Differentiability in $\mathbb{R}^n$

Christos Chatzifountas
Department of Mathematics
University of Crete

Polished English version based on the original Greek undergraduate thesis

Contents

Introduction

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. We define its average over the interval $[x-r, x+r]$ by

$$A_r f(x) = \frac{1}{2r} \int_{x-r}^{x+r} f(y)\, dy.$$

Since f is continuous, $A_r f(x)$ converges to $f(x)$ as r becomes small; more precisely,

$$\lim_{r \to 0^+} A_r f(x) = f(x) \qquad \forall x \in \mathbb{R}.$$

This property of continuous functions remains true in higher dimensions. If $f : \mathbb{R}^n \to \mathbb{R}$, then, in analogy with the one-dimensional case, we set

$$A_r f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy,$$

where we replace the interval $[x-r, x+r] \subset \mathbb{R}$ by the ball $B(x,r) \subset \mathbb{R}^n$ centered at x with radius r, and the length of the interval by Lebesgue measure on $\mathbb{R}^n$, denoted by $|\cdot|$. As in the one-dimensional case,

$$\lim_{r \to 0^+} A_r f(x) = f(x) \qquad \text{for every } x \in \mathbb{R}^n.$$

Proof. Let $\varepsilon > 0$ be arbitrary. Since f is continuous at x, there exists $\delta > 0$ such that for all r with $0 < r < \delta$ we have

$$\forall y \in B(x,r) \implies |f(x) - f(y)| < \varepsilon.$$

Therefore for every such $r > 0$,

$$|A_r f(x) - f(x)| \leq \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(x) - f(y)|\, dy \leq \varepsilon.$$

□

The continuity assumption is necessary. Indeed, if we set

$$A = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_n \geq 0\}$$

and $f(x) = \chi_A(x)$, then for every x on the boundary of A we have

$$A_r f(x) = \frac{|A \cap B(x,r)|}{|B(x,r)|} = \frac{1}{2} \neq 1 = f(x).$$

We observe that the boundary, namely the precise set on which $A_r f$ fails to converge to f, has measure zero. This leads us to ask to what extent averages can approximate "arbitrary" (not necessarily continuous) functions. In the previous example, the set of problematic points has measure zero. So perhaps, more generally, the average behaves well for almost every x. If the answer is yes, is it enough for the function merely to be integrable on every ball ($f \in L^1_{\text{loc}}$)? This is the minimal requirement needed in order for the average to be defined.

The investigation of these questions — which, as it turns out, have affirmative answers — is one of the main goals of the present thesis. In the next chapter we begin with a careful study of several averaging operators, and in particular of the operator

$$M : L^1_{\text{loc}} \to \mathbb{R}, \qquad Mf = \sup_{r>0} A_r |f|.$$

This operator is called the Hardy–Littlewood maximal function. In Chapter 2 we study its action on integrable functions by proving its fundamental property, the weak type inequality. In Chapter 3 we use this inequality to show that approximation by averages is indeed possible (the Lebesgue differentiation theorem). In the last chapter we present several properties of the maximal function which are of independent interest.

1 The Hardy–Littlewood maximal function

The Hardy–Littlewood maximal function is a sublinear operator on the space of locally Lebesgue integrable real-valued functions. It is given by

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

where $f \in L^1_{\text{loc}}(\mathbb{R}^n)$.

We shall also need two auxiliary maximal functions with similar expressions. First, the maximal function over cubes:

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

where $Q_r = [-r, r]^n$.

Second, the dyadic maximal function. For each $k \in \mathbb{Z}$ let $\mathcal{Q}_k$ be the family of all k-dyadic cubes, that is,

$$\mathcal{Q}_k = \left\{ \prod_{i=1}^{n} \left[\frac{j_i}{2^k}, \frac{j_i+1}{2^k} \right) : j_1, \dots, j_n \in \mathbb{Z} \right\}.$$

We set

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

and define

$$M_d f(x) = \sup_k |E_k f(x)|.$$

We observe that Mf and $M^0 f$ are pointwise comparable.

Proposition 1.1. *There exist positive constants c_n, C_n, depending only on the dimension n, such that*

$$Mf \le c_n M^0 f \qquad \textit{and} \qquad M^0 f \le C_n Mf.$$

Proof. Let $r > 0$. Then $B(0,r) \subset [-r, r]^n = Q_r$. Since the integrand is nonnegative,

$$\int_{B(0,r)} |f(x-y)|\, dy \le \int_{Q_r} |f(x-y)|\, dy.$$

Hence, if v_n denotes the volume of the unit ball,

$$\frac{1}{|B(0,r)|} \int_{B(0,r)} |f(x-y)|\, dy = \frac{2^n}{v_n} \frac{1}{|Q_r|} \int_{B(0,r)} |f(x-y)|\, dy \le \frac{2^n}{v_n} \frac{1}{|Q_r|} \int_{Q_r} |f(x-y)|\, dy.$$

Therefore

$$Mf(x) \le \frac{2^n}{v_n} M^0 f(x).$$

On the other hand, $B(0,r) \supset [-r/\sqrt{n}, r/\sqrt{n}]^n = Q_{r/\sqrt{n}}$, so

$$\frac{1}{|Q_{r/\sqrt{n}}|} \int_{Q_{r/\sqrt{n}}} |f(x-y)|\, dy = v_n \left(\frac{\sqrt{n}}{2} \right)^n \frac{1}{|B(0,r)|} \int_{Q_{r/\sqrt{n}}} |f(x-y)|\, dy$$

and hence

$$\frac{1}{|Q_{r/\sqrt{n}}|} \int_{Q_{r/\sqrt{n}}} |f(x-y)|\, dy \le v_n \left(\frac{\sqrt{n}}{2} \right)^n \frac{1}{|B(0,r)|} \int_{B(0,r)} |f(x-y)|\, dy.$$

Taking suprema yields

$$M^0 f(x) \le v_n \left(\frac{\sqrt{n}}{2} \right)^n Mf(x).$$

□

Proposition 1.2. *For every locally integrable f, the function Mf is measurable.*

Proof. We show that the set $A_\lambda = \{x : Mf(x) > \lambda\}$ is measurable for every real number λ. Suppose $x \in A_\lambda$. Then

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

Hence there exists $r_0 > 0$ such that

$$\frac{1}{|B(0,r_0)|} \int_{B(0,r_0)} |f(x-y)|\, dy > \lambda.$$

After a change of variables,

$$\frac{1}{|B(x,r_0)|} \int_{B(x,r_0)} |f(y)|\, dy > \lambda.$$

Choose $\varepsilon > 0$ such that

$$\frac{1}{|B(x,r_0+\varepsilon)|} \int_{B(x,r_0)} |f(y)|\, dy > \lambda.$$

Now if $x_0 \in B(x,\varepsilon)$, then $B(x_0, r_0+\varepsilon) \supset B(x,r_0)$, so

$$Mf(x_0) \geq \frac{1}{|B(x_0,r_0+\varepsilon)|} \int_{B(x_0,r_0+\varepsilon)} |f(y)|\, dy > \frac{1}{|B(x,r_0+\varepsilon)|} \int_{B(x,r_0)} |f(y)|\, dy > \lambda.$$

Thus $B(x,\varepsilon) \subset A_\lambda$, which shows that A_λ is open and therefore measurable. □

In a similar way one proves that $M^0 f$ is measurable. The measurability of $M_d f$ is an immediate consequence of the measurability of the functions $E_k f$. As for integrability, M behaves quite differently: if f is not identically zero, then there exist $R > 0$ and $\varepsilon > 0$ such that

$$\int_{B(0,R)} |f| \geq \varepsilon.$$

For every x with $|x| > R$ we have $B(0,R) \subset B(x, 2|x|)$, hence

$$Mf(x) \geq \frac{1}{|B(x,2|x|)|} \int_{B(x,2|x|)} |f| \geq \frac{1}{2^n v_n |x|^n} \int_{B(0,R)} |f| \geq \frac{\varepsilon}{2^n v_n |x|^n}.$$

Therefore

$$\int_{\mathbb{R}^n} Mf(x)\, dx \geq \frac{\varepsilon}{2^n v_n} \int_{|x|>R} \frac{dx}{|x|^n} = +\infty.$$

So Mf never belongs to L^1, unless f is identically zero.

2 The weak type inequality

As we saw above, applying the maximal operator to an L^1 function does not in general yield another integrable function. What does hold is the weaker, but fundamental, weak $(1,1)$ inequality.

Definition 2.1. Let (X, μ) and (Y, ν) be measure spaces and let T be an operator from $L^p(X, \mu)$ to the space of measurable functions from Y to $\mathbb{R}$. We say that T is of weak type (p, q), with $1 \leq p, q < \infty$, if there exists a constant $C > 0$ such that

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

for every $\lambda > 0$.

When $(X, \mu) = (Y, \nu)$ and $p = q$, the weak inequality is just the usual Chebyshev inequality. It is easy to prove that M_d is weak $(1, 1)$.

Theorem 2.2. *The dyadic maximal operator M_d is weak $(1, 1)$.*

Proof. Let $f \in L^1$. Without loss of generality we may assume $f \geq 0$. For $\lambda > 0$ define

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

and

$$\Omega_k = \{x \in \mathbb{R}^n : E_k f(x) > \lambda \text{ and } E_j f(x) \leq \lambda \text{ for } j < k\}, \qquad k \in \mathbb{Z}.$$

If $k_1 \neq k_2$, then Ω_{k_1} and Ω_{k_2} are disjoint. Indeed, suppose $k_1 > k_2$. If $x \in \Omega_{k_1}$, then $E_{k_2} f(x) \leq \lambda$, so $x \notin \Omega_{k_2}$. Similarly, if $x \in \Omega_{k_2}$, then $E_{k_1} f(x) > \lambda$, so $x \notin \Omega_{k_1}$.

Each Ω_k can be written as a union of k-dyadic cubes, and

$$\Omega = \bigcup_k \Omega_k.$$

Indeed, the inclusion $\bigcup_k \Omega_k \subset \Omega$ is obvious. Conversely, if $x \in \Omega$, let $Q_x^k \in \mathcal{Q}_k$ be the dyadic cube of generation k containing x. Then

$$E_k f(x) = \frac{1}{|Q_x^k|} \int_{Q_x^k} f.$$

Since $M_d f(x) > \lambda$, the set

$$I = \{k \in \mathbb{Z} : E_k f(x) > \lambda\}$$

is nonempty. Because $f \in L^1$, we have

$$\lim_{k \to -\infty} \frac{1}{|Q_x^k|} \int_{Q_x^k} f = 0,$$

so I is bounded below. Let $k(x) = \min I$. Then $x \in \Omega_{k(x)}$.

Therefore,

$$\begin{aligned}
|\{x \in \mathbb{R}^n : M_d f(x) > \lambda\}| = |\Omega| = \sum_k |\Omega_k| &\leq \frac{1}{\lambda} \sum_k \int_{\Omega_k} E_k f \\
&= \frac{1}{\lambda} \sum_k \sum_{Q \in \mathcal{Q}_k} \left(\frac{1}{|Q|} \int_Q f\right) |Q \cap \Omega_k| \\
&= \frac{1}{\lambda} \int_{\bigcup_k \Omega_k} f \leq \frac{1}{\lambda} \|f\|_1.
\end{aligned}$$

This is the weak $(1, 1)$ inequality with $C = 1$. ∎

The weak $(1,1)$ inequality for the Hardy–Littlewood maximal function follows immediately from the corresponding inequality for the dyadic maximal function together with the following lemma.

Lemma 2.3. *If f is integrable, then*

$$|\{x \in \mathbb{R}^n : M^0 f(x) > 4^n\lambda\}| \leq 2^n |\{x \in \mathbb{R}^n : M_d f(x) > \lambda\}|.$$

Proof. We may assume $f \geq 0$. With the notation of the previous proof,

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

where the Q_j are dyadic cubes. For each Q_j consider the cube $2Q_j$ with the same center and twice the side length. Fix

$$x_0 \notin \bigcup_j 2Q_j.$$

Let Q be a cube centered at x_0 with side length $\ell(Q)$. Then there exists an integer k such that

$$2^{-(k+1)} \leq \ell(Q) < 2^{-k}.$$

Since the side length of Q is at most 2^{-k}, it meets at most 2^n dyadic cubes of generation k, say $R_1, \ldots, R_m$. If any of these cubes were contained in some Q_j, then

$$\bigcup_i R_i \subset \bigcup_j 2Q_j \implies x_0 \in \bigcup_j 2Q_j,$$

contrary to assumption. Since two dyadic cubes are either disjoint or one is contained in the other, we conclude that $R_i \cap Q_j = \varnothing$ for every i, j. Hence

$$\frac{1}{|R_i|} \int_{R_i} f \leq \lambda$$

for each i. Therefore

$$\frac{1}{|Q|} \int_Q f = \frac{1}{|Q|} \sum_{i=1}^m \int_{Q \cap R_i} f \leq \sum_{i=1}^m \frac{|R_i|}{|Q|} \cdot \frac{1}{|R_i|} \int_{R_i} f$$
$$\leq 2^n \sum_{i=1}^m \lambda \leq 4^n \lambda,$$

where in the second inequality we used

$$\frac{|R_i|}{|Q|} \leq \frac{2^{-kn}}{2^{-(k+1)n}} = 2^n.$$

This shows that

$$M^0 f(x_0) \leq 4^n \lambda.$$

Hence

$$\{x \in \mathbb{R}^n : M^0 f(x) > 4^n\lambda\} \subset \bigcup_j 2Q_j.$$

Consequently,

$$|\{x \in \mathbb{R}^n : M^0 f(x) > 4^n\lambda\}| \leq \sum_j |2Q_j| = 2^n \sum_j |Q_j| = 2^n |\{x \in \mathbb{R}^n : M_d f(x) > \lambda\}|.$$

□

Corollary 2.4. *The Hardy–Littlewood maximal function is weak* $(1, 1)$.

Proof. From the previous chapter we know that

$$Mf \leq \frac{2^n}{v_n} M^0 f.$$

Combining this with the previous results we obtain

$$|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \leq \left|\left\{x \in \mathbb{R}^n : M^0 f(x) > \frac{v_n}{2^n}\lambda\right\}\right| \leq \frac{16^n}{v_n} \frac{\|f\|_1}{\lambda},$$

which is exactly the required estimate. □

We now give an alternative proof of the weak type inequality. Let

$$E = \{x \in \mathbb{R}^n : Mf(x) > \lambda\},$$

and let $K \subset E$ be compact. For every $x \in K$ there exists a ball B_x centered at x such that

$$\frac{1}{|B_x|} \int_{B_x} |f| > \lambda.$$

Since K is compact, there exists a finite family $\mathcal{F} \subset \{B_x : x \in K\}$ such that

$$K \subset \bigcup_{B \in \mathcal{F}} B.$$

Choose a disjoint subfamily $\mathcal{D} \subset \mathcal{F}$ as follows. Let B_1 be the largest ball in $\mathcal{F}$. Remove from $\mathcal{F}$ the ball B_1 and all balls intersecting it. Then choose B_2 as the largest remaining ball, and continue in the same way until the family is exhausted. Clearly $\mathcal{D}$ is disjoint, and moreover

$$\bigcup_{B \in \mathcal{F}} B \subset \bigcup_{B \in \mathcal{D}} 3B,$$

because at step k all balls intersecting B_k have radius no larger than that of B_k and hence are contained in $3B_k$.

Thus

$$|K| \leq \sum_{B \in \mathcal{D}} |3B| = 3^n \sum_{B \in \mathcal{D}} |B| \leq \frac{3^n}{\lambda} \sum_{B \in \mathcal{D}} \int_B |f| \leq \frac{3^n}{\lambda} \|f\|_1.$$

Since this holds for every compact subset of E, by regularity of Lebesgue measure we get

$$|E| = \sup_{\substack{K \subset E \\ K \text{ compact}}} |K| \leq \frac{3^n}{\lambda} \|f\|_1.$$

We chose to present the more indirect proof via dyadic cubes because its usefulness will become clear in the next chapter.

3 The Lebesgue differentiation theorem

We are now ready to use the weak $(1,1)$ inequality together with elementary tools from analysis to prove the Lebesgue differentiation theorem.

Theorem 3.1 (Lebesgue differentiation theorem)**.** *If f is an integrable real-valued function on $\mathbb{R}^n$, then*

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\,dy = f(x) \qquad \text{for almost every } x.$$

Equivalently, using the notation from the Introduction,

$$\lim_{r\to 0} A_r f = f \qquad \text{almost everywhere.}$$

Proof. Let $f \in L^1$. Define two auxiliary operators:

$$T_r f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|\,dy$$

and

$$Tf(x) = \limsup_{r\to 0} T_r f(x).$$

We will show that $Tf(x) = 0$ for almost every x. We know that the set of continuous functions with compact support is dense in L^1. This means that for every $k \in \mathbb{N}$ there exists $g \in C_0(\mathbb{R}^n)$ with

$$\|f - g\|_1 < \frac{1}{k}.$$

Let $h = f - g$. Then

$$T_r f \le T_r h + T_r g.$$

Hence

$$Tf \le Th + Tg.$$

Now

$$T_r g(x) = A_r(|g - g(x)|)(x).$$

Since for fixed x the function $|g - g(x)|$ is continuous, the result from the Introduction gives

$$\lim_{r\to 0} A_r(|g - g(x)|)(z) = |g(z) - g(x)| \qquad \text{for every } z.$$

Therefore

$$Tg = 0,$$

and hence

$$Tf \le Th. \tag{3.1}$$

Also,

$$\begin{aligned} T_r h(x) &= \frac{1}{|B(x,r)|} \int_{B(x,r)} |h(y) - h(x)|\,dy \\ &\le \frac{1}{|B(x,r)|} \int_{B(x,r)} |h(y)|\,dy + |h(x)| \\ &\le Mh(x) + |h(x)|. \end{aligned}$$

Thus

$$Th \leq Mh + |h|. \tag{3.2}$$

Combining (3.1) and (3.2), we obtain

$$Tf \leq Mh + |h|.$$

Let now $\lambda > 0$. If for some x we have $Tf(x) > 2\lambda$, then $Mh(x) + |h(x)| > 2\lambda$, so either $Mh(x) > \lambda$ or $|h(x)| > \lambda$. Therefore

$$\{x \in \mathbb{R}^n : Tf(x) > 2\lambda\} \subset \{x \in \mathbb{R}^n : Mh(x) > \lambda\} \cup \{x \in \mathbb{R}^n : |h(x)| > \lambda\}.$$

Hence

$$|\{x \in \mathbb{R}^n : Tf(x) > 2\lambda\}| \leq |\{x \in \mathbb{R}^n : Mh(x) > \lambda\}| + |\{x \in \mathbb{R}^n : |h(x)| > \lambda\}|.$$

The weak $(1, 1)$ inequality gives

$$|\{x \in \mathbb{R}^n : Mh(x) > \lambda\}| \leq \frac{16^n}{v_n \lambda} \|h\|_1,$$

while Chebyshev's inequality yields

$$|\{x \in \mathbb{R}^n : |h(x)| > \lambda\}| \leq \frac{1}{\lambda} \|h\|_1.$$

Therefore

$$|\{x \in \mathbb{R}^n : Tf(x) > 2\lambda\}| \leq \frac{16^n + v_n}{v_n \lambda} \|h\|_1 \leq \frac{16^n + v_n}{v_n \lambda} \cdot \frac{1}{k}.$$

Since this inequality holds for every natural number k, the left-hand side must be zero:

$$|\{x \in \mathbb{R}^n : Tf(x) > 2\lambda\}| = 0.$$

Because this holds for every $\lambda > 0$, it follows that

$$|\{x \in \mathbb{R}^n : Tf(x) > 0\}| = 0.$$

Thus $Tf = 0$ almost everywhere. The points at which $Tf = 0$ are called *Lebesgue points.* □

Remarks

1. In the differentiation theorem we assumed that f is integrable so that we could apply the weak type inequality. Nevertheless, the theorem remains valid for any locally integrable function. Indeed, if $f \in L^1_{\text{loc}}$ and we set

   $$f_k = f\chi_{B(0,k)}, \qquad k \in \mathbb{N},$$

   then $f_k \in L^1$. Hence there exists $E_k \subset \mathbb{R}^n$ with $|E_k| = 0$ such that

   $$\lim_{r \to 0} A_r f_k(x) = f_k(x) \qquad \text{for every } x \notin E_k.$$

   Therefore

   $$\lim_{r \to 0} A_r f(x) = f(x) \qquad \text{for every } x \notin \bigcup_{k=1}^{\infty} E_k.$$

2. Every $f \in L^p$, $1 \le p < \infty$, is locally integrable, because for every compact $K \subset \mathbb{R}^n$, Hölder's inequality gives

$$\int_K |f| \le |K|^{1/p'} \|f\|_p < \infty, \qquad p' = \frac{p}{p-1}.$$

Hence, by the previous remark,

$$\lim_{r \to 0} A_r f = f \qquad \text{almost everywhere.}$$

3. An obvious variation of the differentiation theorem is the following. For each $x \in \mathbb{R}^n$, let $\{U_k(x) : k \in \mathbb{N}\}$ be a family of balls or cubes such that

$$x \in U_k(x) \qquad \text{and} \qquad \lim_{k \to \infty} \operatorname{diam}(U_k(x)) = 0.$$

Then for every $f \in L^1_{\mathrm{loc}}$ we have

$$\lim_{k \to \infty} \frac{1}{|U_k(x)|} \int_{U_k(x)} f(y)\, dy = f(x) \qquad \text{for almost every } x.$$

4. Another variation is the following: if $f \in L^p$, $1 \le p < \infty$, then

$$\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|^p\, dy = 0 \qquad \text{for almost every } x.$$

Indeed, for each $t \in \mathbb{Q}$ there exists a null set $E_t \subset \mathbb{R}^n$ such that

$$\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - t|^p\, dy = |f(x) - t|^p \qquad \text{for every } x \notin E_t.$$

Now fix

$$x \notin \bigcup_{t \in \mathbb{Q}} E_t$$

and let $\varepsilon > 0$. Choose $t_0 \in \mathbb{Q}$ such that $|f(x) - t_0| < \varepsilon$. Minkowski's inequality gives

$$\left(\frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|^p\, dy \right)^{1/p} \le \left(\frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - t_0|^p\, dy \right)^{1/p} + |f(x) - t_0|.$$

Taking $\limsup$ as $r \to 0$, we obtain

$$\limsup_{r \to 0} \left(\frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|^p\, dy \right)^{1/p} \le 2|f(x) - t_0| < 2\varepsilon,$$

and the conclusion follows.

5. If $A \subset \mathbb{R}^n$ is a Lebesgue measurable set and we apply the differentiation theorem to the function χ_A, we obtain

$$\lim_{r \to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1 \qquad \text{for almost every } x \in A,$$

and

$$\lim_{r\to 0} \frac{|A\cap B(x,r)|}{|B(x,r)|} = 0 \qquad \text{for almost every } x \notin A.$$

This result is usually called the *Lebesgue density theorem.* Intuitively, it says that, from a measure-theoretic point of view, a measurable set looks like an interval at sufficiently small scales.

A very important application of the differentiation theorem in analysis is the so-called Calderón–Zygmund decomposition. The proof uses ideas developed in the previous chapter.

Theorem 3.2 (Calderón–Zygmund decomposition at height λ)**.** *Let f be a nonnegative integrable function, and let $\lambda > 0$. Then there exists a sequence $\{Q_j\}$ of pairwise disjoint dyadic cubes such that:*

1.

$$\left|\bigcup_j Q_j\right| \le \frac{1}{\lambda}\|f\|_1;$$

2.

$$\lambda < \frac{1}{|Q_j|}\int_{Q_j} f \le 2^n\lambda \qquad \textit{for every } j;$$

3.

$$f(x) \le \lambda \qquad \textit{for almost every } x \notin \bigcup_j Q_j.$$

Proof. Consider the sets Ω_k and the corresponding dyadic cubes Q_j from the previous chapter. The first inequality is exactly the weak $(1,1)$ estimate for M_d. If $Q_j \subset \Omega_k$ is a k-dyadic cube and $\widetilde{Q}_j$ is the unique $(k-1)$-dyadic cube containing it, then by definition of Ω_k,

$$\lambda < \frac{1}{|Q_j|}\int_{Q_j} f, \qquad \frac{1}{|\widetilde{Q}_j|}\int_{\widetilde{Q}_j} f \le \lambda.$$

But $|\widetilde{Q}_j| = 2^n|Q_j|$, and this gives the second estimate.

Finally, if $x \notin \bigcup_j Q_j$ is a Lebesgue point of f, and $\{R_i\}$ is a decreasing sequence of dyadic cubes with $x \in R_i$, then by definition of the sets Ω_k we have

$$\frac{1}{|R_i|}\int_{R_i} f \le \lambda.$$

Passing to the limit as $i \to \infty$ yields the last statement. □

The next chapter presents a standard application of this decomposition.

4 The maximal function on Lebesgue spaces

In the previous chapters we saw the important role played by the weak type $(1,1)$ inequality. This inequality reflects the way the maximal operator acts on L^1 functions. We now examine additional properties, beginning with its action on L^p functions. For the next theorem we will need the following proposition.

Proposition 4.1. *For every real-valued function $f \in L^p$,*

$$\|f\|_p^p = p\int_0^\infty \lambda^{p-1}|\{|f| > \lambda\}|\,d\lambda.$$

Proof. Let $f \in L^p$. By the fundamental theorem of calculus,

$$\|f\|_p^p = \int_{\mathbb{R}^n} |f(x)|^p\,dx = \int_{\mathbb{R}^n}\int_0^{|f(x)|} p\lambda^{p-1}\,d\lambda\,dx.$$

Changing the order of integration gives

$$\begin{aligned}\int_{\mathbb{R}^n}\int_0^{|f(x)|} p\lambda^{p-1}\,d\lambda\,dx &= \int_0^\infty\int_{\mathbb{R}^n} p\lambda^{p-1}\chi_{\{|f|>\lambda\}}(x)\,dx\,d\lambda \\ &= p\int_0^\infty \lambda^{p-1}|\{x\in\mathbb{R}^n : |f(x)| > \lambda\}|\,d\lambda.\end{aligned}$$

□

Theorem 4.2. *The Hardy–Littlewood maximal function is a bounded operator from L^p to L^p for every $1 < p \leq \infty$.*

Proof. Let $f \in L^p$. Decompose f as

$$f_0 = f\,\chi_{\{|f|>\lambda/2\}}, \qquad f_1 = f\,\chi_{\{|f|\leq\lambda/2\}}.$$

Clearly $f = f_0 + f_1$. Moreover, the set on which $|f| > \lambda/2$ has finite measure, otherwise $\|f\|_p = +\infty$. By Hölder's inequality,

$$\|f_0\|_1 = \int_{\mathbb{R}^n} |f|\,\chi_{\{|f|>\lambda/2\}} \leq |\{|f| > \lambda/2\}|^{1/q}\|f\|_p < +\infty,$$

where q is the conjugate exponent of p. Thus $f_0 \in L^1$, while obviously $f_1 \in L^\infty$. By sublinearity,

$$Mf \leq Mf_0 + Mf_1.$$

Hence

$$|\{|Mf| > \lambda\}| \leq |\{|Mf_0| > \lambda/2\}| + |\{|Mf_1| > \lambda/2\}| = |\{|Mf_0| > \lambda/2\}|,$$

because $\|f_1\|_\infty \leq \lambda/2$.

Applying the weak $(1,1)$ inequality to M and using the previous proposition, we get

$$\begin{aligned}\|Mf\|_p^p &= p\int_0^\infty \lambda^{p-1}|\{|Mf| > \lambda\}|\,d\lambda \\ &\leq p\int_0^\infty \lambda^{p-1}|\{|Mf_0| > \lambda/2\}|\,d\lambda \\ &\leq pC_n\int_0^\infty \lambda^{p-2}\|f_0\|_1\,d\lambda \\ &= pC_n\int_0^\infty \lambda^{p-2}\left(\int_{\mathbb{R}^n} |f(x)|\chi_{\{|f|>\lambda/2\}}(x)\,dx\right)d\lambda \\ &= pC_n\int_{\mathbb{R}^n} |f(x)|\left(\int_0^{2|f(x)|}\lambda^{p-2}\,d\lambda\right)dx \\ &= \frac{2^p pC_n}{p-1}\|f\|_p^p.\end{aligned}$$

□

The constant appearing in the previous $L^p \to L^p$ estimate depends on the weak type constant, which in turn depends exponentially on the dimension n if one inspects the proofs in Chapter 2. We now give a proof that improves this dependence from exponential to linear. For that purpose we introduce the following auxiliary functions.

Let $h \in L^1_{\text{loc}}(\mathbb{R})$. For $x \in \mathbb{R}$ define the one-dimensional Hardy–Littlewood maximal function by

$$M_1 h(x) = \sup_{r>0} \frac{1}{2r} \int_{-r}^{r} |h(x-t)|\, dt.$$

If $x \in \mathbb{R}^n$, $\xi \in S^{n-1}$, and $f \in C_0(\mathbb{R}^n)$, define the directional Hardy–Littlewood maximal function in the direction ξ by

$$M_\xi f(x) = \sup_{r>0} \frac{1}{r} \int_0^r |f(x-t\xi)|\, dt.$$

Finally, given $f : \mathbb{R}^n \to \mathbb{R}$ and $x' \in \mathbb{R}^{n-1}$, define $f_{x'} : \mathbb{R} \to \mathbb{R}$ by

$$f_{x'}(t) = f(t, x').$$

Let now $f \in L^p$ for $1 < p < \infty$ and $\xi \in S^{n-1}$. Then there exists a rotation $A_\xi : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$A_\xi(\xi) = e_1 = (1, 0, \dots, 0).$$

That is, A_ξ sends ξ to the first coordinate vector. Observe that

$$\begin{aligned} M_\xi f(A_\xi^{-1} x) &= \sup_{r>0} \frac{1}{r} \int_0^r |f(A_\xi^{-1}(x - te_1))|\, dt \\ &\le 2 M_1 (f \circ A_\xi^{-1})_{x'}(x_1), \end{aligned}$$

where x_1 is the first coordinate of x. Since Lebesgue measure is rotation-invariant,

$$\begin{aligned} \int_{\mathbb{R}^n} (M_\xi f(x))^p\, dx &= \int_{\mathbb{R}^n} \left(M_\xi f(A_\xi^{-1} x)\right)^p\, dx \\ &\le 2^p \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \left(M_1 (f \circ A_\xi^{-1})_{x'}(x_1)\right)^p\, dx_1\, dx'. \end{aligned}$$

For the one-dimensional Hardy–Littlewood maximal function we already know that

$$\|M_1 g\|_p \le C_p \|g\|_p.$$

Combining this with the previous estimate gives

$$\int_{\mathbb{R}^n} (M_\xi f(x))^p\, dx \le 2^p C_p^p \|f\|_p^p.$$

Now, changing to polar coordinates,

$$\begin{aligned} \frac{1}{|B(0,r)|} \int_{B(0,r)} |f(x-y)|\, dy &= \frac{1}{v_n r^n} \int_0^r \int_{S^{n-1}} t^{n-1} |f(x-t\xi)|\, d\sigma(\xi)\, dt \\ &\le \frac{1}{v_n} \int_{S^{n-1}} M_\xi f(x)\, d\sigma(\xi). \end{aligned}$$

Therefore,

$$Mf(x) \le \frac{1}{v_n} \int_{S^{n-1}} M_\xi f(x)\, d\sigma(\xi).$$

Using Minkowski's integral inequality, we get

$$\|Mf\|_p \le \frac{1}{v_n} \int_{S^{n-1}} \|M_\xi f\|_p \, d\sigma(\xi) \le \frac{2C_p}{v_n} \sigma(S^{n-1}) \|f\|_p.$$

Since

$$v_n = \frac{1}{n} \sigma(S^{n-1}),$$

we conclude that

$$\|Mf\|_p \le 2nC_p \|f\|_p.$$

This is the desired estimate.

As we have seen, the theorem above fails when $p = 1$. Nevertheless, one can still estimate the L^1 norm of Mf if some restriction is imposed on f. The following theorem, after the next lemma, has precisely this purpose.

Lemma 4.3. *There exist positive constants C_1, C_2, and c such that for every $f \in L^1$ and every $\lambda > 0$:*

1. $$|\{x : Mf(x) > 2\lambda\}| \le \frac{C_1}{\lambda} \int_{\{|f|>\lambda\}} |f(x)|\, dx;$$

2. $$|\{x : Mf(x) > c\lambda\}| \ge \frac{C_2}{\lambda} \int_{\{|f|>\lambda\}} |f(x)|\, dx.$$

Proof. 1. Set

$$f_1 = f\, \chi_{\{|f|>\lambda\}}.$$

Then

$$\{x : Mf(x) > 2\lambda\} \subset \{x : Mf_1(x) > \lambda\},$$

and the weak $(1, 1)$ inequality gives

$$|\{Mf > 2\lambda\}| \le |\{Mf_1 > \lambda\}| \le \frac{C_1}{\lambda} \int |f_1(x)|\, dx = \frac{C_1}{\lambda} \int_{\{|f|>\lambda\}} |f(x)|\, dx.$$

2. Let $\{Q_j\}$ be a Calderón–Zygmund decomposition of $|f|$ at height λ. For each $x \in Q_j$ let

$$B_j = B(x, \ell(Q_j)\sqrt{n}).$$

Then $Q_j \subset B_j$, hence

$$\lambda < \frac{1}{|Q_j|} \int_{Q_j} |f(x)|\, dx \le \frac{|B_j|}{|Q_j|} \cdot \frac{1}{|B_j|} \int_{B_j} |f(x)|\, dx \le \frac{1}{c} Mf(x),$$

where $1/c = |B_j|/|Q_j|$. This holds for all cubes Q_j, so

$$\bigcup_j Q_j \subset \{x \in \mathbb{R}^n : Mf(x) > c\lambda\}.$$

Hence

$$\left|\bigcup_j Q_j\right| \leq |\{x \in \mathbb{R}^n : Mf(x) > c\lambda\}|.$$

From the properties of the decomposition we know that $|f(x)| \leq \lambda$ for almost every $x \notin \bigcup_j Q_j$. Therefore

$$\begin{aligned}\int_{\{|f|>\lambda\}} |f(x)|\,dx &\leq \int_{\bigcup_j Q_j} |f(x)|\,dx = \sum_j \int_{Q_j} |f(x)|\,dx \\ &\leq 2^n\lambda \sum_j |Q_j| = 2^n\lambda \left|\bigcup_j Q_j\right| \\ &\leq 2^n\lambda|\{x \in \mathbb{R}^n : Mf(x) > c\lambda\}|.\end{aligned}$$

□

The previous lemma says that, in a certain sense, the weak type inequality can be inverted. We are now in a position to prove the following theorem.

Theorem 4.4. *Let $B \subset \mathbb{R}^n$ be a ball, and let $f \in L^1$ be a function on $\mathbb{R}^n$ with support contained in B. Then*

$$Mf \in L^1(B) \quad \textit{if and only if} \quad \int_B |f| \log^+ |f| < +\infty,$$

where

$$\log^+ |f| = \max\{0, \log |f|\}.$$

Proof. For simplicity we denote all constants by C.

Assume first that

$$\int_B |f| \log^+ |f| < +\infty.$$

Then

$$\int_B Mf(x)\,dx = 2\int_0^\infty |\{x \in B : Mf(x) > 2\lambda\}|\,d\lambda.$$

Now

$$\int_0^1 |\{x \in B : Mf(x) > 2\lambda\}|\,d\lambda \leq |B|,$$

while by the previous lemma,

$$\int_1^\infty |\{x \in B : Mf(x) > 2\lambda\}|\,d\lambda \leq C\int_1^\infty \frac{1}{\lambda}\left(\int_{\{x\in B:|f(x)|>\lambda\}} |f(x)|\,dx\right)d\lambda.$$

Hence

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

Conversely, assume that

$$\int_B Mf(x)\,dx < +\infty.$$

We claim that

$$\int_{\frac{3}{2}B} Mf(x)\,dx < +\infty.$$

Indeed, if $B = B(x_0, r_0)$, then for $x \in \frac{3}{2}B \setminus B$ and $r > |x - x_0| - r_0$ we have

$$B(x, r) \subset B(x_0', 3r),$$

where x_0' is the point symmetric to x with respect to the boundary of B. Since the support of f is contained in B, it follows that

$$Mf(x) \leq CMf(x_0').$$

Using polar coordinates, we conclude that

$$\int_{\frac{3}{2}B\setminus B} Mf(x)\,dx \leq C \int_{B\setminus\frac{1}{2}B} Mf(x)\,dx < +\infty.$$

Thus $Mf \in L^1(\frac{3}{2}B)$.

Repeating the same argument with $\frac{3}{2}B$ in place of B, and then inductively, we conclude that for every $k \in \mathbb{N}$,

$$\int_{(3/2)^k B} Mf(x)\,dx < +\infty.$$

That is, Mf is locally integrable.

We also observe that

$$Mf(x) \to 0 \qquad \text{as } |x| \to +\infty.$$

Indeed, let $\varepsilon > 0$. Choose $R > 0$ such that $B \subset B(0, R)$ and $1/R < \varepsilon$. Then for every x with $|x| > 2R$,

$$Mf(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f| = \sup_{r>R} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f| \leq C\frac{\|f\|_1}{R^n} < C\|f\|_1\varepsilon^n.$$

Since ε is arbitrary,

$$\lim_{|x|\to\infty} Mf(x) = 0.$$

Hence there exists a sufficiently large ball B_0 such that

$$\{x \in \mathbb{R}^n : Mf(x) > C\} \subset B_0.$$

Therefore

$$\int_{\{x\in\mathbb{R}^n : Mf(x)>C\}} Mf(x)\,dx \leq \int_{B_0} Mf(x)\,dx < +\infty,$$

because Mf is locally integrable. Applying the previous lemma, we obtain

$$\begin{aligned}
\infty &> \int_{\{x\in\mathbb{R}^n : Mf(x)>C\}} Mf(x)\,dx \\
&\geq \int_C^\infty |\{x \in \mathbb{R}^n : Mf(x) > \lambda\}|\,d\lambda \\
&\geq C\int_1^\infty \frac{1}{\lambda}\left(\int_{\{x\in\mathbb{R}^n : |f(x)|>\lambda\}} |f(x)|\,dx\right) d\lambda \\
&= C\int_B |f| \log^+ |f|.
\end{aligned}$$

Thus

$$\int_B |f| \log^+ |f| < +\infty.$$

□

The space of functions satisfying the condition in the previous theorem is called the $L \log L$ *space*. It lies between L^1 and all the L^p spaces with $p > 1$.

Final remarks

The maximal operator for $n = 1$ was introduced by Hardy and Littlewood [3], who also gave the first proof of the weak type inequality. The proof we presented via dyadic cubes rests on ideas of Calderón and Zygmund [1], to whom the decomposition bearing their names is also due.

The alternative proof at the end of Chapter 2 was given by Wiener [6].

Theorem 4.1 is a special case of the Marcinkiewicz interpolation theorem, which allows one to pass from weak type inequalities to L^p inequalities.

The technique by which we improved the constant in Theorem 4.1 is called the *method of rotations* and is due to Calderón and Zygmund [2]. In fact, the constant can be shown [5] not to depend on the dimension n. It remains an open problem whether the same is true for the weak type constant.

Theorem 4.2 is called the $L \log L$ *theorem*. The forward implication is due to Stein [4], and the converse to Wiener [6].

References

[1] A. Calderón and A. Zygmund, *On the existence of certain singular integrals*, Acta Math. 88 (1952), 85–139.

[2] A. Calderón and A. Zygmund, *On singular integrals*, Amer. J. Math. 18 (1956), 289–309.

[3] G. H. Hardy and J. E. Littlewood, *A maximal theorem with function-theoretic applications*, Acta Math. 54 (1930), 81–116.

[4] E. M. Stein, *Note on the class* $L \log L$, Studia Math. 32 (1969), 305–310.

[5] E. M. Stein and J. O. Strömberg, *Behavior of maximal functions in* $\mathbb{R}^n$ *for large* n, Ark. Mat. 21 (1983), 259–269.

[6] N. Wiener, *The ergodic theorem*, Duke Math. J. 5 (1939), 1–18.