t_0} |T_t f(x) - f(x)| > 1/k\}) \\
&\quad = 0.
\end{aligned}$$

□

By the same technique we can also prove that the set

$$\{f \in L^p(X,\mu) : \lim_{t\to t_0} T_t f(x) \text{ exists a.e.}\}$$

is closed in $L^p(X,\mu)$. It suffices to show that

$$\mu(\{x\in X : \limsup_{t\to t_0} T_t f(x) - \liminf_{t\to t_0} T_t f(x) > \lambda\}) = 0,$$

and this follows as in the above argument since

$$\limsup_{t\to t_0} T_t f(x) - \liminf_{t\to t_0} T_t f(x) \le 2T^* f(x).$$

(If $T_t f(x)$ is complex, we apply this argument to its real and imaginary parts separately.)

Since for $f \in \mathcal{S}$ approximations of the identity converge pointwise to f, we can apply this theorem to show pointwise convergence almost everywhere for $f \in L^p$, $1 \le p < \infty$, or for $f \in C_0$ if we can show that the maximal operator $\sup_{t>0} |\phi_t * f(x)|$ is weakly bounded.

3. The Marcinkiewicz interpolation theorem

Let (X, μ) be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. We call the function $a_f : (0, \infty) \to [0, \infty]$, given by

$$a_f(\lambda) = \mu(\{x \in X : |f(x)| > \lambda\}),$$

the distribution function of f (associated with μ).

Proposition 2.3. *Let $\phi : [0, \infty) \to [0, \infty)$ be differentiable, increasing and such that $\phi(0) = 0$. Then*

$$\int_X \phi(|f(x)|)\, d\mu = \int_0^\infty \phi'(\lambda) a_f(\lambda)\, d\lambda.$$

To prove this it is enough to observe that the left-hand side is equivalent to

$$\int_X \int_0^{|f(x)|} \phi'(\lambda)\, d\lambda\, d\mu$$

and then change the order of integration. If, in particular, $\phi(\lambda) = \lambda^p$ then

$$\|f\|_p^p = p \int_0^\infty \lambda^{p-1} a_f(\lambda)\, d\lambda. \tag{2.1}$$

Since weak inequalities measure the size of the distribution function, this representation of the L^p norm is ideal for proving the following interpolation theorem, which will let us deduce L^p boundedness from weak inequalities. It applies to a larger class of operators than linear ones (note that maximal operators are not linear): an operator T from a vector space of measurable functions to measurable functions is sublinear if

$$|T(f_0 + f_1)(x)| \le |Tf_0(x)| + |Tf_1(x)|,$$
$$|T(\lambda f)| = |\lambda||Tf|, \quad \lambda \in \mathbb{C}.$$

Theorem 2.4 (Marcinkiewicz Interpolation). *Let (X,μ) and (Y,ν) be measure spaces, $1 \le p_0 < p_1 \le \infty$, and let T be a sublinear operator from $L^{p_0}(X,\mu)+L^{p_1}(X,\mu)$ to the measurable functions on Y that is weak (p_0,p_0) and weak (p_1,p_1). Then T is strong (p,p) for $p_0 < p < p_1$.*

Proof. Given $f \in L^p$, for each $\lambda > 0$ decompose f as $f_0 + f_1$, where

$$f_0 = f\chi_{\{x:|f(x)|>c\lambda\}},$$
$$f_1 = f\chi_{\{x:|f(x)|\le c\lambda\}};$$

the constant c will be fixed below. Then $f_0 \in L^{p_0}(\mu)$ and $f_1 \in L^{p_1}(\mu)$. Furthermore,

$$|Tf(x)| \le |Tf_0(x)| + |Tf_1(x)|,$$

so

$$a_{Tf}(\lambda) \le a_{Tf_0}(\lambda/2) + a_{Tf_1}(\lambda/2).$$

We consider two cases.

Case 1: $p_1 = \infty$. Choose $c = 1/(2A_1)$, where A_1 is such that $\|Tg\|_\infty \le A_1\|g\|_\infty$. Then $a_{Tf_1}(\lambda/2) = 0$. By the weak (p_0,p_0) inequality,

$$a_{Tf_0}(\lambda/2) \le \left(\frac{2A_0}{\lambda}\|f_0\|_{p_0}\right)^{p_0};$$

hence,

$$\begin{aligned}\|Tf\|_p^p &\le p\int_0^\infty \lambda^{p-1-p_0}(2A_0)^{p_0}\int_{\{x:|f(x)|>c\lambda\}} |f(x)|^{p_0}\,d\mu\,d\lambda \\ &= p(2A_0)^{p_0}\int_X |f(x)|^{p_0}\int_0^{|f(x)|/c}\lambda^{p-1-p_0}\,d\lambda\,d\mu \\ &= \frac{p}{p-p_0}(2A_0)^{p_0}(2A_1)^{p-p_0}\|f\|_p^p.\end{aligned}$$

Case 2: $p_1 < \infty$. We now have the pair of inequalities

$$a_{Tf_i}(\lambda/2) \le \left(\frac{2A_i}{\lambda}\|f_i\|_{p_i}\right)^{p_i}, \quad i = 0,1.$$

From these we get (arguing as above) that

$$\begin{aligned}\|Tf\|_p^p &\le p\int_0^\infty \lambda^{p-1-p_0}(2A_0)^{p_0}\int_{\{x:|f(x)|>c\lambda\}} |f(x)|^{p_0}\,d\mu\,d\lambda \\ &\quad + p\int_0^\infty \lambda^{p-1-p_1}(2A_1)^{p_1}\int_{\{x:|f(x)|\le c\lambda\}} |f(x)|^{p_1}\,d\mu\,d\lambda \\ &= \left(\frac{p2^{p_0}}{p-p_0}\frac{A_0^{p_0}}{c^{p-p_0}} + \frac{p2^{p_1}}{p_1-p}\frac{A_1^{p_1}}{c^{p-p_1}}\right)\|f\|_p^p.\end{aligned}$$

□

We can write the strong (p,p) norm inequality in this theorem more precisely as

$$\|Tf\|_p \leq 2p^{1/p}\left(\frac{1}{p-p_0}+\frac{1}{p_1-p}\right)^{1/p} A_0^{1-\theta}A_1^{\theta}\|f\|_p, \tag{2.2}$$

where

$$\frac{1}{p}=\frac{\theta}{p_1}+\frac{1-\theta}{p_0}, \quad 0<\theta<1.$$

When $p_1=\infty$ this is the constant which appears in the proof; when $p_1<\infty$ it is enough to take c such that $(2A_0c)^{p_0}=(2A_1c)^{p_1}$ and then simplify.

4. The Hardy-Littlewood maximal function

Let $B_r = B(0,r)$ be the Euclidean ball of radius r centered at the origin. The Hardy-Littlewood maximal function of a locally integrable function f on $\mathbb{R}^n$ is defined by

$$Mf(x)=\sup_{r>0}\frac{1}{|B_r|}\int_{B_r}|f(x-y)|\,dy. \tag{2.3}$$

(This can equal $+\infty$.) If we let $\phi=|B_1|^{-1}\chi_{B_1}$, then (2.3) coincides for non-negative f with the maximal operator associated with the approximation of the identity $\{\phi_t\}$ as in Theorem 2.2.

Sometimes we will define the maximal function with cubes in place of balls. If Q_r is the cube $[-r,r]^n$, define

$$M'f(x)=\sup_{r>0}\frac{1}{(2r)^n}\int_{Q_r}|f(x-y)|\,dy. \tag{2.4}$$

When $n=1$, M and M' coincide; if $n>1$ then there exist constants c_n and C_n, depending only on n, such that

$$c_nM'f(x)\leq Mf(x)\leq C_nM'f(x). \tag{2.5}$$

Because of inequality (2.5), the two operators M and M' are essentially interchangeable, and we will use whichever is more appropriate, depending on the circumstances. In fact, we can define a more general maximal function

$$M''f(x)=\sup_{Q\ni x}\frac{1}{|Q|}\int_{Q}|f(y)|\,dy, \tag{2.6}$$

where the supremum is taken over all cubes containing x. Again, M'' is pointwise equivalent to M. One sometimes distinguishes between M' and M'' by referring to the former as the centered and the latter as the non-centered maximal operator. Alternatively, we could define the non-centered maximal function with balls instead of cubes.

Theorem 2.5. *The operator M is weak $(1,1)$ and strong (p,p), $1 < p \le \infty$.*

We remark that by inequality (2.5), the same result is true for M' (and also for M'').

It is immediate from the definition that

$$\|Mf\|_\infty \le \|f\|_\infty, \tag{2.7}$$

so by the Marcinkiewicz interpolation theorem, to prove Theorem 2.5 it will be enough to prove that M is weak $(1,1)$. Here we will prove this when $n = 1$; we will prove the general case in Section 6. In the one-dimensional case we need the following covering lemma whose simple proof we leave to the reader.

Lemma 2.6. *Let $\{I_\alpha\}_{\alpha\in A}$ be a collection of intervals in $\mathbb{R}$ and let K be a compact set contained in their union. Then there exists a finite subcollection $\{I_j\}$ such that*

$$K \subset \bigcup_j I_j, \quad \text{and} \quad \sum_j \chi_{I_j}(x) \le 2, \quad x \in \mathbb{R}.$$

Proof of Theorem 2.5 for $n = 1$**.** Let $E_\lambda = \{x \in \mathbb{R} : Mf(x) > \lambda\}$. If $x \in E_\lambda$ then there exists an interval I_x centered at x such that

$$\frac{1}{|I_x|}\int_{I_x} |f| > \lambda. \tag{2.8}$$

Let $K \subset E_\lambda$ be compact. Then $K \subset \bigcup I_x$, so by Lemma 2.6 there exists a finite collection $\{I_j\}$ of intervals such that $K \subset \bigcup_j I_j$ and $\sum_j \chi_{I_j} \le 2$. Hence,

$$|K| \le \sum_j |I_j| \le \sum_j \frac{1}{\lambda}\int_{I_j} |f| \le \frac{1}{\lambda}\int_{\mathbb{R}} \sum_j \chi_{I_j}|f| \le \frac{2}{\lambda}\|f\|_1;$$

the second inequality follows from (2.8). Since this inequality holds for every compact $K \subset E_\lambda$, the weak $(1,1)$ inequality for M follows immediately. □

Lemma 2.6 is not valid in dimensions greater than 1, and though one could replace it with similar results, this is not the approach we are going to take here. (For such a proof, see Section 8.6.)

The importance of the maximal function in the study of approximations of the identity comes from the following result.

Proposition 2.7. *Let ϕ be a function which is positive, radial, decreasing (as a function on $(0,\infty)$) and integrable. Then*

$$\sup_{t>0} |\phi_t * f(x)| \le \|\phi\|_1 Mf(x).$$

Proof. If we assume in addition to the given hypotheses that ϕ is a simple function, that is, it can be written as

$$\phi(x) = \sum_j a_j \chi_{B_{r_j}}(x)$$

with $a_j > 0$, then

$$\phi * f(x) = \sum_j a_j |B_{r_j}| \frac{1}{|B_{r_j}|} \chi_{B_{r_j}} * f(x) \leq \|\phi\|_1 M f(x)$$

since $\|\phi\|_1 = \sum a_j |B_{r_j}|$.

An arbitrary function ϕ satisfying the hypotheses can be approximated by a sequence of simple functions which increase to it monotonically. Any dilation ϕ_t is another positive, radial, decreasing function with the same integral, and it will satisfy the same inequality. The desired conclusion follows at once. □

Corollary 2.8. *If $|\phi(x)| \leq \psi(x)$ almost everywhere, where ψ is positive, radial, decreasing and integrable, then the maximal function $\sup_t |\phi_t * f(x)|$ is weak $(1,1)$ and strong (p,p), $1 < p \leq \infty$.*

This is an immediate consequence of Proposition 2.7 and Theorem 2.5. If we combine this corollary with Theorem 2.2 we get the following result.

Corollary 2.9. *Under the hypotheses of the previous corollary, if $f \in L^p$, $1 \leq p < \infty$, or if $f \in C_0$, then*

$$\lim_{t \to 0} \phi_t * f(x) = \left(\int \phi \right) \cdot f(x) \ \text{a.e.}$$

In particular, the summability methods discussed in Chapter 1, Section 9 (Cesàro, Abel-Poisson and Gauss-Weierstrass summability), each converge to $f(x)$ almost everywhere if f is in one of the given spaces.

Proof. Since we have convergence for $f \in \mathcal{S}$, by Theorem 2.2 we have convergence for $f \in \overline{\mathcal{S}} = L^p$ (or $f \in C_0$ if $p = \infty$). The Poisson kernel (1.30) and the Gauss-Weierstrass kernel (1.31) are decreasing; the Féjer kernel (1.24) is not but $F_1(x) \leq \min(1, (\pi x)^{-2})$. □

5. The dyadic maximal function

In $\mathbb{R}^n$ we define the unit cube, open on the right, to be the set $[0,1)^n$, and we let $\mathcal{Q}_0$ be the collection of cubes in $\mathbb{R}^n$ which are congruent to $[0,1)^n$ and whose vertices lie on the lattice $\mathbb{Z}^n$. If we dilate this family of cubes by a factor of 2^{-k} we get the collection $\mathcal{Q}_k$, $k \in \mathbb{Z}$; that is, $\mathcal{Q}_k$ is the family of cubes, open on the right, whose vertices are adjacent points of the lattice $(2^{-k}\mathbb{Z})^n$. The cubes in $\bigcup_k \mathcal{Q}_k$ are called dyadic cubes.

From this construction we immediately get the following properties:

(1) given $x \in \mathbb{R}^n$ there is a unique cube in each family $\mathcal{Q}_k$ which contains it;

(2) any two dyadic cubes are either disjoint or one is wholly contained in the other;

(3) a dyadic cube in $\mathcal{Q}_k$ is contained in a unique cube of each family $\mathcal{Q}_j$, $j < k$, and contains 2^n dyadic cubes of $\mathcal{Q}_{k+1}$.

Given a function $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, define

$$E_k f(x) = \sum_{Q \in \mathcal{Q}_k} \left(\frac{1}{|Q|} \int_Q f \right) \chi_Q(x);$$

$E_k f$ is the conditional expectation of f with respect to the σ-algebra generated by $\mathcal{Q}_k$. It satisfies the following fundamental identity: if Ω is the union of cubes in $\mathcal{Q}_k$, then

$$\int_\Omega E_k f = \int_\Omega f.$$

$E_k f$ is a discrete analog of an approximation of the identity. The following theorem makes this precise; first, define the dyadic maximal function by

$$M_d f(x) = \sup_k |E_k f(x)|. \tag{2.9}$$

Theorem 2.10.

(1) *The dyadic maximal function is weak* $(1,1)$.

(2) *If* $f \in L^1_{\mathrm{loc}}$, $\lim_{k\to\infty} E_k f(x) = f(x)$ *a.e.*

Proof. (1) Fix $f \in L^1$; we may assume that f is non-negative: if f is real, it can be decomposed into its positive and negative parts, and if it is complex, into its real and imaginary parts.

Now let

$$\{x \in \mathbb{R}^n : M_d f(x) > \lambda\} = \bigcup_k \Omega_k,$$

where

$$\Omega_k = \{x \in \mathbb{R}^n : E_k f(x) > \lambda \text{ and } E_j f(x) \le \lambda \text{ if } j < k\};$$

that is, $x \in \Omega_k$ if $E_k f(x)$ is the first conditional expectation of f which is greater than λ. (Since $f \in L^1$, $E_k f(x) \to 0$ as $k \to -\infty$, so such a k exists.)

The sets Ω_k are clearly disjoint, and each one can be written as the union of cubes in $\mathcal{Q}_k$. Hence,

$$\begin{aligned}|\{x \in \mathbb{R}^n : M_d f(x) > \lambda\}| &= \sum_k |\Omega_k| \\ &\leq \sum_k \frac{1}{\lambda} \int_{\Omega_k} E_k f \\ &= \frac{1}{\lambda} \sum_k \int_{\Omega_k} f \\ &\leq \frac{1}{\lambda} \|f\|_1.\end{aligned}$$

(2) This limit is clearly true if f is continuous, and so by Theorem 2.2 it holds for $f \in L^1$. To complete the proof, note that if $f \in L^1_{\text{loc}}$ then $f\chi_Q \in L^1$ for any $Q \in \mathcal{Q}_0$. Hence, (2) holds for almost every $x \in Q$, and so for almost every $x \in \mathbb{R}^n$. □

This proof uses a decomposition of $\mathbb{R}^n$ which has proved to be extremely useful. It is called the Calderón-Zygmund decomposition and we state it precisely as follows.

Theorem 2.11. *Given a function f which is integrable and non-negative, and given a positive number λ, there exists a sequence $\{Q_j\}$ of disjoint dyadic cubes such that*

(1) $f(x) \leq \lambda$ *for almost every* $x \notin \bigcup_j Q_j$;

(2) $\displaystyle\Big|\bigcup_j Q_j\Big| \leq \frac{1}{\lambda}\|f\|_1$;

(3) $\displaystyle\lambda < \frac{1}{|Q_j|}\int_{Q_j} f \leq 2^n \lambda.$

Proof. As in the proof of Theorem 2.10, form the sets Ω_k and decompose each into disjoint dyadic cubes contained in $\mathcal{Q}_k$; together, all of these cubes form the family $\{Q_j\}$.

Part (2) of the theorem is then just the weak $(1, 1)$ inequality of Theorem 2.10.

If $x \notin \bigcup_j Q_j$ then for every k, $E_k f(x) \leq \lambda$, and so by part (2) of Theorem 2.10, $f(x) \leq \lambda$ at almost every such point.

Finally, by the definition of the sets Ω_k, the average of f over Q_j is greater than λ; this is the first inequality in (3). Furthermore, if $\tilde{Q}_j$ is the dyadic cube containing Q_j whose sides are twice as long, then the average of f over $\tilde{Q}_j$ is at most λ. Therefore,

$$\frac{1}{|Q_j|}\int_{Q_j} f \le \frac{|\tilde{Q}_j|}{|Q_j|}\frac{1}{|\tilde{Q}_j|}\int_{\tilde{Q}_j} f \le 2^n\lambda.$$

□

6. The weak $(1,1)$ inequality for the maximal function

We are now going to use Theorem 2.10 to prove Theorem 2.5. In fact, it is an immediate consequence of the following lemma and inequality (2.5). (Recall that M' is the maximal operator on cubes defined by (2.4).)

Lemma 2.12. *If f is a non-negative function, then*

$$|\{x \in \mathbb{R}^n : M'f(x) > 4^n\lambda\}| \le 2^n|\{x \in \mathbb{R}^n : M_d f(x) > \lambda\}|.$$

Given this lemma, by the weak $(1,1)$ inequality for M_d proved in Theorem 2.10,

$$|\{x \in \mathbb{R}^n : M'f(x) > \lambda\}| \le 2^n|\{x \in \mathbb{R}^n : M_d f(x) > 4^{-n}\lambda\}| \le \frac{8^n}{\lambda}\|f\|_1.$$

(Since $M'f = M'(|f|)$, we may assume that f is non-negative.)

Proof of Lemma 2.12. As before, we form the decomposition

$$\{x \in \mathbb{R}^n : M_d f(x) > \lambda\} = \bigcup_j Q_j.$$

Let $2Q_j$ be the cube with the same center as Q_j and whose sides are twice as long. To complete the proof it will suffice to show that

$$\{x \in \mathbb{R}^n : M'f(x) > 4^n\lambda\} \subset \bigcup_j 2Q_j.$$

Fix $x \notin \bigcup_j 2Q_j$ and let Q be any cube centered at x. Let $l(Q)$ denote the side length of Q, and choose $k \in \mathbb{Z}$ such that $2^{k-1} \le l(Q) < 2^k$. Then Q intersects $m \le 2^n$ dyadic cubes in $\mathcal{Q}_k$; call them $R_1, R_2, \dots, R_m$. None of these cubes is contained in any of the Q_j's, for otherwise we would have $x \in \bigcup_j 2Q_j$. Hence, the average of f on each R_i is at most λ, and so

$$\frac{1}{|Q|}\int_Q f = \frac{1}{|Q|}\sum_{i=1}^m \int_{Q\cap R_i} f \le \sum_{i=1}^m \frac{2^{kn}}{|Q|}\frac{1}{|R_i|}\int_{R_i} f \le 2^n m\lambda \le 4^n\lambda.$$

□

As a consequence of the weak $(1,1)$ inequality and Theorem 2.2 we get a continuous analog of the second half of Theorem 2.10.

Corollary 2.13 (Lebesgue Differentiation Theorem)**.** *If* $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ *then*

$$\lim_{r\to 0^+} \frac{1}{|B_r|} \int_{B_r} f(x-y)\,dy = f(x) \ \textit{a.e.}$$

From this we see that $|f(x)| \leq Mf(x)$ almost everywhere. The same is true if we replace M by M' or M''.

We can make the conclusion of Corollary 2.13 sharper:

$$\lim_{r\to 0^+} \frac{1}{|B_r|} \int_{B_r} |f(x-y) - f(x)|\,dy = 0 \text{ a.e.} \tag{2.10}$$

This follows immediately from the fact that

$$\frac{1}{|B_r|} \int_{B_r} |f(x-y) - f(x)|\,dy \leq Mf(x) + |f(x)|.$$

The points in $\mathbb{R}^n$ for which limit (2.10) equals 0 are called the Lebesgue points of f. If x is a Lebesgue point and if $\{B_j\}$ is a sequence of balls such that $B_1 \supset B_2 \supset \cdots$ and $\bigcap_j B_j = \{x\}$ (note that the balls need not be centered at x), then

$$\lim_{j\to\infty} \frac{1}{|B_j|} \int_{B_j} f = f(x).$$

This follows immediately from the inclusion $B_j \subset B(x, 2r_j)$, where r_j is the radius of B_j. A similar argument shows that the set of Lebesgue points of f does not change if we take cubes instead of balls.

The weak $(1,1)$ inequality for M is a substitute for the strong $(1,1)$ inequality, which is false. In fact, it never holds, as the following result shows.

Proposition 2.14. *If* $f \in L^1$ *and is not identically* 0, *then* $Mf \notin L^1$.

The proof is simple: since f is not identically 0, there exists $R > 0$ such that

$$\int_{B_R} |f| \geq \epsilon > 0.$$

Now if $|x| > R$, $B_R \subset B(x, 2|x|)$, so

$$Mf(x) \geq \frac{1}{(2|x|)^n} \int_{B_R} |f| \geq \frac{\epsilon}{2^n|x|^n}.$$

Nevertheless, we do have the following.

Theorem 2.15. *If B is a bounded subset of $\mathbb{R}^n$, then*

$$\int_B Mf \leq 2|B| + C\int_{\mathbb{R}^n} |f| \log^+ |f|,$$

where $\log^+ t = \max(\log t, 0)$.

Proof.

$$\begin{aligned}\int_B Mf &\leq 2\int_0^\infty |\{x \in B : Mf(x) > 2\lambda\}|\, d\lambda \\ &\leq 2|B| + 2\int_1^\infty |\{x \in B : Mf(x) > 2\lambda\}|\, d\lambda.\end{aligned}$$

Decompose f as $f_1 + f_2$, where $f_1 = f\chi_{\{x:|f(x)|>\lambda\}}$ and $f_2 = f - f_1$. Then

$$\{x \in B : Mf(x) > 2\lambda\} \subset \{x \in B : Mf_1(x) > \lambda\}.$$

Hence,

$$\begin{aligned}\int_1^\infty |\{x \in B : Mf(x) > 2\lambda\}|\, d\lambda &\leq \int_1^\infty \frac{C}{\lambda}\int_{\{x:|f(x)|>\lambda\}} |f(x)|\, dx\, d\lambda \\ &\leq C\int_{\mathbb{R}^n} |f(x)| \int_1^{\max(|f(x)|,1)} \frac{d\lambda}{\lambda}\, dx \\ &= C\int_{\mathbb{R}^n} |f(x)| \log^+ |f(x)|\, dx.\end{aligned}$$

□

7. A weighted norm inequality

Theorem 2.16. *If w is a non-negative, measurable function and $1 < p < \infty$, then there exists a constant C_p such that*

$$\int_{\mathbb{R}^n} Mf(x)^p w(x)\, dx \leq C_p \int_{\mathbb{R}^n} |f(x)|^p Mw(x)\, dx.$$

Furthermore,

$$\int_{\{x:Mf(x)>\lambda\}} w(x)\, dx \leq \frac{C_1}{\lambda}\int_{\mathbb{R}^n} |f(x)| Mw(x)\, dx. \tag{2.11}$$

Proof. It will suffice to show that $\|Mf\|_{L^\infty(w)} \leq \|f\|_{L^\infty(Mw)}$ and that the weak $(1,1)$ inequality holds; the strong (p,p) inequality then follows from the Marcinkiewicz interpolation theorem. If $Mw(x) = 0$ for any x, then

$w(x) = 0$ almost everywhere and there is nothing to prove. Therefore, we may assume that for every x, $Mw(x) > 0$. If $a > \|f\|_{L^\infty(Mw)}$ then

$$\int_{\{x:|f(x)|>a\}} Mw(x)\,dx = 0,$$

and so $|\{x \in \mathbb{R}^n : |f(x)| > a\}| = 0$; that is, $|f(x)| \leq a$ almost everywhere. From this it follows that $Mf(x) \leq a$ a.e., so $\|Mf\|_{L^\infty(w)} \leq a$. Hence, $\|Mf\|_{L^\infty(w)} \leq \|f\|_{L^\infty(Mw)}$.

To prove the weak $(1,1)$ inequality we may assume that f is non-negative and $f \in L^1(\mathbb{R}^n)$. (If $f \in L^1(Mw)$ then $f_j = f\chi_{B(0,j)}$ is a sequence of integrable functions which increase pointwise to f.) If $\{Q_j\}$ is the Calderón-Zygmund decomposition of f at height $\lambda > 0$, then as we showed in the proof of Lemma 2.12,

$$\{x \in \mathbb{R}^n : M'f(x) > 4^n\lambda\} \subset \bigcup_j 2Q_j;$$

hence,

$$\begin{aligned}\int_{\{x:M'f(x)>4^n\lambda\}} w(x)\,dx &\leq \sum_j \int_{2Q_j} w(x)\,dx \\ &= \sum_j 2^n|Q_j|\frac{1}{|2Q_j|}\int_{2Q_j} w(x)\,dx \\ &\leq \frac{2^n}{\lambda}\sum_j \int_{Q_j} f(y)\left(\frac{1}{|2Q_j|}\int_{2Q_j} w(x)\,dx\right)dy \\ &\leq \frac{2^nC}{\lambda}\int_{\mathbb{R}^n} f(y)M''w(y)\,dy.\end{aligned}$$

Since $M''w(y) \leq C_nMw(y)$, we get the desired inequality. □

If w is such that $Mw(x) \leq w(x)$ a.e., then these inequalities hold with the same weight w on both sides. Functions which satisfy this condition are called A_1 weights, and we will consider them in greater detail in Chapter 7.

8. Notes and further results

8.1. References.

The maximal function for $n = 1$ was introduced by G. H. Hardy and J. E. Littlewood (*A maximal theorem with function-theoretic applications*, Acta Math. **54** (1930), 81–116), and for $n > 1$ by N. Wiener (*The ergodic theorem*, Duke Math. J. **5** (1939), 1–18). In their article, Hardy and Littlewood first consider the discrete case, about which they say: "The problem is most easily grasped when stated in the language of cricket, or any other game in which a player compiles a series of scores of which an average is recorded." Their proof, which uses decreasing rearrangements (see 8.2 below) can be found in Zygmund [**21**]. Our proof follows the ideas of Calderón and Zygmund (*On the existence of certain singular integrals*, Acta Math. **88** (1952), 85–139). The decomposition which bears their name also first appeared in this paper. The method of rotations, discussed in Chapter 4, lets us deduce the strong (p,p) inequality for $n > 1$ from the one-dimensional result. For a discussion of questions related to Theorem 2.2, and in particular the necessity of the condition in that theorem, see de Guzmán [**7**]. The Marcinkiewicz interpolation theorem was announced by J. Marcinkiewicz (*Sur l'interpolation d'opérations*, C. R. Acad. Sci. Paris **208** (1939), 1272–1273). However, he died in World War II and a complete proof was finally given by A. Zygmund (*On a theorem of Marcinkiewicz concerning interpolation of operations*, J. Math. Pures Appl. **34** (1956), 223–248). Both Marcinkiewicz interpolation and Riesz-Thorin interpolation in Chapter 1 have been generalized considerably; see, for example the book by C. Bennett and R. Sharpley (*Interpolation of Operators*, Academic Press, New York, 1988). The weighted norm inequality for the maximal operator is due to C. Fefferman and E. M. Stein (*Some maximal inequalities*, Amer. J. Math. **93** (1971), 107–115).

8.2. The Hardy operator, one-sided maximal functions, and decreasing rearrangements of functions.

Given a function g on $\mathbb{R}^+ = (0,\infty)$ the Hardy operator acting on g is defined by

$$Tg(t) = \frac{1}{t}\int_0^t g(s)\,ds, \quad t \in \mathbb{R}^+.$$

If $g \in L^1(\mathbb{R}^+)$ is non-negative, then, since Tg is continuous, one can show that

$$E(\lambda) = \frac{1}{\lambda}\int_{E(\lambda)} g(t)\,dt, \tag{2.12}$$

where $E(\lambda) = \{t \in \mathbb{R}^+ : Tg(t) > \lambda\}$. From this and (2.1) we get $\|Tg\|_p \leq p'\|g\|_p$, $1 < p \leq \infty$. Other proofs of this result and generalizations of the operator can be found in the classical book by G. H. Hardy, J. E. Littlewood and G. Pólya (*Inequalities*, Cambridge Univ. Press, Cambridge, 1987, first edition in 1932), or in Chapter 2 of the book by Bennett and Sharpley cited above.

For a function f on $\mathbb{R}$, the one-sided Hardy-Littlewood maximal functions are defined by

$$M^+f(t) = \sup_{h>0} \frac{1}{h} \int_t^{t+h} |f(s)|\, ds \quad \text{and} \quad M^-f(t) = \sup_{h>0} \frac{1}{h} \int_{t-h}^{t} |f(s)|\, ds.$$

The maximal function as defined by Hardy and Littlewood corresponds to M^- for functions on $\mathbb{R}^+$ (with $0 < h < t$ in the definition). When $|f|$ is decreasing this maximal function coincides with the Hardy operator acting on $|f|$.

If f is a measurable function on $\mathbb{R}^n$, we can define a decreasing function f^* on $(0, \infty)$, called the decreasing rearrangement of f, that has the same distribution function as f:

$$f^*(t) = \inf\{\lambda : a_f(\lambda) \le t\}.$$

Because f and f^* have the same distribution function, by (2.1) their L^p norms are equal, as well as any other quantity which depends only on their distribution function. (Cf. Section 8.3 below.) The action of the Hardy operator on f^* is usually denoted by f^{**}.

Hardy and Littlewood showed that for functions on $\mathbb{R}^+$,

$$|\{x : M^-f(x) > \lambda\}| \le |\{x : f^{**}(x) > \lambda\}|, \tag{2.13}$$

so the weak $(1, 1)$ and strong (p, p) inequalities for M^- follow from the corresponding ones for T.

A beautiful proof of the weak $(1, 1)$ inequality for M^+ was given by F. Riesz as an application of his "rising sun lemma" (*Sur un théorème du maximum de MM. Hardy et Littlewood*, J. London Math. Soc. **7** (1932), 10–13). Given a function F on $\mathbb{R}$, a point x is a shadow point of F if there exists $y > x$ such that $F(y) > F(x)$. The set of shadow points of F is denoted by $S(F)$.

Lemma 2.17. *Let F be a continuous function such that*

$$\lim_{x\to+\infty} F(x) = -\infty \quad \textit{and} \quad \lim_{x\to-\infty} F(x) = +\infty.$$

Then $S(F)$ is open and can be written as the disjoint union $\bigcup_j (a_j, b_j)$ of finite open intervals such that $F(a_j) = F(b_j)$.

Given $f \in L^1(\mathbb{R})$ and $\lambda > 0$, define $F(x) = \int_0^x |f(t)|\, dt - \lambda x$. Then $E(\lambda) = \{x : M^+f(x) > \lambda\}$ equals $S(F)$. Using Lemma 2.17 one can deduce that

$$E(\lambda) = \frac{1}{\lambda} \int_{E(\lambda)} |f(x)|\, dx,$$

which is similar to (2.12) for the Hardy operator. The strong (p, p) inequality now follows as before with constant p' (which is sharp). We leave the details of the proof of this and of Lemma 2.17 to the reader.

An inequality similar to (2.13) holds for the maximal operator acting on functions on $\mathbb{R}^n$; in fact, we have the following pointwise inequality: there exist positive constants c_n and C_n such that

$$c_n(Mf)^*(t) \leq f^{**}(t) \leq C_n(Mf)^*(t), \quad t \in \mathbb{R}^+.$$

The left-hand inequality is due to F. Riesz; the right-hand inequality is due to C. Herz ($n = 1$) and C. Bennet and R. Sharpley ($n > 1$). See Chapter 3 of the above-cited book by these authors for a proof and further references.

8.3. The Lorentz spaces $L^{p,q}$.

Let (X, μ) be a measure space. $L^{p,q}(X)$ denotes the space of measurable functions f which satisfy

$$\|f\|_{p,q} = \left(\frac{q}{p}\int_0^\infty [t^{1/p}f^*(t)]^q \frac{dt}{t}\right)^{1/q} < \infty$$

when $1 \leq p < \infty$, $1 \leq q < \infty$, and

$$\|f\|_{p,\infty} = \sup_{t>0} t^{1/p}f^*(t) < \infty$$

when $1 \leq p \leq \infty$. When $p = q$,

$$\|f\|_{p,p} = \|f^*\|_p = \|f\|_p$$

and we recover L^p. In general, however, $\|\cdot\|_{p,q}$ is not a norm since the triangle inequality only holds when $1 \leq q \leq p < \infty$ or $p = q = \infty$. But when $1 < p \leq \infty$ and $1 \leq q \leq \infty$, if we replace f^* in the definition of $\|f\|_{p,q}$ with f^{**}, we get a quantity which is equivalent to $\|f\|_{p,q}$ and which defines a norm.

If $q_1 \leq q_2$ then $\|f\|_{p,q_2} \leq \|f\|_{p,q_1}$, so $L^{p,q_1} \subset L^{p,q_2}$.

An operator T is weak (p, p) precisely when

$$\|Tf\|_{p,\infty} \leq C\|f\|_p.$$

The Marcinkiewicz interpolation theorem can be generalized to these spaces; this lets us, for example, give a version of the Hausdorff-Young inequality which is stronger than Corollary 1.20: if $f \in L^p(\mathbb{R}^n)$, $1 < p \leq 2$, then $\hat{f} \in L^{p',p}$ and there exists a constant B_p such that

$$\|\hat{f}\|_{p',p} \leq B_p\|f\|_p.$$

For more information on decreasing rearrangements and the $L^{p,q}$ spaces, see Stein and Weiss [**18**, Chapter 5] and the book by Bennett and Sharpley cited above. Also see this latter book for more on interpolation theorems, particularly the so-called real method of interpolation which is well suited to Lorentz spaces. These spaces were introduced by G. G. Lorentz in two papers (*Some new functional spaces*, Ann. of Math. **51** (1950), 37–55, and *On the theory of spaces* Λ, Pacific J. Math. **1** (1951), 411–429).

8.4. $L\log L$**.**

In the proof of Proposition 2.14, the integrability of Mf failed at infinity, and did not exclude local integrability. However, the example $f(x) = x^{-1}(\log x)^{-2}\chi_{(0,1/2]}$ shows that even local integrability can fail. A partial converse of Theorem 2.15 is true and characterizes when Mf is locally integrable.

Theorem 2.18. *If f is an integrable function supported on a compact set B, then $Mf \in L^1(B)$ if and only if $f\log^+ f \in L^1(B)$.*

This is due to E. M. Stein (*Note on the class* $L\log L$, Studia Math. **32** (1969), 305–310); a proof can also be found in García-Cuerva and Rubio de Francia [**6**, p. 146].

At the heart of the proof is a stronger version of the weak $(1,1)$ inequality and a "reverse" weak $(1,1)$ inequality for the maximal function:

$$|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \leq \frac{C}{\lambda}\int_{\{x:|f(x)|>\lambda/2\}} |f(x)|\,dx, \tag{2.14}$$

$$|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \geq \frac{c}{\lambda}\int_{\{x:|f(x)|>\lambda\}} |f(x)|\,dx. \tag{2.15}$$

Inequality (2.14) follows from the weak $(1,1)$ inequality applied to $f_1 = f\chi_{\{x:|f(x)|>\lambda/2\}}$; inequality (2.15) follows from the Calderón-Zygmund decomposition if we replace M by M''.

In this and related problems it would be useful to consider functions f such that $f\log^+ f \in L^1$ as members of a Banach space. Unfortunately, the expression $\int f\log^+ f$ does not define a norm. One way around this is to introduce the Luxemburg norm: given a set $\Omega \subseteq \mathbb{R}^n$, define

$$\|f\|_{L\log L(\Omega)} = \inf\left\{\lambda > 0 : \int_\Omega \frac{|f(x)|}{\lambda}\log^+\left(\frac{|f(x)|}{\lambda}\right)dx \leq 1\right\}. \tag{2.16}$$

By using this we can strengthen the conclusion of Theorem 2.18 to the following inequality: $\|Mf\|_{L^1(B)} \leq C\|f\|_{L\log L(B)}$. (See Zygmund [**21**, Chapter 4].)

By replacing the function $t\log^+ t$ in (2.16) by any convex, increasing function Φ, we get a class of Banach function spaces, $L^\Phi(\Omega)$, which generalize the L^p spaces and are referred to as Orlicz spaces. These have a rich theory; for further information, consult the books by M. A. Krasnosel'skiĭ and Ya. B. Rutickiĭ, (*Convex functions and Orlicz spaces*, P. Noordhoff, Groningen, 1961), and M. M. Rao and Z. D. Ren, (*Theory of Orlicz Spaces*, Marcel Dekker, New York, 1991).

8.5. The size of constants.

In the proofs in this chapter, the constants which appear in the weak $(1,1)$ and strong (p,p) inequalities, $1 < p < \infty$, for M and M' are of exponential type with exponent n. For the strong (p,p) inequality, the method of rotations (see Corollary 4.7) gives a better constant of the form $C_p n$ for M. Furthermore, one can show that there exist constants C_p, independent of n, such that

$$\|Mf\|_p \leq C_p \|f\|_p, \qquad 1 < p < \infty.$$

This result is due to E. M. Stein, and the proof appeared in an article by E. M. Stein and J.-O. Strömberg (*Behavior of maximal functions in* $\mathbb{R}^n$ *for large* n, Ark. Mat. **21** (1983), 259–269); also see E. M. Stein, *Three variations on the theme of maximal functions* (Recent Progress in Fourier Analysis, I. Peral and J. L. Rubio de Francia, eds., pp. 229–244, North-Holland, Amsterdam, 1985). A partial generalization of this result is possible: let B be a convex set which is symmetric about the origin and define

$$M_B f(x) = \sup_{r>0} \frac{1}{|rB|} \int_{rB} |f(x-y)|\, dy.$$

Then if $p > 3/2$ there exists a constant C_p independent of B and n such that

$$\|M_B f\|_p \leq C_p \|f\|_p.$$

For the weak $(1,1)$ inequality, the best constant known for M is of order n. (See the article by Stein and Strömberg cited above.)

For the non-centered maximal function defined by (2.6) (but with balls instead of cubes), the best constant in the strong (p,p) inequality must grow exponentially in n. See L. Grafakos and S. Montgomery-Smith (*Best constants for uncentered maximal functions*, Bull. London Math. Soc. **29** (1997), 60–64).

We note that when $n = 1$ the best constants are known for the one-sided and for the non-centered maximal operator M'' but not for M. In the case of the weak $(1,1)$ inequality for M'', one can use Lemma 2.6 to get $C = 2$, and the function $f(t) = \chi_{[0,1]}$ shows this is the best possible. For the strong (p,p) inequality for M'', see the article by Grafakos and Montgomery-Smith cited above. For lower and upper bounds on the weak $(1,1)$ constants for M, see the article by J. M. Aldaz (*Remarks on the Hardy-Littlewood maximal function*, Proc. Roy. Soc. Edinburgh Sect. A **128** (1998), 1–9).

8.6. Covering lemmas.

A standard approach to proving that the maximal function in $\mathbb{R}^n$ is weak $(1,1)$ is to use covering lemmas. Here we give two; the first is a Vitali-type lemma due to N. Wiener (in the paper cited above). If $B = B(x,r)$ and $t > 0$, then we let $tB = B(x,tr)$.

Theorem 2.19. *Let $\{B_j\}_{j\in\mathcal{J}}$ be a collection of balls in $\mathbb{R}^n$. Then there exists an at most countable subcollection of disjoint balls $\{B_k\}$ such that*

$$\bigcup_{j\in\mathcal{J}} B_j \subset \bigcup_k 5B_k.$$

The second is due independently to A. Besicovitch and A. P. Morse; for a proof and further references, see the book by M. de Guzmán (*Differentiation of Integrals in* $\mathbb{R}^n$, Lecture Notes in Math. **481**, Springer-Verlag, Berlin, 1985).

Theorem 2.20. *Let A be a bounded set in $\mathbb{R}^n$, and suppose that $\{B_x\}_{x\in A}$ is a collection of balls such that $B_x = B(x, r_x)$, $r_x > 0$. Then there exists an at most countable subcollection of balls $\{B_j\}$ and a constant C_n, depending only on the dimension, such that*

$$A \subset \bigcup_j B_j \quad \textit{and} \quad \sum_j \chi_{B_j}(x) \le C_n.$$

The same result holds if balls are replaced by cubes; more generally, the point x need not be the center of the ball or cube but must be uniformly close to the center.

Using the Besicovitch-Morse lemma, we can extend our results for the maximal function to L^p spaces with respect to other measures. Given a non-negative Borel measure μ, define the maximal function

$$M_\mu f(x) = \sup_{r>0} \frac{1}{\mu(B_r)} \int_{B_r} |f(x-y)|\, d\mu(y).$$

(If $\mu(B_r) = 0$, define the μ-average of f on B_r to be zero.) Then one can show that M_μ is weak $(1,1)$ with respect to μ; hence, by interpolation it is bounded on $L^p(\mu)$, $1 < p < \infty$.

Note that if we define M'_μ as the maximal operator with respect to cubes, then M_μ and M'_μ need not be pointwise equivalent unless μ satisfies an additional doubling condition: there exists C such that for any ball B, $\mu(2B) \le C\mu(B)$. Furthermore, if we define the non-centered maximal operator M''_μ, then this need not be weak $(1,1)$ if $n > 1$ unless μ satisfies a doubling condition. In this case the weak $(1,1)$ inequality follows from

Wiener's lemma; when $n = 1$, Lemma 2.6 can be applied even to non-doubling measures. An example of μ such that M''_μ is not weak $(1,1)$ is due to P. Sjögren (*A remark on the maximal function for measures on* $\mathbb{R}^n$, Amer. J. Math. **105** (1983), 1231–1233). For certain kinds of measures a weaker hypothesis than doubling implies that M''_μ is weak $(1,1)$; see, for example, the paper by A. Vargas (*On the maximal function for rotation invariant measures in* $\mathbb{R}^n$, Studia Math. **110** (1994), 9–17).

8.7. The non-tangential Poisson maximal function.

Given $f \in L^p(\mathbb{R}^n)$, $u(x,t) = P_t * f(x)$ defines the harmonic extension of f to the upper half-space $\mathbb{R}^{n+1}_+ = \mathbb{R}^n \times (0,\infty)$, and $\lim_{t\to 0} P_t * f(x)$ is the limit of this function on the boundary as we approach x "vertically", that is, along the line perpendicular to $\mathbb{R}^n$. More generally, one can consider a non-tangential approach: fix $a > 0$ and let

$$\Gamma_a(x) = \{(y,t) : |y - x| < at\}$$

be the cone with vertex x and aperture a. We can then ask whether

$$\lim_{\substack{(y,t)\to(x,0)\\ (y,t)\in\Gamma_a(x)}} u(y,t) = f(x) \qquad \text{a.e. } x \in \mathbb{R}^n.$$

Since this limit holds for all x if f is continuous and has compact support, it suffices to consider the maximal function

$$u_a^*(x) = \sup_{(y,t)\in\Gamma_a(x)} |u(y,t)|,$$

so as to apply Theorem 2.2. But one can show that there exists a constant C_a, depending on a, such that

$$u_a^*(x) \le C_a Mf(x),$$

and so u_a is weak $(1,1)$ and strong (p,p), $1 < p \le \infty$.

These results can be generalized to "tangential" approach regions, provided the associated maximal function is weakly bounded. For results in this direction, see, for example, the articles by A. Nagel and E. M. Stein (*On certain maximal functions and approach regions*, Adv. in Math. **54** (1984), 83–106) and D. Cruz-Uribe, C. J. Neugebauer and V. Olesen (*Norm inequalities for the minimal and maximal operator, and differentiation of the integral*, Publ. Mat. **41** (1997), 577–604).

8.8. The strong maximal function.

Let $R(h_1,\dots,h_n) = [-h_1,h_1] \times \cdots \times [-h_n,h_n]$. If $f \in L^1_{\text{loc}}$, define the strong maximal function of f by

$$M_s f(x) = \sup_{h_1,\dots,h_n>0} \frac{1}{|R(h_1,\dots,h_n)|} \int_{R(h_1,\dots,h_n)} |f(x-y)|\,dy.$$

Then M_s is bounded on $L^p(\mathbb{R}^n)$, $p > 1$; this is a consequence of the one-dimensional result. However, M_s is not weak $(1,1)$: the best possible inequality is

$$|\{x \in \mathbb{R}^n : M_s f(x) > \lambda\}| \leq C \int \frac{|f(x)|}{\lambda} \left(1 + \log^+ \frac{|f(x)|}{\lambda}\right)^{n-1} dx.$$

This result is due to B. Jessen, J. Marcinkiewicz and A. Zygmund (*Note on the differentiability of multiple integrals*, Fund. Math. **25** (1935), 217–234). A geometric proof was given by A. Córdoba and R. Fefferman (*A geometric proof of the strong maximal theorem*, Ann. of Math. **102** (1975), 95–100); also see the article by R. J. Bagby (*Maximal functions and rearrangements: some new proofs*, Indiana Univ. Math. J. **32** (1983), 879–891).

The analog of the Lebesgue differentiation theorem,

$$\lim_{\max(h_i)\to 0} \frac{1}{|R(h_1,\dots,h_n)|} \int_{R(h_1,\dots,h_n)} f(x-y)\,dy = f(x) \text{ a.e.}$$

is false for some $f \in L^1$, but is true if $f(1+\log^+|f|)^{n-1}$ is locally integrable, and in particular if $f \in L^p_{\text{loc}}$ for some $p > 1$.

If in the definition of M_s we allow rectangles (that is, parallelepipeds) with arbitrary orientation (and not just with edges parallel to the coordinate axes), then the resulting operator is not bounded on any L^p, $p < \infty$, and the associated differentiation theorem does not hold even for f bounded.

For these and other problems related to differentiation of the integral see de Guzmán [**7**], the book by the same author cited above, and the monograph by A. M. Bruckner (*Differentiation of integrals*, Amer. Math. Monthly **78** (1971), Slaught Memorial Papers, 12).

8.9. The Kakeya maximal function.

Given $N > 1$, let $\mathcal{R}_N$ be the set of all rectangles in $\mathbb{R}^n$ with $n-1$ sides of length h and one side of length Nh, $h > 0$. The Kakeya maximal function is defined by

$$\mathcal{K}_N f(x) = \sup_{x\in R\in\mathcal{R}_N} \frac{1}{|R|}\int_R |f|.$$

Since each rectangle in $\mathcal{R}_N$ is contained in a ball of radius $c_n Nh$, the Kakeya maximal function is bounded pointwise by the Hardy-Littlewood maximal function:

$$\mathcal{K}_N f(x) \leq C_n N^{n-1} M f(x)\,. \tag{2.17}$$

It follows immediately that $\mathcal{K}_N$ is bounded on L^p, $1 < p \leq \infty$.

An important problem is to determine the size of the constants as functions of N for the L^p estimates for the Kakeya maximal function. Inequality

(2.17) gives the trivial estimate N^{n-1} for the constant in the weak $(1,1)$ inequality, and it is clear that in L^∞, $\mathcal{K}_N$ has norm 1. Interpolating between these values we get

$$\|\mathcal{K}_N f\|_p \leq C_n N^{(n-1)/p}\|f\|_p, \quad 1 < p < \infty.$$

It is conjectured that the constant can be improved to a power of $\log N$ when $p = n$, which, again by interpolation, would give the following bounds:

(1) if $p \geq n$ then $\|\mathcal{K}_N f\|_p \leq C_n(\log N)^{\alpha_p}\|f\|_p$;

(2) if $1 < p \leq n$ then $\|\mathcal{K}_N f\|_p \leq C_n N^{n/p-1}(\log N)^{\alpha_p}\|f\|_p$.

This conjecture is closely related to two other problems: the Hausdorff dimension of the Kakeya set—a set with Lebesgue measure zero which contains a line segment in each direction (see Stein [**17**, p. 434]); and the boundedness properties of Bochner-Riesz multipliers (see Chapter 8, Sections 5 and 8.3).

This conjecture has been proved completely only when $n = 2$: see the paper by A. Córdoba (*The Kakeya maximal function and the spherical summation multipliers*, Amer. J. Math. **99** (1977), 1–22). In this paper he also discussed the connection with Bochner-Riesz multipliers. Córdoba later proved the result in all dimensions when $1 < p \leq 2$; see *A note on Bochner-Riesz operators* (Duke Math. J. **46** (1979), 505–511). M. Christ, J. Duoandikoetxea and J. L. Rubio de Francia (*Maximal operators related to the Radon transform and the Calderón-Zygmund method of rotations*, Duke Math. J. **53** (1986), 189–209) extended the range to $1 < p \leq (n+1)/2$.

A great deal of activity on this problem has been spurred by the work of J. Bourgain (*Besicovitch type maximal operators and applications to Fourier analysis*, Geom. Funct. Anal. **1** (1991), 147–187). In this paper he proved the conjecture for $1 < p \leq (n+1)/2 + \epsilon_n$, where ϵ_n is given by an inductive formula (for instance, $\epsilon_3 = 1/3$). T. Wolff (*An improved bound for Kakeya type maximal functions*, Rev. Mat. Iberoamericana **11** (1995), 651–674) improved this to $1 < p \leq (n+2)/2$. Recently, J. Bourgain (*On the dimension of Kakeya sets and related maximal inequalities*, Geom. Funct. Anal. **9** (1999), 256–282) has shown that there exists $c > 1/2$ such that the conjecture is true for $p \leq cn$ if n is large enough.

For a discussion of recent results on this problem and its connection with the Kakeya set, see the survey article by T. Wolff (*Recent work connected with the Kakeya problem*, Prospects in Mathematics, H. Rossi, ed., pp. 129–162, Amer. Math. Soc., Providence, 1999).