Singular Integral Operators

Damian Dąbrowski

December 1, 2023

Abstract

Notes for the course *Singular Integral Operators* lectured at the University of Jyväskylä in Autumn 2023.

Contents

1 Introduction

This course will focus on singular integral operators, which are operators of the form

$$Tf(x) = \int K(x,y)f(y)\,dy,$$

where the kernel $K(x,y)$ has a singularity on the diagonal $x = y$. These operators appear naturally e.g. in the theory of partial differential equations, and they have been studied for over a century. The prototypical example is the Hilbert transform

$$Hf(x) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\,dy.$$

The basic questions we will study concern the mapping properties of singular integral operators: for which $1 \le p \le \infty$ and under what hypotheses on the kernel K is the operator T bounded on L^p, in the sense that

$$\|Tf\|_{L^p} \le C\|f\|_{L^p}.$$

The material we will cover reflects both the long tradition of this field, and the fact that it is still an active area of research. We will begin by studying the Hilbert and Riesz transforms, which date back almost 100 years back. Then, we will move on to the Calderón-Zygmund theory, which revolutionized the field in the 1950s. Finally, we will discuss singular integrals in the weighted setting, which is a much more recent topic. The grand finale will be the proof of the A_2 theorem, which was shown by Tuomas Hytönen in 2012 [Hyt12]. We will follow a short and elegant proof from [Ler16] which uses a cutting-edge technique called *sparse domination.*

The field of singular integral operators is huge, and we will only scratch the surface in this course. We refer interested readers to the textbooks [Duo01, Gra14a, Gra14b, Ste70, Ste93] for more thorough treatments of the subject.

We claim no originality for any of the proofs. While preparing these lecture notes we used the books mentioned above, as well as [Con13, Ler16, LN19, Par20].

2 Preliminaries

Before getting started in earnest, we recall briefly some useful facts and definitions. For proofs and details, see e.g. Chapters 1 and 2 of [Gra14a].

In these notes we sometimes use the notation $A \lesssim B$, which stands for "there exists a dimensional constant $C \geq 1$ such that $A \leq CB$." We write $A \sim B$ instead of $A \lesssim B \lesssim A$.

2.1 Schwartz functions and tempered distributions

Definition 2.1 (Schwartz functions). A function $f \in C^\infty(\mathbb{R}^n)$ is a *Schwartz function*, denoted by $f \in \mathcal{S}(\mathbb{R}^n)$, if for every pair of multi-indices $\alpha = (\alpha_1, \dots, \alpha_n)$, $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{N}^n$ we have

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

We will say that a function decays rapidly if it decays at ∞ faster than any polynomial. Hence, Schwartz functions are precisely those $C^\infty(\mathbb{R}^n)$ functions which decay rapidly and whose all partial derivatives decay rapidly.

Example 2.2. Any smooth and compactly supported function is a Schwartz function, so that $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$. A simple example of a non-compactly supported Schwartz function is $e^{-|x|^2}$.

One of the reasons Schwartz functions are useful is the following density result.

Lemma 2.3. *The Schwartz functions are dense in $L^p(\mathbb{R}^n)$ for all $1 \leq p < \infty$.*

Note that $\mathcal{S}(\mathbb{R}^n)$ is a vector space. A topology on $\mathcal{S}(\mathbb{R}^n)$ can be defined using the family of semi-norms $\rho_{\alpha,\beta}$, and it is compatible with the following notion of convergence.

Definition 2.4 (convergence in $\mathcal{S}(\mathbb{R}^n)$). Given $f \in \mathcal{S}(\mathbb{R}^n)$ and a sequence $f_k \in \mathcal{S}(\mathbb{R}^n)$, we say that f_k coverges to f in $\mathcal{S}(\mathbb{R}^n)$ if for all multi-indices $\alpha, \beta \in \mathbb{N}^n$

$$\lim_{k \to \infty} \rho_{\alpha,\beta}(f_k - f) = 0.$$

Definition 2.5 (tempered distributions). We denote by $\mathcal{S}'(\mathbb{R}^n)$ the dual space of $\mathcal{S}(\mathbb{R}^n)$, i.e., the space of all continuous linear functionals $T : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$. The elements of $\mathcal{S}'(\mathbb{R}^n)$ are called *tempered distributions.*

Given $T \in \mathcal{S}'(\mathbb{R}^n)$ and $f \in \mathcal{S}(\mathbb{R}^n)$, instead of writing $T(f)$ we will write $\langle T, f \rangle$, and we will call it *the action of T on f*.

We have the following useful characterization of tempered distributions:

Lemma 2.6. *A linear functional $T : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ is a tempered distribution if and only if there exist $m, k \in \mathbb{N}$ and $C > 0$ such that for all $f \in \mathcal{S}(\mathbb{R}^n)$*

$$|\langle T, f\rangle| \leq C \sum_{|\alpha|\leq m,\ |\beta|\leq k} \rho_{\alpha,\beta}(f).$$

Example 2.7. Any function $g \in L^p(\mathbb{R}^n), 1 \leq p \leq \infty$, gives rise to a tempered distribution $T_g \in \mathcal{S}'(\mathbb{R}^n)$ defined via $\langle T_g, f\rangle = \int f(x)g(x)\ dx$.

Example 2.8. Any finite Borel measure μ gives rise to a tempered distribution $T_\mu \in \mathcal{S}'(\mathbb{R}^n)$ defined via $\langle T_\mu, f\rangle = \int f\ d\mu$.

In the case of tempered distributions as above, we will often identify T_g with g, and T_μ with μ. For example, the statement "$T \in \mathcal{S}'(\mathbb{R}^n)$ is a $C^\infty(\mathbb{R}^n)$ function" should be understood as "there exists $f \in C^\infty(\mathbb{R}^n)$ such that $T = T_f$." The Hilbert transform we will define shortly will provide us with an example of a tempered distribution which is neither a locally integrable function, nor a measure.

Many common operations performed on functions can be extended by duality to tempered distributions. For example, given $h \in \mathcal{S}(\mathbb{R}^n)$ and $T \in \mathcal{S}'(\mathbb{R}^n)$, we define their convolution as a tempered distribution $T * h \in \mathcal{S}'(\mathbb{R}^n)$ given by

$$\langle T * h, f\rangle \coloneqq \langle T, \tilde{h} * f\rangle,$$

where $\tilde{h}(x) = h(-x)$. Similarly, the product of $h \in \mathcal{S}(\mathbb{R}^n)$ and $T \in \mathcal{S}'(\mathbb{R}^n)$ can be defined as a tempered distribution $hT \in \mathcal{S}'(\mathbb{R}^n)$ given by

$$\langle hT, f\rangle \coloneqq \langle T, hf\rangle.$$

Proposition 2.9. *Given $h \in \mathcal{S}(\mathbb{R}^n)$ and $T \in \mathcal{S}'(\mathbb{R}^n)$ the convolution $T * h$ belongs to $C^\infty(\mathbb{R}^n)$. Moreover,*

$$T * h(x) = \langle T, h(x - \cdot)\rangle.$$

2.2 Fourier transform

Definition 2.10. The Fourier transform of $f \in \mathcal{S}(\mathbb{R}^n)$ is defined as

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

Sometimes we will denote it by $\mathcal{F}(f)$ instead of $\hat{f}$.

The Fourier transform is a homeomorphism of $\mathcal{S}(\mathbb{R}^n)$ to itself, and its inverse is given by

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

sometimes denoted by $\mathcal{F}^{-1}(f)$.

The Plancherel identity asserts that for any $f \in \mathcal{S}(\mathbb{R}^n)$

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

By the density of $\mathcal{S}(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$, this allows us to extend the Fourier transform to an isometry of $L^2(\mathbb{R}^n)$.

One may further extend the definition of Fourier transform to all tempered distributions using duality: for any $T \in \mathcal{S}'(\mathbb{R}^n)$ we define $\widehat{T} \in \mathcal{S}'(\mathbb{R}^n)$ via

$$\langle \widehat{T}, f\rangle := \langle T, \hat{f}\rangle.$$

We list a few properties of the Fourier transform we will use later on.

Lemma 2.11. *If $f \in \mathcal{S}(\mathbb{R}^n)$ and $T \in \mathcal{S}'(\mathbb{R}^n)$, then*

(i) $\mathcal{F}(\partial^\alpha f) = (2\pi i \xi)^\alpha \hat{f}$,

(ii) $\partial^\alpha \hat{f} = \mathcal{F}((-2\pi i x)^\alpha f)$,

(iii) $\widehat{T * f} = \hat{T}\hat{f}$.

2.3 Weak and strong type inequalities

In this subsection we assume that (X, μ) and (Y, ν) are two measure spaces.

Definition 2.12. Given $1 \leq p, q \leq \infty$ and an operator T mapping functions from a dense subset of $L^p(X, \mu)$ to measurable functions on (Y, ν), we say that T *is of strong type* (p, q) if there exists $C > 0$ such that

$$\|Tf\|_{L^q(Y,\nu)} \leq C\|f\|_{L^p(X,\mu)}.$$

We say that T *is of weak type* (p, q) if there exists $C > 0$ such that for all $\lambda > 0$

$$\nu(\{y \in Y \ : \ |Tf(y)| > \lambda\}) \leq C\left(\frac{\|f\|_{L^p(X,\mu)}}{\lambda}\right)^q.$$

It is easy to see that strong type (p, q) implies weak type (p, q).

Definition 2.13 (sublinear operator). An operator T defined on a linear space of measurable functions on (X, μ) and taking values in measurable functions on (Y, ν) is sub-linear if

$$|T(f + g)| \leq |Tf| + |Tg| \quad \text{and} \quad |T(\lambda f)| = |\lambda||Tf|.$$

The Marcinkiewicz interpolation theorem stated below plays a crucial role in the theory of singular integral operators.

Theorem 2.14. *Let $1 \leq p_0 < p_1 \leq \infty$. Suppose that T is a sub-linear operator mapping $L^{p_0}(X, \mu) + L^{p_1}(X, \mu)$ to the set of measurable functions on (Y, ν). If T is of weak type (p_0, p_0) and (p_1, p_1), then it is of strong type (p, p) for all $p_0 < p < p_1$.*

3 The Hilbert and Riesz transforms

In this section we will study the prototypical singular integral operator, the Hilbert transform, as well as its higher dimensional counterparts, the Riesz transforms. The Hilbert transform arises naturally e.g. in the study of boundary values of analytic functions, in questions regarding the convergence of Fourier transform, or in signal processing. While we will not study these applications, they may be chosen as a presentation topic to pass the course.

3.1 The Hilbert transform on $\mathcal{S}(\mathbb{R})$

The Hilbert transform is the singular integral operator associated with kernel $K(x,y) = \frac{1}{\pi(x-y)}$. We begin by defining it for Schwartz functions.

As a first attempt at defining it, one could try to simply integrate against the kernel:

$$\frac{1}{\pi}\int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy.$$

However, the expression above is highly problematic. Even for a very nice function f, say, $f \in C_c^\infty(\mathbb{R})$, it is easy to see that as soon as $f(x) \neq 0$, the integral above is not well-defined! This is because $(x-y)^{-1}$ has a singularity at x which is not integrable.

To avoid this issue, we first consider the following *truncated Hilbert transform*.

Definition 3.1 (truncated Hilbert transform). For $f \in \mathcal{S}(\mathbb{R})$ and $\varepsilon > 0$, we define the truncated Hilbert transform of f as

$$H_\varepsilon f(x) \coloneqq \frac{1}{\pi}\int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\, dy = \frac{1}{\pi}\int_{|y|>\varepsilon} \frac{f(x-y)}{y}\, dy.$$

Note that, by the rapid decay of Schwartz functions, $H_\varepsilon f(x)$ is well-defined for every $x \in \mathbb{R}$.

Definition 3.2 (Hilbert transform). For $f \in \mathcal{S}(\mathbb{R})$, we define the Hilbert transform of f as

$$Hf(x) \coloneqq \lim_{\varepsilon\to 0} H_\varepsilon f(x) = \lim_{\varepsilon\to 0} \frac{1}{\pi}\int_{|y|>\varepsilon} \frac{f(x-y)}{y}\, dy.$$

Clearly, for $x \notin \operatorname{supp} f$ this is well-defined, and in fact

$$Hf(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{f(y)}{x-y}\, dy \qquad \text{for } x \notin \operatorname{supp} f. \tag{3.1}$$

Let us show that $Hf(x)$ is well-defined also for $x \in \operatorname{supp} f$.

Lemma 3.3. *For any $f \in \mathcal{S}(\mathbb{R})$ and $x \in \mathbb{R}$ the limit $\lim_{\varepsilon\to 0} H_\varepsilon f(x)$ exists, and we have*

$$Hf(x) = \frac{1}{\pi}\int_{-1}^{1} \frac{f(x-y)-f(x)}{y}\, dy + \frac{1}{\pi}\int_{|y|>1} \frac{f(x-y)}{y}\, dy. \tag{3.2}$$

Proof. Fix $\varepsilon > 0$. Note that, since the kernel $\frac{1}{y}$ is odd, it has zero mean on any symmetric pair of intervals around the origin, and in particular

$$\int_{\varepsilon<|y|<1} \frac{1}{y}\, dy = 0.$$

It follows that

$$\int_{|y|>\varepsilon} \frac{f(x-y)}{y}\, dy = \int_{\varepsilon<|y|<1} \frac{f(x-y)-f(x)}{y}\, dy + \int_{|y|>1} \frac{f(x-y)}{y}\, dy.$$

The second integral on the right hand side is just a constant that does not depend on ε. Concerning the first integral, observe that by the mean value theorem the integrand is uniformly bounded

$$\left|\frac{f(x-y)-f(x)}{y}\right| \le \|f'\|_{L^\infty(\mathbb{R})},$$

and so the limit exists and we have

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

□

We showed that the Hilbert transform is a well-defined, linear operator defined on $\mathcal{S}(\mathbb{R})$. Later on, we will be interested in extending it to the L^p spaces for $1 < p < \infty$. One way to do that is by showing that H is of strong type (p,p), i.e. that for all $f \in \mathcal{S}(\mathbb{R})$ we have

$$\|Hf\|_{L^p(\mathbb{R})} \le C_p \|f\|_{L^p(\mathbb{R})}.$$

After establishing such inequality, we may use the density of $\mathcal{S}(\mathbb{R})$ in $L^p(\mathbb{R})$ to extend the Hilbert transform to functions in $L^p(\mathbb{R})$. The exercise below shows that we may only hope for the strong type (p,p) inequality to hold for $1 < p < \infty$.

Exercise 3.4 (1 point). Let $f = \mathbf{1}_{[0,1]}$. Show that for $x \in \mathbb{R}\setminus\{0,1\}$

$$\lim_{\varepsilon\to 0}\int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\, dy = \log\left|\frac{x}{x-1}\right|.$$

Conclude that the Hilbert transform is neither of strong type (∞,∞) nor of strong type $(1,1)$.

So our goal is estimating $\|Hf\|_{L^p(\mathbb{R})}$. As a warm-up, we prove that for $f \in \mathcal{S}(\mathbb{R})$ we have $Hf \in L^p(\mathbb{R})$ for all $1 < p \leq \infty$. This is a consequence of the following asymptotic identity.

Lemma 3.5. *For $f \in \mathcal{S}(\mathbb{R})$ we have*

$$\lim_{|x|\to\infty} x \cdot Hf(x) = \frac{1}{\pi} \int_{\mathbb{R}} f(y)\, dy.$$

Proof. The proof is similar to that of (3.2). We use the oddness of kernel $\frac{1}{y}$ once again to get that for any $x \in \mathbb{R}$ with $|x| > 0$

$$\begin{aligned}
\pi x \cdot Hf(x) &= \lim_{\varepsilon\to 0} x \int_{|y|>\varepsilon} \frac{f(x-y)}{y}\, dy \\
&= \lim_{\varepsilon\to 0} x \int_{\varepsilon<|y|<\frac{|x|}{2}} \frac{f(x-y)-f(x)}{y}\, dy + x \int_{\frac{|x|}{2}<|y|<2|x|} \frac{f(x-y)}{y}\, dy \\
&\qquad + x \int_{|y|>2|x|} \frac{f(x-y)}{y}\, dy = I_1 + I_2 + I_3.
\end{aligned}$$

Regarding I_1, note that for $|y| < |x|/2$ we have $|x|/2 \leq |x-y| \leq 3|x|/2$, and so by the mean value theorem

$$|I_1| \leq |x|^2 \sup_{|x|/2\leq|\xi|\leq 3|x|/2} |f'(\xi)| \sim \sup_{|x|/2\leq|\xi|\leq 3|x|/2} |\xi^2 f'(\xi)| \xrightarrow{|x|\to\infty} 0,$$

where in the last step we used the rapid decay of Schwartz functions.

Concerning I_3, we have $|x-y| \geq |x|$ whenever $|y| > 2|x|$, and so

$$|I_3| \leq |x| \int_{|y|>2|x|} \frac{|f(x-y)|}{2|x|}\, dy \leq \int_{|z|>|x|} |f(z)|\, dz \xrightarrow{|x|\to\infty} 0,$$

since f is integrable.

Finally,

$$I_2 - \int f(x-y)\, dy = \int_{\frac{|x|}{2}<|y|<2|x|} \left(\frac{x}{y} - 1\right) f(x-y)\, dy - \int_{|y|<|x|/2, \text{ or } |y|>2|x|} f(x-y)\, dy,$$

which gives

$$\begin{aligned}
\Big|I_2 - \int f(x-y)\, dy\Big| &\leq \int_{\frac{|x|}{2}<|y|<2|x|} \left|\frac{x-y}{y}\right| |f(x-y)|\, dy + \int_{|y|<|x|/2, \text{ or } |y|>2|x|} |f(x-y)|\, dy \\
&\lesssim \frac{1}{|x|} \int |zf(z)|\, dy + \int_{|z|>|x|/2} |f(z)|\, dy \xrightarrow{|x|\to\infty} 0.
\end{aligned}$$

□

Corollary 3.6. *For every $f \in \mathcal{S}(\mathbb{R})$ we have $Hf \in L^p(\mathbb{R})$ for all $1 < p \leq \infty$.*

Proof. Note that by (3.2) and the mean value theorem we have

$$\|Hf\|_{L^\infty(\mathbb{R})} \lesssim \|f'\|_{L^\infty(\mathbb{R})} + \sup_{x \in \mathbb{R}} |x \cdot f(x)|, \tag{3.3}$$

so the Hilbert transform of a Schwartz function is bounded. Thus, whether $Hf \in L^p$ for $1 \leq p < \infty$ depends only on the decay rate of Hf at infinity. By Lemma 3.5, for $|x|$ large enough we have $|Hf(x)| \lesssim_f x^{-1}$, and it follows that $Hf \in L^p(\mathbb{R})$ for all $p > 1$. □

Exercise 3.7. Let $f \in \mathcal{S}(\mathbb{R})$. Show that $Hf \in L^1(\mathbb{R})$ if and only if $\int_{\mathbb{R}} f(y)\, dy = 0$. A hint: Modify the proof of Lemma 3.5 to estimate the asymptotics of $x^2 \cdot Hf(x)$ as $|x| \to \infty$.

3.2 The Hilbert transform on $L^2(\mathbb{R})$

In this subsection we extend the Hilbert transform to $L^2(\mathbb{R})$. We begin by computing the Fourier transform of Hf.

First, since for any $f \in \mathcal{S}(\mathbb{R})$ we have $Hf \in L^2(\mathbb{R})$ by Corollary 3.6, the Fourier transform $\widehat{Hf}$ is well-defined as a function in L^2. Below we compute its precise value.

Proposition 3.8. *For any $f \in \mathcal{S}(\mathbb{R})$ we have*

$$\widehat{Hf}(\xi) = -i \operatorname{sgn}(\xi) \hat{f}(\xi) \quad \textit{for a.e. } \xi \in \mathbb{R}. \tag{3.4}$$

To prove this, we start by taking a slightly more abstract point of view. Since the Hilbert transform is linear, and we have the estimate (3.3), we can define a tempered distribution $T_0 \in \mathcal{S}'(\mathbb{R})$ by

$$\langle T_0, f \rangle := -Hf(0) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_{|y| > \varepsilon} \frac{f(y)}{y}\, dy.$$

Note that

$$Hf(x) = \langle T_0, f(x - \cdot) \rangle = T_0 * f(x).$$

Taking the Fourier transform (in the sense of distributions), we see that

$$\widehat{Hf} = \widehat{T_0} \cdot \hat{f}, \tag{3.5}$$

where the product is also understood in the sense of distributions: for any $\varphi \in \mathcal{S}(\mathbb{R})$ we have $\langle \widehat{Hf}, \varphi \rangle = \langle \widehat{T_0}, \hat{f}\varphi \rangle$.

As a consequence of (3.5), to prove (3.4) it suffices to show that $\widehat{T_0}$, which *a priori* is just a tempered distribution, is in fact a function, and that $\widehat{T_0}(\xi) = -i \operatorname{sgn}(\xi)$.

Lemma 3.9. *We have* $\widehat{T_0}(\xi) = -i\,\mathrm{sgn}(\xi)$.

Proof. An exercise. Some hints:

(i) Let $K_\varepsilon(y) = \frac{1}{y}\mathbf{1}_{|y|>\varepsilon}$, so that $\langle T_0, f\rangle = \frac{1}{\pi}\lim_{\varepsilon\to 0}\langle K_\varepsilon, f\rangle$, and consider $Q_\varepsilon(y) = \frac{y}{y^2+\varepsilon^2}$. Show that

$$\lim_{\varepsilon\to 0}(K_\varepsilon - Q_\varepsilon) = 0 \quad \text{in } \mathcal{S}'(\mathbb{R}).$$

(ii) Using the above, argue that $\widehat{T_0} = \frac{1}{\pi}\lim_{\varepsilon\to 0}\widehat{Q_\varepsilon}$, in the sense of distributions.

(iii) Show that $Q_\varepsilon(x) = \mathcal{F}^{-1}(-\pi i\,\mathrm{sgn}(\xi)e^{-2\pi\varepsilon|\xi|})(x)$. Conclude that $\widehat{T_0}$ is given by a function, and that $\widehat{T_0}(\xi) = -i\,\mathrm{sgn}(\xi)$.

□

As a corollary of Proposition 3.8 and Plancherel's identity, we can define the Hilbert transform of functions in $L^2(\mathbb{R})$.

Corollary 3.10. *For any* $f \in \mathcal{S}(\mathbb{R})$ *we have*

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

Consequently, the Hilbert transform extends to an isometry of $L^2(\mathbb{R})$. *Moreover, for any* $f \in L^2(\mathbb{R})$ *its Hilbert transform satisfies*

$$\widehat{Hf}(\xi) = -i\,\mathrm{sgn}(\xi)\hat{f}(\xi).$$

Recall that for $f \in \mathcal{S}(\mathbb{R})$ we have a nice formula for $Hf(x)$ assuming $x \notin \mathrm{supp}\, f$, see (3.1). It is easy to see that the same formula holds for $f \in L^2(\mathbb{R})$.

Exercise 3.11. Show that if $f \in L^2(\mathbb{R})$, then for a.e. $x \notin \mathrm{supp}(f)$

$$Hf(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{f(y)}{x-y}\,dy.$$

Here, $\mathrm{supp}\, f$ denotes the essential support of f.

3.2.1 Truncated Hilbert transform

In Definition 3.1 we introduced the truncated Hilbert transform

$$H_\varepsilon f(x) = \int_{|x-y|>\varepsilon} \frac{f(y)}{x-y}\,dy$$

for $f \in \mathcal{S}(\mathbb{R})$. However, the same definition makes sense for $f \in L^p(\mathbb{R})$ for all $1 \leq p < \infty$ To see that, we use Hölder's inequality to show that the integral defining $H_\varepsilon f$ converges absolutely:

$$\int_{|x-y|>\varepsilon} \left| \frac{f(y)}{x-y} \right| \, dy \leq \|f\|_{L^p} \left\| \frac{\mathbf{1}_{|x-y|>\varepsilon}}{x-y} \right\|_{L^q} < \infty,$$

where $1/p + 1/q = 1$, so that $1 < q \leq \infty$.

By the definition of Hilbert transform, we have $Hf(x) = \lim_{\varepsilon \to 0} H_\varepsilon f(x)$ for all $f \in \mathcal{S}(\mathbb{R})$ and $x \in \mathbb{R}$. It is natural to ask for a counterpart of this statement for $f \in L^2(\mathbb{R})$; for example, do we have $H_\varepsilon f \to Hf$ in L^2 sense? We are able to show this if we assume that all truncated Hilbert transforms are of strong type $(2, 2)$, in a uniform way.

Proposition 3.12. *Suppose that there exists a constant $C > 0$ such that*

$$\sup_{\varepsilon>0} \|H_\varepsilon f\|_{L^2(\mathbb{R})} \leq C\|f\|_{L^2(\mathbb{R})} \qquad \textit{for all } f \in L^2. \tag{3.6}$$

Then, for every $f \in L^2(\mathbb{R})$ we have $H_\varepsilon f \to Hf$ in L^2.

Proof. Let $f_n \in \mathcal{S}(\mathbb{R})$ be such that $f_n \to f$ in L^2. Then, $Hf_n \to Hf$ in L^2, and we have

$$\|H_\varepsilon f - Hf\|_{L^2} \leq \|H_\varepsilon f - H_\varepsilon f_n\|_{L^2} + \|H_\varepsilon f_n - Hf_n\|_{L^2} + \|Hf_n - Hf\|_{L^2} =: I_1 + I_2 + I_3.$$

The term I_3 converges to 0 because $Hf_n \to H_f$ in L^2, whereas I_1 converges to 0 because

$$\|H_\varepsilon f - H_\varepsilon f_n\|_{L^2} = \|H_\varepsilon(f - f_n)\|_{L^2} \overset{(3.6)}{\leq} C\|f - f_n\|_{L^2} \xrightarrow{n\to\infty} 0.$$

It remains to estimate $I_2 = \|H_\varepsilon f_n - Hf_n\|_{L^2}$. By (3.2) we have

$$|H_\varepsilon f_n(x) - Hf_n(x)| = \left| \frac{1}{\pi} \int_{-\varepsilon}^{\varepsilon} \frac{f_n(x-y) - f_n(x)}{y} \, dy \right| \leq \frac{2\varepsilon}{\pi} \sup_{z \in (x-\varepsilon, x+\varepsilon)} |f_n'(z)|.$$

Set $g_n(x) \coloneqq \sup_{z \in (x-\varepsilon, x+\varepsilon)} |f_n'(z)|$. Since f_n' decays rapidly, we get that g_n also decays rapidly, and so

$$I_2 = \|H_\varepsilon f_n - Hf_n\|_{L^2} \leq \frac{2\varepsilon}{\pi} \|g_n\|_{L^2}.$$

Hence, for any $\delta > 0$ we may take n large enough such that $I_1 + I_3 \leq \delta$, and then $\varepsilon > 0$ small enough so that $I_2 \leq \delta$. Then, we have $\|H_\varepsilon f - Hf\|_{L^2} \leq 2\delta$, and taking $\delta \to 0$ concludes the proof. $\square$

The question remains, how to show the estimate (3.6)? We will address this later on when we prove the so-called *Cotlar's inequality* for general singular integral operators.

3.3 The Riesz transform

Before moving on to general singular integral operators and their L^p-theory, we briefly discuss another important family of operators, the Riesz transforms. They are higher dimensional counterparts of the Hilbert transform.

Definition 3.13. For $f \in \mathcal{S}(\mathbb{R}^n)$ and $1 \leq j \leq n$, we define the j-th Riesz transform of f as

$$R_j f(x) := \lim_{\varepsilon \to 0} C_n \int_{|x-y|>\varepsilon} \frac{x_j - y_j}{|x-y|^{n+1}} f(y)\, dy,$$

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

As in the case of the Hilbert transform, there is a simple formula for the Fourier transform of $R_j f$.

Proposition 3.14. *For any $f \in \mathcal{S}(\mathbb{R}^n)$*

$$\widehat{R_j f}(\xi) = -i \frac{\xi_j}{|\xi|} \hat{f}(\xi). \tag{3.7}$$

The proof is similar to that of Proposition 3.4, although there are additional difficulties. The interested reader can find the full proof e.g. in [Gra14a, Proposition 5.1.14].

As an immediate corollary of (3.7), we get that for all $f \in \mathcal{S}(\mathbb{R}^n)$

$$\|R_j f\|_{L^2(\mathbb{R}^n)} \leq \|f\|_{L^2(\mathbb{R}^n)}, \tag{3.8}$$

and we may extend the Riesz transforms to $L^2(\mathbb{R}^n)$.

Finally, we give a simple application of (3.8), which also motivates the study of L^p-bounds for the Riesz transforms for $1 < p < \infty$.

Proposition 3.15. *For $f \in \mathcal{S}(\mathbb{R}^n)$ and $1 \leq j, k \leq n$ we have*

$$\partial_j \partial_k f = -R_j R_k \Delta f. \tag{3.9}$$

In consequence, for any $1 < p < \infty$ such that the bound $\|R_j f\|_{L^p(\mathbb{R}^n)} \leq C_p \|f\|_{L^p(\mathbb{R}^n)}$ holds for all $1 \leq j \leq n$, we have

$$\|\partial_j \partial_k f\|_{L^p(\mathbb{R}^n)} \leq (C_p)^2 \|\Delta f\|_{L^p(\mathbb{R}^n)}. \tag{3.10}$$

Proof. By taking the Fourier transform of $\partial_j \partial_k f$ we get

$$\begin{aligned}
\mathcal{F}(\partial_j \partial_k f)(\xi) &= (2\pi i \xi_j)(2\pi i \xi_k) \hat{f}(\xi) \\
&= -\left(-i\frac{\xi_j}{|\xi|}\right)\left(-i\frac{\xi_k}{|\xi|}\right)(-4\pi^2|\xi|^2)\hat{f}(\xi) \\
&= -\mathcal{F}(R_j R_k \Delta f)(\xi).
\end{aligned}$$

Taking the inverse Fourier transform finishes the proof of identity (3.9). The estimate (3.10) follows immediately. $\square$

4 Calderón-Zygmund theory

In this section we begin the study of general singular integral operators.

4.1 Standard kernels and Calderón-Zygmund operators

The operators we will consider will be associated to the following kernels.

Definition 4.1 (standard kernel). We say that a Borel function $K : \mathbb{R}^n \times \mathbb{R}^n \setminus \{(x,x) : x \in \mathbb{R}^n\} \to \mathbb{C}$ is a *standard kernel* if there exists $\delta > 0$ and $C > 0$ such that

$$|K(x,y)| \leq \frac{C}{|x-y|^n}, \tag{4.1}$$

$$|K(x,y) - K(x,y')| \leq C \frac{|y-y'|^\delta}{|x-y|^{n+\delta}} \qquad \text{if } |x-y| > 2|y-y'|, \tag{4.2}$$

$$|K(x,y) - K(x',y)| \leq C \frac{|x-x'|^\delta}{|x-y|^{n+\delta}} \qquad \text{if } |x-y| > 2|x-x'|. \tag{4.3}$$

The bound (4.1) will be referred to as the *size condition*, while the other two estimates will be called the *smoothness conditions*.

Remark 4.2. The estimate $|x-y| > 2|y-y'|$ appearing in the smoothness condition can be interpreted in the following way: it is the estimate ensuring that $\frac{1}{2}|x-y| \leq |x-y'| \leq 2|x-y|$ (this follows easily from the triangle inequality).

We give a few examples.

Example 4.3. The Hilbert transform kernel $K(x,y) = \frac{1}{x-y}$ is a standard kernel on $\mathbb{R}$. More generally, the kernels $K(x,y) = \frac{x_j - y_j}{|x-y|^{n+1}}$ associated to the Riesz transforms are standard kernels on $\mathbb{R}^n$.

Example 4.4. Given $f \in C_c^\infty(\mathbb{R}^2)$ the solution to the Poisson equation $\Delta u = -2\pi f$ is given by the logarithmic potential of f

$$u(x) = \int_{\mathbb{R}^2} f(y) \log\left(\frac{1}{|x-y|}\right) \, dy.$$

It can be shown that the mixed partial derivative $\partial_{x_1}\partial_{x_2} u$ is given by the singular integral operator

$$\partial_{x_1}\partial_{x_2} u(x) = \lim_{\varepsilon \to 0} \int_{|x-y|>\varepsilon} \frac{\Omega_0\left(\frac{x-y}{|x-y|}\right)}{|x-y|^2} f(y) \, dy,$$

where $\Omega_0(x) = \frac{2x_1x_2}{|x|^2}$, see [CZ52, p. 130]. By the exercise below, the kernel associated to Ω_0 is a standard kernel.

Exercise 4.5. Show that for every Hölder continuous $\Omega : \mathbb{S}^{n-1} \to \mathbb{C}$ the kernel defined by

$$K(x,y) = \frac{\Omega\left(\frac{x-y}{|x-y|}\right)}{|x-y|^n}$$

is a standard kernel on $\mathbb{R}^n$.

Example 4.6. The kernel

$$K(z,w) = \frac{1}{(z-w)^2} \qquad z,w \in \mathbb{C},$$

is a standard kernel. It is associated to the *Beurling-Ahlfors transform*, which plays a fundamental role in the theory of quasiconformal mappings, see [Ast94].

The three examples above are kernels of convolution type, in the sense that $K(x,y) = K_0(x-y)$ for some $K_0 : \mathbb{R}^n \setminus \{0\} \to \mathbb{C}$. The next example shows that there are interesting kernels of non-convolution type, which justifies developing the theory in this generality.

Example 4.7 (Cauchy integral along a Lipschitz graph). Let $A : \mathbb{R} \to \mathbb{R}$ be a Lipschitz function, and let $\Gamma = \{(t, A(t)) : t \in \mathbb{R}\} \subset \mathbb{C}$. Given $f \in \mathcal{S}(\mathbb{R})$ let $F : \Gamma \to \mathbb{C}$ be given by $F(t+iA(t)) = f(t)$. The Cauchy integral of f is defined as

$$C_\Gamma f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{F(w)}{w-z}\,dw = \frac{1}{2\pi i}\int_\mathbb{R} \frac{f(t)(1+iA'(t))}{t+iA(t)-z}\,dt,$$

and it defines an analytic function on $\mathbb{C}\setminus\Gamma$. One can compute the boundary values of $C_\Gamma f(z)$ on Γ:

$$\begin{aligned}\lim_{\varepsilon\to 0} C_\Gamma f(x+i(A(x)+\varepsilon)) &= \frac{1}{2\pi i}\lim_{\varepsilon\to 0}\int_{|x-t|>\varepsilon} \frac{f(t)(1+iA'(t))}{t-x+i(A(t)-A(x))}\,dt + \frac{1}{2}f(x)\\ \lim_{\varepsilon\to 0} C_\Gamma f(x+i(A(x)-\varepsilon)) &= \frac{1}{2\pi i}\lim_{\varepsilon\to 0}\int_{|x-t|>\varepsilon} \frac{f(t)(1+iA'(t))}{t-x+i(A(t)-A(x))}\,dt - \frac{1}{2}f(x),\end{aligned}$$

see [Gra14b, Chapter 4.6]. This leads to the study of the Cauchy transform

$$Tf(x) = \lim_{\varepsilon\to 0}\int_{|x-y|>\varepsilon} \frac{f(y)}{x-y+i(A(x)-A(y))}\,dy,$$

whose kernel

$$K(x,y) = \frac{1}{x-y+i(A(x)-A(y))} \tag{4.4}$$

is a standard kernel of non-convolution type. For more information and the history of the Cauchy transform see [Tol14, Ver21].

Exercise 4.8. Prove that if A is Lipschitz, then the Cauchy kernel (4.4) is standard with $\delta = 1$.

We are ready to define our main object of study in this course: the Calderón-Zygmund operators.

Definition 4.9 (Calderón-Zygmund operator). We say that a linear operator $T : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a *Calderón-Zygmund operator* if

(i) T is of strong type $(2, 2)$,

(ii) there exists a standard kernel K such that for all $f \in L^2(\mathbb{R}^n)$ with compact support

$$Tf(x) = \int K(x,y)f(y)\, dy \qquad \text{for } x \notin \operatorname{supp} f. \tag{4.5}$$

Whenever (4.5) holds, we will say that T *is associated to the kernel* K.

We make a few clarifying remarks regarding the definition of Calderón-Zygmund operators.

Remark 4.10. We stress that the definition of a Calderón-Zygmund operator *assumes* that the operator is bounded on L^2. We already know that this is true for the Hilbert transform and the Riesz transforms, and so they are Calderón-Zygmund operators (the property (ii) was shown in Exercise 3.11).

While the other operators mentioned in Examples 4.4, 4.6, 4.7 are also bounded on L^2, in general it is far from obvious. For example, proving the L^2-boundedness of the Cauchy transform on Lipschitz graphs was a major open problem for decades, and it was only solved in [CMM82]. We will not cover this result.

There are some sufficient conditions on kernels K that imply the L^2-boundedness of associated operators, see [Duo01, Chapter 4]. This may be a topic for a presentation.

The following exercise shows that a Calderón-Zygmund operator uniquely determines its kernel.

Exercise 4.11. If T is a Calderón-Zygmund operator such that (4.5) holds with two kernels K_1 and K_2, then $K_1 = K_2$ a.e.

The converse is not true. The trivial kernel $K = 0$ is associated both with the zero operator $T = 0$ and with the identity operator $T = I$. In general, for any $b \in L^\infty(\mathbb{R}^n)$ the pointwise multiplication operator

$$Tf(x) = b(x)f(x)$$

is a Calderón-Zygmund operator associated with the kernel $K = 0$. However, this is the only ambiguity.

Lemma 4.12. *Suppose that T_1 and T_2 are two Calderón-Zygmund operators associated with the same kernel K. Then, there exists $b \in L^\infty(\mathbb{R}^n)$ such that*

$$T_1 f = T_2 f + bf.$$

Proof. Let $T = T_1 - T_2$, so that T is a Calderón-Zygmund operator associated with the kernel $K = 0$. Our aim is to show that $Tf = bf$ for some $b \in L^\infty$. We will only prove this identity for characteristic functions, the case of general $f \in L^2$ follows by the density of simple functions in L^2.

First, we claim that for all measurable sets $E, F \subset \mathbb{R}^n$ with $0 < |E|, |F| < \infty$ we have $T(\mathbf{1}_E) = \mathbf{1}_E T(\mathbf{1}_E)$ and

$$\mathbf{1}_F T(\mathbf{1}_E) = T(\mathbf{1}_{E\cap F}). \tag{4.6}$$

Indeed, we have $T(\mathbf{1}_E)(x) = 0$ for a.e. $x \notin E$, since T is associated to $K = 0$. This gives $T(\mathbf{1}_E) = \mathbf{1}_E T(\mathbf{1}_E)$, and also it shows that $\mathbf{1}_F T(\mathbf{1}_E) = \mathbf{1}_{E\cap F} T(\mathbf{1}_E)$. By linearity of T,

$$\begin{aligned}\mathbf{1}_{E\cap F} T(\mathbf{1}_E) &= \mathbf{1}_{E\cap F} T(\mathbf{1}_{E\cap F}) + \mathbf{1}_{E\cap F} T(\mathbf{1}_{E\setminus F}) \\ &= \mathbf{1}_{E\cap F} T(\mathbf{1}_{E\cap F}) + \mathbf{1}_{E\cap F}\mathbf{1}_{E\setminus F} T(\mathbf{1}_{E\setminus F}) = \mathbf{1}_{E\cap F} T(\mathbf{1}_{E\cap F}) + 0.\end{aligned}$$

This gives (4.6).

Formally, we would like to define $b = T1$, but since $1 \notin L^2$, we have to work a bit to make this rigorous. Let $\{Q\}_{Q\in\mathcal{Q}}$ be a family of closed unit cubes tiling $\mathbb{R}^n$. Let $b_Q = T(\mathbf{1}_Q)$. Note that $\operatorname{supp} b_Q = \operatorname{supp} T(\mathbf{1}_Q) \subset Q$.

By the Lebesgue differentiation theorem, for a.e. $x \in \mathbb{R}^n$ we have

$$|b_Q(x)| = \lim_{r\to 0} \frac{\left|\int_{B(x,r)} b_Q \, dy\right|}{|B(x,r)|}. \tag{4.7}$$

We use the Cauchy-Schwarz inequality and the L^2-boundedness of T to get

$$\begin{aligned}\left|\int_{B(x,r)} b_Q \, dy\right| &= \left|\int_{B(x,r)} \mathbf{1}_{B(x,r)} T(\mathbf{1}_Q) \, dy\right| \overset{(4.6)}{=} \left|\int_{B(x,r)} T(\mathbf{1}_{Q\cap B(x,r)}) \, dy\right| \\ &\le |B(x,r)|^{1/2} \|T(\mathbf{1}_{Q\cap B(x,r)})\|_{L^2} \le C|B(x,r)|^{1/2}|Q\cap B(x,r)|^{1/2}.\end{aligned}$$

Together with (4.7) this gives $|b_Q(x)| \le C$ for a.e. $x \in \mathbb{R}^n$, so that $b_Q \in L^\infty$. Recalling that $\operatorname{supp} b_Q \subset Q$, we get that

$$b := \sum_{Q\in\mathcal{Q}} b_Q \in L^\infty.$$

We claim that for any bounded measurable $E \subset \mathbb{R}^n$ with $0 < |E| < \infty$ we have $T\mathbf{1}_E = b\mathbf{1}_E$. Indeed, such E intersects only a finite number of $Q \in \mathcal{Q}$, and then

$$T(\mathbf{1}_E) = \sum_{Q \in \mathcal{Q}} T(\mathbf{1}_{E \cap Q}) \overset{(4.6)}{=} \mathbf{1}_E \sum_{Q \in \mathcal{Q}} T(\mathbf{1}_Q) = \mathbf{1}_E b.$$

□

Our goal is to prove the following fundamental result due to Calderón and Zygmund.

Theorem 4.13. *Suppose that T is a Calderón-Zygmund operator. Then, T is of weak type $(1,1)$, and of strong type (p,p) for $1 < p < \infty$.*

We will prove it over the following two subsections.

4.2 Calderón-Zygmund decomposition

Definition 4.14. The family of *dyadic cubes* in $\mathbb{R}^n$, denoted by $\mathcal{D}(\mathbb{R}^n)$, is defined as

$$\mathcal{D}(\mathbb{R}^n) = \left\{ 2^{-k}(m + [0,1)^n) = \prod_{i=1}^{n} [2^{-k}m_i, 2^{-k}m_i + 2^{-k}) \ : \ m \in \mathbb{Z}^n, k \in \mathbb{Z} \right\}.$$

Given $Q \in \mathcal{D}(\mathbb{R}^n)$, we will denote its sidelength by $\ell(Q)$. We set

$$\mathcal{D}_k(\mathbb{R}^n) = \{Q \in \mathcal{D}(\mathbb{R}^n) : \ell(Q) = 2^{-k}\}.$$

When the ambient space $\mathbb{R}^n$ is clear from context, we will write $\mathcal{D}$ instead of $\mathcal{D}(\mathbb{R}^n)$. Note that in our definition dyadic cubes are half-open, half-closed, so that for a fixed $k \in \mathbb{Z}$ the family $\mathcal{D}_k(\mathbb{R}^n)$ consists of pairwise-disjoint cubes, and it is a partition of $\mathbb{R}^n$.

We point out several important properties of the dyadic cubes:

(i) For any $Q, P \in \mathcal{D}(\mathbb{R}^n)$ we have either $Q \cap P = \varnothing$, or $Q \subset P$, or $P \subset Q$.

(ii) For every $Q \in \mathcal{D}_k(\mathbb{R}^n)$ there is a unique $\widehat{Q} \in \mathcal{D}_{k-1}(\mathbb{R}^n)$ such that $Q \subset \widehat{Q}$. We will call $\widehat{Q}$ *the parent of* Q.

(iii) Every $Q \in \mathcal{D}_k(\mathbb{R}^n)$ contains exactly 2^n cubes from $\mathcal{D}_{k+1}(\mathbb{R}^n)$. We will call these cubes *the children of* Q.

These properties endow $\mathcal{D}(\mathbb{R}^n)$ with a natural tree structure based on the parent-child relation.

The following is the main result of this subsection, and it is crucial for the proof of Theorem 4.13.

Proposition 4.15. *Let $f \in L^1(\mathbb{R}^n)$ and $\alpha > 0$. There exists a decomposition of f of the form*

$$f = g + \sum_{Q \in \mathcal{B}} b_Q,$$

where $\mathcal{B}$ is a collection of disjoint dyadic cubes, and which satisfies the following:

(i) the "good part" g satisfies $\|g\|_{L^1} \le \|f\|_{L^1}$ and $\|g\|_{L^\infty} \le 2^n \alpha$,

(ii) each "bad function" b_Q is supported on $\overline{Q}$, satisfies $\int_Q b_Q = 0$, and

$$\|b_Q\|_{L^1} \le 2^{n+1} \alpha |Q|, \tag{4.8}$$

(iii) for each $Q \in \mathcal{B}$ we have

$$\alpha \le \frac{1}{|Q|} \int_Q |f| \le 2^n \alpha, \tag{4.9}$$

(iv) we can estimate the total measure of cubes in $\mathcal{B}$ by

$$\sum_{Q \in \mathcal{B}} |Q| \le \frac{\|f\|_{L^1}}{\alpha}.$$

Proof. We will say that a cube $Q \in \mathcal{D}$ is *bad* if

$$\frac{1}{|Q|} \int_Q |f| > \alpha.$$

A bad cube Q is called maximal if there in no other bad cube Q' such that $Q \subsetneq Q'$.

We claim that every bad cube is contained in some maximal bad cube. If that was not true, then there would be a sequence of bad cubes $Q_1, Q_2, \dots$ such that $\ell(Q_k) \to \infty$. At the same time,

$$\alpha < \frac{1}{|Q_k|} \int_{Q_k} |f| \le \frac{\|f\|_{L^1}}{|Q_k|} \xrightarrow{k \to \infty} 0,$$

which is a contradiction.

Let $\mathcal{B}$ be the family of maximal bad cubes. Since they are dyadic and maximal, they are disjoint. For every $Q \in \mathcal{B}$ we define

$$b_Q := \left(f - \frac{1}{|Q|} \int_Q f \right) \mathbf{1}_Q,$$

and

$$g := f - \sum_{Q \in \mathcal{B}} b_Q.$$

We begin by proving (iii). The lower bound in (4.9) is just the definition of bad cubes. The upper bound follows from maximality: for every $Q \in \mathcal{B}$ its parent $\widehat{Q}$ is not a bad cube, and so

$$\frac{1}{|Q|}\int_Q |f| \le \frac{|\widehat{Q}|}{|Q|}\frac{1}{|\widehat{Q}|}\int_{\widehat{Q}} |f| \le 2^n\alpha.$$

Concerning (ii), the first two properties follow immediately from the definition, and

$$\|b_Q\|_{L^1} \le \int_Q |f|dx + \int_Q \left|\frac{1}{|Q|}\int_Q f dx\right| dy \le 2\int_Q |f|dx \overset{(4.9)}{\le} 2^{n+1}\alpha|Q|.$$

We move on to (i). Note that

$$g(x) = \begin{cases} f(x) & \text{for } x \notin \bigcup_{Q\in\mathcal{B}} Q \\ \frac{1}{|Q|}\int_Q f & \text{for } x \in Q \in \mathcal{B}. \end{cases}$$

Hence,

$$\begin{aligned} \|g\|_{L^1} = \int_{\mathbb{R}^n\setminus\bigcup_{Q\in\mathcal{B}} Q} |f|dx + \sum_{Q\in\mathcal{B}}\int_Q \left|\frac{1}{|Q|}\int_Q f dy\right| dx \\ = \int_{\mathbb{R}^n\setminus\bigcup_{Q\in\mathcal{B}} Q} |f|dx + \sum_{Q\in\mathcal{B}}\left|\int_Q f dy\right| \le \|f\|_{L^1}. \end{aligned}$$

To see $\|g\|_{L^\infty} \le 2^n\alpha$, note that $|g(x)| \le 2^n\alpha$ for $x \in Q \in \mathcal{B}$ by (4.9). Let $x \notin \bigcup_{Q\in\mathcal{B}} Q$, so that $g(x) = f(x)$. Then, for all dyadic cubes containing x we have $\frac{1}{|Q|}\int_Q |f|dx \le \alpha$. By the (dyadic version of) Lebesgue differentiation theorem for a.e. $y \in \mathbb{R}^n$ we have

$$|f(y)| = \lim_{\ell(Q)\to 0,\, y\in Q\in\mathcal{D}} \frac{1}{|Q|}\int_Q |f|dz$$

Together with the estimate $\frac{1}{|Q|}\int_Q |f| \le \alpha$, this shows that for a.e. $x \notin \bigcup_{Q\in\mathcal{B}} Q$ we have $|g(x)| = |f(x)| \le \alpha < 2^n\alpha$.

Finally, we show (iv). By the definition of bad cubes,

$$\sum_{Q\in\mathcal{B}} |Q| \le \sum_{Q\in\mathcal{B}} \frac{\int_Q |f|}{\alpha} \le \frac{\|f\|_{L^1}}{\alpha},$$

where in the last inequality we also used that the cubes in $\mathcal{B}$ are disjoint. $\square$

4.3 The L^p theory for Calderón-Zygmund operators

In this subsection we prove Theorem 4.13, whose statement we repeat below.

Theorem. *Suppose that T is a Calderón-Zygmund operator. Then, T is of weak type $(1,1)$, and of strong type (p,p) for $1 < p < \infty$.*

We begin by reducing matters to the weak type (1,1) estimate.

Weak type (1,1) implies strong type (p,p). Fix a Calderón-Zygmund operator T. By definition, it is of strong type (2,2). Hence, as soon as we know that it is of weak type (1,1), it follows from the Marcinkiewicz interpolation theorem (Theorem 2.14) that T is of strong type (p,p) for all $1 < p < 2$. To get the same for $2 < p < \infty$ we argue by duality as follows.

Given a Calderón-Zygmund operator $T : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ we consider its adjoint operator $T^t : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ defined by

$$\langle T^t(g), f\rangle_{L^2} = \int T^t(g) \cdot \bar{f}\, dx = \int g \cdot \overline{T(f)}\, dx = \langle g, T(f)\rangle_{L^2}.$$

This is well-defined by the Riesz representation theorem.

Exercise 4.16. Prove that if a Calderón-Zygmund operator T is associated to a standard kernel K, then its adjoint is also a Calderón-Zygmund operator, and it is associated to the standard kernel

$$K^t(x,y) = \overline{K(y,x)}.$$

Since T^t is a Calderón-Zygmund operator, it follows by the argument above that T^t is of strong type (q,q) for all $1 < q < 2$. Fix $f \in \mathcal{S}(\mathbb{R}^n)$, $2 < p < \infty$, and let $1 < q < 2$ be such that $1/p + 1/q = 1$. Then, using that the dual of $L^p(\mathbb{R}^n)$ is $L^q(\mathbb{R}^n)$, we get

$$\begin{aligned}\|Tf\|_{L^p} &= \sup_{g\in\mathcal{S},\, \|g\|_{L^q}\le 1} \left|\int T(f)\cdot \overline{g}\right| = \sup_{g\in\mathcal{S},\, \|g\|_{L^q}\le 1} \left|\int f \cdot \overline{T^t(g)}\right| \\ &\le \|f\|_{L^p} \sup_{g\in\mathcal{S},\, \|g\|_{L^q}\le 1} \|T^t(g)\|_{L^q} \le C_q \|f\|_{L^p} \sup_{g\in\mathcal{S},\, \|g\|_{L^q}\le 1} \|g\|_{L^q} = C_q\|f\|_{L^p}.\end{aligned}$$

Hence, T is of strong type (p,p). □

Proof of the weak type (1,1) estimate. Let[1] $f \in L^1 \cap L^2$. Our goal is to show that there exists a dimensional constant C such that for any $\alpha > 0$

$$|\{x \in \mathbb{R}^n \;:\; |Tf(x)| > \alpha\}| \le C\frac{\|f\|_{L^1}}{\alpha}.$$

[1] We only assume $f \in L^2$ so that Tf is well-defined, our estimates will be independent of $\|f\|_{L^2}$.

We apply the Calderón-Zygmund decomposition (Proposition 4.15) to f at level α, so that $f = g + b = g + \sum_{Q\in\mathcal{B}} b_Q$. By the linearity of T, we have $Tf = Tg + Tb$, and so

$$\begin{aligned} &|\{x \in \mathbb{R}^n \ : \ |Tf(x)| > \alpha\}| \\ &\qquad \le |\{x \in \mathbb{R}^n \ : \ |Tg(x)| > \alpha/2\}| + |\{x \in \mathbb{R}^n \ : \ |Tb(x)| > \alpha/2\}|. \end{aligned} \tag{4.10}$$

To estimate the term corresponding to g, we use Chebyshev's inequality and the fact that $\|g\|_{L^1} \le \|f\|_{L^1}$, $\|g\|_{L^\infty} \le 2^n\alpha$:

$$\begin{aligned} |\{x \in \mathbb{R}^n \ : \ |Tg(x)| > \alpha/2\}| &\le \frac{\|Tg\|_{L^2}^2}{(\alpha/2)^2} \lesssim \frac{\|g\|_{L^2}^2}{\alpha^2} \\ &\le \frac{\|g\|_{L^1}\|g\|_{L^\infty}}{\alpha^2} \le \frac{2^n\alpha\|f\|_{L^1}}{\alpha^2} \sim \frac{\|f\|_{L^1}}{\alpha}. \end{aligned} \tag{4.11}$$

So the first term from the RHS of (4.10) satisfies the desired inequality. We move on to the second term, which is more difficult to estimate.

For every $Q \in \mathcal{B}$ let Q^* be the cube with the same center as Q, and with sidelength $\ell(Q^*) = 2\sqrt{n}\,\ell(Q)$. We have

$$|\{x \in \mathbb{R}^n \ : \ |Tb(x)| > \alpha/2\}| \le \sum_{Q\in\mathcal{B}} |Q^*| + |\{x \in \mathbb{R}^n \setminus \bigcup_{Q\in\mathcal{B}} Q^* \ : \ |Tb(x)| > \alpha/2\}|.$$

The first term satisfies

$$\sum_{Q\in\mathcal{B}} |Q^*| \lesssim \sum_{Q\in\mathcal{B}} |Q| \le \frac{\|f\|_{L^1}}{\alpha}$$

by Proposition 4.15 (iv). Concerning the second term, by Chebyshev's inequality

$$\begin{aligned} |\{x \in \mathbb{R}^n \setminus \bigcup_{Q\in\mathcal{B}} Q^* \ : \ |Tb(x)| > \alpha/2\}| &\le \frac{2}{\alpha}\int_{(\bigcup_{Q\in\mathcal{B}} Q^*)^c} |Tb(x)| \ dx \\ &\le \frac{2}{\alpha}\sum_{Q'\in\mathcal{B}} \int_{(\bigcup_{Q\in\mathcal{B}} Q^*)^c} |Tb_{Q'}(x)| \ dx \le \frac{2}{\alpha}\sum_{Q\in\mathcal{B}} \int_{(Q^*)^c} |Tb_Q(x)| \ dx. \end{aligned}$$

It remains to show that the sum above is bounded by $C\|f\|_{L^1}$.

Fix $Q \in \mathcal{B}$ and let y_Q denote the center of Q. Since $\operatorname{supp} b_Q \subset \overline{Q}$ and $\int_Q b_Q = 0$, we get that for $x \in (Q^*)^c$

$$Tb_Q(x) = \int_Q K(x,y) b_Q(y) \ dy = \int_Q (K(x,y) - K(x,y_Q)) b_Q(y) \ dy.$$

Observe that for $x \in (Q^*)^c$ and $y \in Q$ we have $|x-y| \geq \ell(Q^*)/2 = \sqrt{n}\ell(Q)$ and $|y-y_Q| \leq \mathrm{diam}(Q)/2 = \sqrt{n}\ell(Q)/2$, so that $|x-y| \geq 2|y-y_Q|$. It follows that we may use the smoothness condition on K (4.2) to estimate

$$|Tb_Q(x)| \lesssim \int_Q \frac{|y-y_Q|^\delta}{|x-y_Q|^{n+\delta}} |b_Q(y)| \, dy \lesssim \frac{\ell(Q)^\delta}{|x-y_Q|^{n+\delta}} \|b_Q\|_{L^1}.$$

Hence,

$$\int_{(Q^*)^c} |Tb_Q(x)| \, dx \lesssim \ell(Q)^\delta \|b_Q\|_{L^1} \int_{(Q^*)^c} \frac{1}{|x-y_Q|^{n+\delta}} \, dy \leq C(\delta) \|b_Q\|_{L^1},$$

which gives

$$\sum_{Q \in \mathcal{B}} \int_{(Q^*)^c} |Tb_Q(x)| \, dx \lesssim_\delta \sum_{Q \in \mathcal{B}} \|b_Q\|_{L^1} = \|b\|_{L^1} \leq \|f\|_{L^1} + \|g\|_{L^1} \leq 2\|f\|_{L^1}.$$

This finishes the proof. $\square$

5 Truncations of Calderón-Zygmund operators

5.1 Convergence of truncated operators

Definition 5.1. Given a Calderón-Zygmund operator T associated to a standard kernel K, for every $\varepsilon > 0$ we define the truncated operator T_ε as

$$T_\varepsilon f(x) = \int_{|x-y|>\varepsilon} K(x,y) f(y)\, dy,$$

where $f \in \bigcup_{1\le p<\infty} L^p(\mathbb{R}^n)$.

The integral defining $T_\varepsilon f$ makes sense for any $f \in \bigcup_{1\le p<\infty} L^p(\mathbb{R}^n)$ by the size condition (4.1) and Hölder's inequality.

Definition 5.2. Given a Calderón-Zygmund operator T, the *maximal operator associated to* T is defined as

$$T_* f(x) := \sup_{\varepsilon>0} |T_\varepsilon f(x)|.$$

We will prove the following result in the next subsection.

Theorem 5.3. *If T is a Calderón-Zygmund operator, then the maximal operator T_* is of weak type $(1,1)$ and of strong type (p,p) for all $1 < p < \infty$.*

We give an application of this result to the study of convergence of truncated operators.

Definition 5.4. A Calderón-Zygmund operator T is called a *Calderón-Zygmund singular integral operator* if for all $f \in C_c^\infty(\mathbb{R}^n)$ and a.e. $x \in \mathbb{R}^n$

$$Tf(x) = \lim_{\varepsilon\to 0} T_\varepsilon f(x).$$

Example 5.5. The Hilbert and Riesz transforms are Calderón-Zygmund singular integral operators.

Not all Calderón-Zygmund operators are Calderón-Zygmund singular integral operators. For some Calderón-Zygmund operators the limit $\lim_{\varepsilon\to 0} T_\varepsilon f(x)$ does not exist, see Example 5.9 and Proposition 5.12 in [Duo01]. For others, the limit exists but is different from Tf. For example, if $T = I$ is the identity operator, than $T_\varepsilon = 0$ for all $\varepsilon > 0$. See also Proposition 4.1.11 in [Gra14b] for a related result.

The following is a more general and stronger version of Proposition 3.12.

Proposition 5.6. *Suppose that T is a Calderón-Zygmund singular integral operator. For every $1 \leq p < \infty$ and $f \in L^p(\mathbb{R}^n)$ we have*

$$\lim_{\varepsilon \to 0} T_\varepsilon f(x) = Tf(x) \qquad \text{for a.e. } x \in \mathbb{R}^n. \tag{5.1}$$

Moreover, for $1 < p < \infty$ and $f \in L^p(\mathbb{R}^n)$ we have

$$\lim_{\varepsilon \to 0} \|T_\varepsilon f - Tf\|_{L^p} = 0. \tag{5.2}$$

Proof. For any $f \in L^p(\mathbb{R}^n)$, $1 \leq p < \infty$, define

$$\Lambda f(x) := \limsup_{\varepsilon \to 0} |T_\varepsilon f(x) - Tf(x)|.$$

Note that $\Lambda f \leq T_* f + Tf$. Since T is a Calderón-Zygmund singular integral operator, we have $\Lambda f = 0$ a.e. for $f \in C_c^\infty(\mathbb{R}^n)$.

For a general $f \in L^p(\mathbb{R}^n)$, let $f_n \in C_c^\infty(\mathbb{R}^n)$ be such that $f_n \to f$ in L^p. Then,

$$\Lambda f(x) \leq \Lambda f_n(x) + \Lambda(f - f_n)(x) = \Lambda(f - f_n)(x)$$

for a.e. $x \in \mathbb{R}^n$. If $1 < p < \infty$, we can estimate

$$\|\Lambda f\|_{L^p} = \|\Lambda(f - f_n)\|_{L^p} \leq \|T_*(f - f_n)\|_{L^p} + \|T(f - f_n)\|_{L^p} \leq C\|f_n - f\|_{L^p},$$

where in the last inequality we used the strong type (p,p) estimates for T_* and T. Letting $n \to \infty$ we get $\|\Lambda f\|_{L^p} = 0$, and so $\Lambda f = 0$ a.e. This gives (5.1).

The remaining conclusions are left as an exercise. □

Exercise 5.7. Prove (5.1) for $p = 1$, and (5.2) for $1 < p < \infty$.

Hint: For (5.1) use the weak type $(1,1)$ estimates of T and T_*. For (5.2) use (5.1) and the dominated convergence theorem.

5.2 Cotlar's inequality

Recall that the Hardy-Littlewood maximal function of $f \in L^1_{loc}(\mathbb{R}^n)$ is defined as

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

This operator satisfies weak type $(1,1)$ estimate, and strong type (p,p) estimates for $1 < p \leq \infty$. Moreover, the weak $(1,1)$ estimate can be refined to

$$|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \leq C \frac{\int_{\{Mf > \lambda\}} |f|}{\lambda}. \tag{5.3}$$

See Chapter 2.4 in [Duo01] or Chapter 2.1 in [Gra14a] for details.

The following estimate is sometimes referred to as Cotlar's inequality.

Theorem 5.8. *Suppose that T is a Calderón-Zygmund operator. For any $0 < r \leq 1$ there exists a constant $C = C(n, r, T)$ such that for any $f \in C_c^\infty(\mathbb{R}^n)$*

$$T_* f(x) \leq C(M(|Tf|^r)(x)^{1/r} + Mf(x)). \tag{5.4}$$

To prove this inequality, we need the following auxiliary estimate due to Kolmogorov.

Lemma 5.9. *Suppose that S is a weak type $(1,1)$ operator, and $E \subset \mathbb{R}^n$ is a measurable set with $|E| < \infty$. Then, there exists $C > 0$ (depending only on the weak $(1,1)$ constant) such that such that for all $f \in L^1(\mathbb{R}^n)$ and $0 < r < 1$*

$$\int_E |Sf(x)|^r \, dx \leq C \frac{1}{1-r} |E|^{1-r} \|f\|_{L^1}^r.$$

Proof. The layer cake formula and the weak type $(1,1)$ estimate for S give

$$\begin{aligned}
\int_E |Sf(x)|^r \, dx &= r \int_0^\infty \lambda^{r-1} |\{x \in E : |Sf(x)| > \lambda\}| \, d\lambda \\
&\leq r \int_0^\infty \lambda^{r-1} \min(|E|, C\|f\|_{L^1}/\lambda) \, d\lambda \\
&= r \int_0^{C\|f\|_{L^1}/|E|} \lambda^{r-1} |E| \, d\lambda + Cr \int_{C\|f\|_{L^1}/|E|}^\infty \lambda^{r-2} \|f\|_{L^1} \, d\lambda \\
&= (C\|f\|_{L^1}/|E|)^r |E| + C \frac{r}{1-r} (C\|f\|_{L^1}/|E|)^{r-1} \|f\|_{L^1}.
\end{aligned}$$

□

Proof of Theorem 5.8. Fix $f \in C_c^\infty(\mathbb{R}^n)$, $x \in \mathbb{R}^n$ and $\varepsilon > 0$. We will show that

$$|T_\varepsilon f(x)| \leq C(M(|Tf|^r)(x)^{1/r} + Mf(x)), \tag{5.5}$$

with C independent of ε.

Let $B = B(x, \varepsilon/2)$ and $2B = B(x, \varepsilon)$. Let $f_1 = f\mathbf{1}_{2B}$ and $f_2 = f\mathbf{1}_{2B^c} = f - f_1$. Then,

$$T_\varepsilon f(x) = \int_{|x-y|>\varepsilon} K(x,y) f(y) \, dy = T(f\mathbf{1}_{2B^c})(x) = Tf_2(x),$$

where in the second equality we used the fact that $x \notin \operatorname{supp}(f\mathbf{1}_{2B^c})$ and the representation formula (4.5).

For $x' \in B$ and $y \in 2B^c$ we have $|x' - x| \le |x - y|/2$, and so

$$\begin{aligned}
|Tf_2(x) - Tf_2(x')| &= \left| \int_{|x-y|>\varepsilon} (K(x,y) - K(x',y)) f(y)\, dy \right| \\
&\lesssim \int_{|x-y|>\varepsilon} \frac{|x - x'|^\delta}{|x-y|^{n+\delta}} |f(y)|\, dy \\
&\lesssim \varepsilon^\delta \sum_{k=0}^{\infty} \int_{2^k\varepsilon<|x-y|\le 2^{k+1}\varepsilon} (2^k\varepsilon)^{-n-\delta} |f(y)|\, dy \\
&\lesssim \sum_{k=0}^{\infty} (2^k)^{-\delta} \frac{1}{(2^{k+1}\varepsilon)^n} \int_{|x-y|\le 2^{k+1}\varepsilon} |f(y)|\, dy \\
&\le \sum_{k=0}^{\infty} (2^k)^{-\delta} Mf(x) \le C(\delta) Mf(x).
\end{aligned}$$

Thus, for any $x' \in B$

$$|T_\varepsilon f(x)| = |Tf_2(x)| \le |Tf_2(x')| + CMf(x) \le |Tf(x')| + |Tf_1(x')| + CMf(x). \quad (5.6)$$

If $|T_\varepsilon f(x)| \le 3CMf(x)$ then (5.5) holds, so suppose that $|T_\varepsilon f(x)| > 3CMf(x) > 0$. We define

$$\begin{aligned}
B_1 &= \{x' \in B : |Tf(x')| \ge |T_\varepsilon f(x)|/3\}, \\
B_2 &= \{x' \in B : |Tf_1(x')| \ge |T_\varepsilon f(x)|/3\}.
\end{aligned}$$

Note that $B = B_1 \cup B_2$. By Chebyshev's inequality

$$|B_1| \lesssim \frac{1}{|T_\varepsilon f(x)|} \int_B |Tf(x')|\, dx' \le \frac{1}{|T_\varepsilon f(x)|} |B|\, M(|Tf|)(x).$$

By the weak $(1,1)$ estimate for T

$$|B_2| \lesssim \frac{1}{|T_\varepsilon f(x)|} \|f_1\|_{L_1} = \frac{1}{|T_\varepsilon f(x)|} \int_{2B} |f| \lesssim \frac{1}{|T_\varepsilon f(x)|} |B|\, Mf(x).$$

Summing the two inequalities above we get

$$|B| \lesssim \frac{1}{|T_\varepsilon f(x)|} |B|\, (M(|Tf|)(x) + Mf(x)),$$

which gives (5.5) for $r = 1$.

To get (5.5) for $0 < r < 1$, we raise (5.6) to power r, so that

$$|T_\varepsilon f(x)|^r \lesssim |Tf(x')|^r + |Tf_1(x')|^r + Mf(x)^r.$$

Averaging over $x' \in B$ and then raising to power $1/r$ we get

$$|T_\varepsilon f(x)| \lesssim M(|Tf|^r)(x)^{1/r} + \left(\frac{1}{|B|}\int_B |Tf_1(x')|^r \, dx'\right)^{1/r} + Mf(x).$$

By Lemma 5.9,

$$\left(\frac{1}{|B|}\int_B |Tf_1(x')|^r \, dx'\right)^{1/r} \lesssim_r |B|^{-1}\|f_1\|_{L^1} = \frac{1}{|B|}\int_{2B} |f| \lesssim Mf(x),$$

which finishes the proof. □

We are ready to prove Theorem 5.3, which asserted that T_* is of weak type $(1,1)$ and strong type (p,p) for $1 < p < \infty$.

Proof of Theorem 5.3. If $1 < p < \infty$, the strong type (p,p) estimate for T_* follows from Cotlar's inequality (5.4) with $r = 1$ and the strong type (p,p) estimates for M and T.

To get the weak type $(1,1)$ estimate for T_*, we use (5.4) with $r = 1/2$ to estimate

$$\begin{aligned}|\{x \in \mathbb{R}^n : |T_* f(x)| \geq \alpha\}| \leq |\{x \in \mathbb{R}^n &: M(|Tf|^{1/2})(x)^2 \geq \alpha/(2C)\}| \\ &+ |\{x \in \mathbb{R}^n : Mf(x) \geq \alpha/(2C)\}|.\end{aligned}$$

The second term is bounded by $C'\|f\|_{L^1}/\alpha$ by the weak type $(1,1)$ estimate for M.

To bound the first term, let $E = |\{x \in \mathbb{R}^n : M(|Tf|^{1/2})(x) \geq \alpha^{1/2}/(2C)^{1/2}\}|$. We use the refined weak type $(1,1)$ estimate for M (5.3), and then Lemma 5.9 to get

$$|E| \lesssim \frac{\int_E |Tf|^{1/2}}{\alpha^{1/2}} \lesssim \frac{|E|^{1/2}\|f\|_{L^1}^{1/2}}{\alpha^{1/2}}.$$

Rearranging this inequality finishes the proof. □

6 Weighted inequalities

6.1 The A_p weights

Definition 6.1 (weight). We define *weights* as locally integrable functions $w : \mathbb{R}^n \to [0, \infty]$. Each weight gives rise to a locally finite measure, still denoted by w, via

$$w(A) = \int_A w.$$

We are interested in studying singular integral operators in the weighted setting $(\mathbb{R}^n, w)$. Given the importance of the Hardy-Littlewood maximal operator M in this theory, it is reasonable to start our investigation by determining the weights for which M is of weak type (p, p) with respect to w, $1 \leq p < \infty$. By definition, M is of weak type (p, p) with respect to w if and only if for every $\lambda > 0$ and $f \in L^p(w)$

$$w(\{x \in \mathbb{R}^n : \; Mf(x) > \lambda\}) \leq \frac{C}{\lambda^p} \int |f(x)|^p w(x) \; dx. \tag{6.1}$$

Instead of the usual Hardy-Littlewood maximal operator, it will be convenient for us to study its non-centered variant associated to cubes. For any $f \in L^1_{loc}(\mathbb{R}^n)$ we define

$$M_c f(x) = \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |f|,$$

where the supremum is taken over all axis-parallel cubes containing x (from now on when we write "cubes" we always assume they are axis-parallel).

Exercise 6.2. Show that for any $f \in L^1_{loc}(\mathbb{R}^n)$ and $x \in \mathbb{R}^n$ we have

$$C^{-1} Mf(x) \leq M_c f(x) \leq C Mf(x).$$

Conclude that M is of weak type (p, p) with respect to w for some $1 \leq p < \infty$ if and only if M_c is of weak type (p, p) with respect to w.

We now derive a necessary condition for w so that M_c is of weak type (p, p) with respect to w. Suppose that (6.1) holds. Let Q be a cube, and $f \in L^1_{loc}(\mathbb{R}^n)$ be such that $\int_Q f > 0$. Fix $0 < \lambda < \int_Q f/|Q|$. Then,

$$Q \subset \{x \in \mathbb{R}^n : M_c(f \mathbf{1}_Q)(x) > \lambda\},$$

and so the weak type (p, p) estimate implies

$$w(Q) \leq \frac{C}{\lambda^p} \int_Q |f(x)|^p w(x) dx.$$

Taking $\lambda \to \int_Q f/|Q|$ we arrive at

$$w(Q)\left(\frac{\int_Q f}{|Q|}\right)^p \leq C\int_Q |f(x)|^p w(x)dx. \tag{6.2}$$

Let $S \subset Q$ be measurable with $|S| > 0$. Taking $f = \mathbf{1}_S$, the inequality above gives

$$w(Q)\left(\frac{|S|}{|Q|}\right)^p \leq Cw(S). \tag{6.3}$$

Since this holds for all cubes Q and all $S \subset Q$ with $|S| > 0$, we get that either $w \equiv 0$ (which is not too interesting), or $w(x) > 0$ for a.e. $x \in \mathbb{R}^n$.

Now there are two cases to consider.

Case $p = 1$. If $p = 1$, (6.3) becomes

$$\frac{w(Q)}{|Q|} \leq C\frac{w(S)}{|S|}.$$

Let $a = \operatorname{ess\,inf}\{w(x) : x \in Q\}$. Then, for every $\varepsilon > 0$ there exists $S_\varepsilon \subset Q$ with $|S_\varepsilon| > 0$ and such that for all $x \in S_\varepsilon$ we have $w(x) \leq a + \varepsilon$. It follows that

$$\frac{w(Q)}{|Q|} \leq C\frac{w(S_\varepsilon)}{|S_\varepsilon|} = C\frac{\int_{S_\varepsilon} w}{|S_\varepsilon|} \leq C(a+\varepsilon).$$

Taking $\varepsilon \to 0$, we get that

$$\frac{w(Q)}{|Q|} \leq C \operatorname*{ess\,inf}_{x\in Q} w(x).$$

Hence, for every cube $Q \subset \mathbb{R}^n$

$$\frac{w(Q)}{|Q|} \leq Cw(x), \quad \text{for a.e. } x \in Q. \tag{6.4}$$

Definition 6.3 (A_1 weights). A weight w satisfies the A_1 *condition* if (6.4) holds for every cube $Q \subset \mathbb{R}^n$. The positive weights w satisfying the A_1 condition are called the A_1 *weights,* and we will write $w \in A_1$ for such weights.

The smallest constant C such that (6.4) holds is called the A_1 *character of* w, and it is denoted by

$$[w]_{A_1} := \sup_{Q\subset\mathbb{R}^n} \frac{w(Q)}{|Q|}\|w^{-1}\|_{L^\infty(Q)}.$$

Exercise 6.4. Show that the A_1 condition is equivalent to

$$M_c w(x) \leq Cw(x) \quad \text{for a.e. } x \in \mathbb{R}^n. \tag{6.5}$$

Case $1 < p < \infty$. Let $1 < p' < \infty$ be such that $1/p + 1/p' = 1$. We plug into (6.2) the function $f = w^{1-p'}\mathbf{1}_Q$,[2] so that

$$w(Q)\left(\frac{1}{|Q|}\int_Q w^{1-p'}\right)^p \leq C\int_Q w^{(1-p')p+1} = C\int_Q w^{1-p'}.$$

Rearranging, we get

$$\left(\frac{1}{|Q|}\int_Q w\right)\left(\frac{1}{|Q|}\int_Q w^{1-p'}\right)^{p-1} \leq C. \tag{6.6}$$

Definition 6.5 (A_p weights). A weight w satisfies the A_p *condition* if (6.6) holds for every cube $Q \subset \mathbb{R}^n$. The positive weights w satisfying the A_p condition are called the A_p *weights,* and we will write $w \in A_p$ for such weights.

The smallest constant C such that (6.6) holds is called the A_p *character of* w, and it is denoted by

$$[w]_{A_p} \coloneqq \sup_{Q\subset\mathbb{R}^n}\left(\frac{1}{|Q|}\int_Q w\right)\left(\frac{1}{|Q|}\int_Q w^{1-p'}\right)^{p-1}.$$

Note that the definition for $p = 2$ is particularly nice, since the A_2 condition is just

$$\left(\frac{1}{|Q|}\int_Q w\right)\left(\frac{1}{|Q|}\int_Q w^{-1}\right) \leq C.$$

The A_1 and A_p conditions are often called *Muckenhoupt conditions.* It turns out that they are not only necessary for the weak estimates for M_c, but also sufficient.

Proposition 6.6. *For* $1 \leq p < \infty$ *the Hardy-Littlewood maximal operator is of weak type* (p, p) *with respect to a weight* w *if and only if* $w \in A_p$.

Proof. Assume that $w \in A_p$ and $f \in L^p(w)$. Our goal is to show

$$w(\{x \in \mathbb{R}^n : M_c f(x) > 4^n\lambda\}) \leq \frac{C}{\lambda^p}\int |f(x)|^p w(x)\, dx. \tag{6.7}$$

Assume additionally that $f \in L^1(\mathbb{R}^n)$. Let $\mathcal{B} \subset \mathcal{D}(\mathbb{R}^n)$ be the family of cubes given by the Calderón-Zygmund decomposition of f at level λ (see Proposition 4.15). We claim that

$$\{x \in \mathbb{R}^n : M_c f(x) > 4^n\lambda\} \subset \bigcup_{Q\in\mathcal{B}} 3Q \tag{6.8}$$

[2]Here we implicitly assume that $w^{1-p'}$ is locally integrable. To avoid this, we could consider $\min(w^{1-p'}, N)$ instead, and at the end take $N \to \infty$.

Let $x \in \mathbb{R}^n$ be such that $M_c f(x) > 4^n\lambda$, and let $x \in P \subset \mathbb{R}^n$ be a cube such that $\frac{1}{|P|}\int_P |f| > 4^n\lambda$. Fix $k \in \mathbb{Z}$ such that $2^{-k-1} \le \ell(P) < 2^{-k}$. Note that P may intersect at most 2^n cubes from $\mathcal{D}_k(\mathbb{R}^n)$, and we denote them by $R_1, \dots, R_m$, $m \le 2^n$.

If one of the R_j's is contained in some $Q \in \mathcal{B}$, then $P \subset 3R_j \subset 3Q$ and we are done with (6.8). So suppose that none of R_j's is contained in any $Q \in \mathcal{B}$. Then, by the definition of bad cubes $\mathcal{B}$ (see the proof of Proposition 4.15) we get that for all $1 \le j \le m$

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

Hence,

$$\frac{1}{|P|}\int_P |f| \le \sum_{j=1}^m \frac{\ell(R_j)^n}{\ell(P)^n}\frac{1}{|R_j|}\int_{R_j} |f| \le m\, 2^n\lambda \le 4^n\lambda,$$

which is a contradiction with $\frac{1}{|P|}\int_P |f| > 4^n\lambda$. This finishes the proof of (6.8).

Now we argue separately for $p = 1$ and $p > 1$. Suppose first that $p = 1$. It follows from (6.8) and the Calderón-Zygmund decomposition property (4.9) that

$$\begin{aligned} w(\{x \in \mathbb{R}^n : M_c f(x) > 4^n\lambda\}) &\le \sum_{Q\in\mathcal{B}} w(3Q) \le \frac{1}{\lambda}\sum_{Q\in\mathcal{B}} \frac{w(3Q)}{|Q|}\int_Q |f| \\ &\le \frac{C}{\lambda}\sum_{Q\in\mathcal{B}}\int_Q |f(x)|\frac{w(3Q)}{|3Q|}\,dx. \end{aligned}$$

By the A_1 condition, we have $\frac{w(3Q)}{|3Q|} \le w(x)$ for a.e. $x \in 3Q$, and so

$$w(\{x \in \mathbb{R}^n : M_c f(x) > 4^n\lambda\}) \le \frac{C}{\lambda}\sum_{Q\in\mathcal{B}}\int_Q |f(x)|w(x)\,dx \le \frac{C}{\lambda}\int_{\mathbb{R}^n} |f(x)|w(x)\,dx,$$

which gives (6.7) for $p = 1$ and $f \in L^1(w) \cap L^1(\mathbb{R}^n)$.

Assume now $p > 1$. By Hölder's inequality and the defintion of A_p weights, for any cube P

$$\begin{aligned} \left(\frac{1}{|P|}\int_P |f|\right)^p &= \left(\frac{1}{|P|}\int_P |f|w^{1/p}w^{-1/p}\right)^p \\ \le \left(\frac{1}{|P|}\int_P |f|^p w\right)\left(\frac{1}{|P|}\int_P w^{-p'/p}\right)^{p/p'} &= \left(\frac{1}{|P|}\int_P |f|^p w\right)\left(\frac{1}{|P|}\int_P w^{1-p'}\right)^{p-1} \\ &\le C\left(\frac{1}{|P|}\int_P |f|^p w\right)\left(\frac{|P|}{w(P)}\right) = C\frac{1}{w(P)}\int_P |f|^p w. \quad (6.9) \end{aligned}$$

Taking $P = 3Q$ and $f = \mathbf{1}_Q$ for some cube Q, it follows that $w(3Q) \le C3^{np}w(Q)$. We use again (6.8), the estimate (6.9), and (4.9), to get

$$\begin{aligned}
w(\{x \in \mathbb{R}^n : M_c f(x) > 4^n\lambda\}) &\le \sum_{Q\in\mathcal{B}} w(3Q) \lesssim \sum_{Q\in\mathcal{B}} w(Q) \\
&\lesssim \sum_{Q\in\mathcal{B}} \left(\frac{1}{|Q|}\int_Q |f|\right)^{-p} \int_Q |f|^p w \\
&\le \sum_{Q\in\mathcal{B}} \lambda^{-p} \int_Q |f|^p w \le \lambda^{-p} \int_{\mathbb{R}^n} |f|^p w.
\end{aligned}$$

This gives (6.7) for $1 < p < \infty$ and $f \in L^p(w) \cap L^1(\mathbb{R}^n)$.

It remains to show that $L^1(\mathbb{R}^n)$ is dense in $L^p(w)$ for $1 \le p < \infty$, and we leave this as an exercise. $\square$

Exercise 6.7. Let $1 \le p < \infty$ and $w \in A_p$. Show that $L^1(\mathbb{R}^n)$ is dense in $L^p(w)$.

Hint: For any $f \in L^p(w)$ prove that $f_R \coloneqq f\mathbf{1}_{B(0,R)} \in L^1(\mathbb{R}^n)$ for all $R > 0$, and that $f_R \to f$ in $L^p(w)$ as $R \to \infty$. The estimate (6.9) and its modification for $p = 1$ may be helpful.

We list a few basic properties of the A_p weights. Below $\mathcal{L}^n$ denotes the Lebesgue measure on $\mathbb{R}^n$.

Lemma 6.8. *We have $A_p \subset A_q$ for $1 \le p \le q < \infty$. Moreover, for any $w \in A_p$ we have*

(i) for any cube Q and $E \subset Q$ measurable

$$\left(\frac{|E|}{|Q|}\right)^p \le C\frac{w(E)}{w(Q)}. \tag{6.10}$$

In particular, $\mathcal{L}^n \ll w$.

(ii) For every $\alpha \in (0,1)$ there exists $\beta \in (0,1)$ such that for every cube Q and $E \subset Q$ measurable

$$|E| \le \alpha|Q| \quad \Rightarrow \quad w(E) \le \beta w(Q).$$

In particular, $w \ll \mathcal{L}^n$.

(iii) w is doubling: for any ball B we have $w(2B) \le Cw(B)$.

(iv) if $p > 1$, then $w^{1-p'} \in A_{p'}$.

Proof. Suppose that $w \in A_p$ and $q > p$. If $p = 1$, then

$$\left(\frac{1}{|Q|}\int_Q w^{1-q'}\right)^{q-1} \le \operatorname*{ess\,sup}_{x\in Q} w(x)^{-1} = \left(\operatorname*{ess\,inf}_{x\in Q} w(x)\right)^{-1} \le C\left(\frac{w(Q)}{|Q|}\right)^{-1},$$

so $w \in A_q$. For $p > 1$, it follows from Hölder's inequality that

$$\left(\frac{1}{|Q|}\int_Q w^{1-q'}\right)^{q-1} \le \left(\frac{1}{|Q|}\left(\int_Q w^{1-p'}\right)^{\frac{1-q'}{1-p'}} |Q|^{1-\frac{1-q'}{1-p'}}\right)^{q-1} = \left(\frac{1}{|Q|}\int_Q w^{1-p'}\right)^{p-1}.$$

Thus, $A_p \subset A_q$.

To show (i) for $w \in A_1$, note that by integrating the A_1 condition (6.4) over E we have

$$|E| \cdot \frac{w(Q)}{|Q|} \le w(E).$$

If $w \in A_p$ with $p > 1$, then by plugging $f = \mathbf{1}_E$ into (6.9) we get the desired inequality.

To get (ii), observe that replacing E by $Q \setminus E$ in (6.10) gives

$$\left(1 - \frac{|E|}{|Q|}\right)^p \leq C\left(1 - \frac{w(E)}{w(Q)}\right),$$

and so $|E| \leq \alpha|Q|$ implies

$$(1-\alpha)^p \leq C\left(1 - \frac{w(E)}{w(Q)}\right),$$

which is equivalent to

$$w(E) \leq \left(1 - \frac{(1-\alpha)^p}{C}\right) w(Q).$$

This gives the desired inequality with $\beta = 1 - \frac{(1-\alpha)^p}{C}$.

The doubling property (iii) follows immediately from (6.10) by taking $E = B$ and Q a cube containing $2B$ with $\ell(Q) \sim r(B)$.

Finally, to get (iv) observe that the $A_{p'}$ condition for $w^{1-p'}$ is

$$\left(\frac{1}{|Q|}\int_Q w^{1-p'}\right)\left(\frac{1}{|Q|}\int_Q w^{(1-p')(1-p)}\right)^{p'-1} \leq C,$$

and since $(1-p')(1-p) = 1$, this is the A_p condition raised to power $p' - 1$. $\square$

Exercise 6.9. Prove that in the definition of the A_p condition we may replace cubes by balls and still get the same class of weights. More specifically,

$$\sup_{Q \subset \mathbb{R}^n}\left(\frac{1}{|Q|}\int_Q w\right)\left(\frac{1}{|Q|}\int_Q w^{1-p'}\right)^{p-1} \sim \sup_{B \subset \mathbb{R}^n}\left(\frac{1}{|B|}\int_B w\right)\left(\frac{1}{|B|}\int_B w^{1-p'}\right)^{p-1},$$

where Q are cubes and B are balls.

Exercise 6.10. Prove that $w(x) = |x|^a$ is an A_p weight on $\mathbb{R}^n$, $1 < p < \infty$, if and only if $-n < a < n(p-1)$.

Hint: Show first that $w(x) = |x|^a$ is a doubling weight ($w(2B) \leq Cw(B)$ for all balls) if and only if $a > -n$. Consider separately balls $B = B(x_0, R)$ such that $|x_0| \geq 3R$, and such that $|x_0| < 3R$.

Exercise 6.11. Show that

$$w(x) = \begin{cases} \log \frac{1}{|x|} & |x| \leq e^{-1} \\ 1 & |x| > e^{-1} \end{cases}$$

is an A_1 weight.

6.2 Reverse Hölder inequality

In this subsection we will talk about weighted strong type estimates for the Hardy-Littlewood maximal operator M.

Suppose that $w \in A_p$ for some $p \geq 1$. Observe that by Lemma 6.8 (i), we have $L^\infty(\mathbb{R}^n) = L^\infty(w)$ with equality of norms. In particular, the Hardy-Littlewood maximal operator is of strong type (∞, ∞) with respect to w.

By Proposition 6.6 we also have that M is of weak type (p, p) with respect to w, and so by the Marcinkiewicz interpolation theorem we get that M is of strong type (q, q) with respect to w for all $p < q < \infty$, in the sense that

$$\int |Mf(x)|^q \, w(x) \, dx \leq C \int |f(x)|^q \, w(x) \, dx.$$

It turns out that the same is true at the endpoint $q = p$, and we have the following improvement over Proposition 6.6.

Theorem 6.12. *For $1 < p < \infty$ the Hardy-Littlewood maximal operator is of strong type (p, p) with respect to a weight w if and only if $w \in A_p$.*

To prove this, we will establish an important property of Muckenhoupt weights called the *reverse Hölder inequality.*

Theorem 6.13. *Let $1 \leq p < \infty$ and $w \in A_p$. There exist constants $C \geq 1$ and $\varepsilon > 0$, depending only on p and $[w]_{A_p}$, such that for any cube Q*

$$\left(\frac{1}{|Q|} \int_Q w^{1+\varepsilon} \right)^{1/(1+\varepsilon)} \leq \frac{C}{|Q|} \int_Q w.$$

Note that the converse estimate holds (with $C = 1$) by Hölder's inequality, hence the name "reverse Hölder inequality".

Proof. Fix $w \in A_p$ and a cube Q. Without loss of generality, we may assume that Q is a dyadic cube (otherwise we replace w by a translated and dilated weight w').

Consider an increasing sequence $\lambda_k \to \infty$, with $\lambda_0 = w(Q)/|Q|$. For every $k \in \mathbb{N}$ let $\mathcal{B}_k$ be the family of dyadic sub-cubes of Q given by the Calderón-Zygmund decomposition of $w\mathbf{1}_Q$ at the level λ_k (see Proposition 4.15). That is, $\mathcal{B}_k$ is the family of maximal sub-cubes of Q satisfying

$$\lambda_k < \frac{w(P)}{|P|} \leq 2^n \lambda_k, \quad P \in \mathcal{B}_k. \tag{6.11}$$

Let $\Omega_k := \bigcup_{P \in \mathcal{B}_k} P$, and observe that

$$w(x) \leq \lambda_k \quad \text{for a.e. } x \notin \Omega_k. \tag{6.12}$$

Note that every cube in $\mathcal{B}_k$ is contained in some cube from $\mathcal{B}_{k-1}$ (this follows from the definition of $\mathcal{B}_k$ and the fact that $\lambda_k > \lambda_{k-1}$). In particular, $\Omega_k \subset \Omega_{k-1}$.

Given $P \in \mathcal{B}_{k-1}$ let $\mathcal{B}_k(P)$ be the family of cubes from $\mathcal{B}_k$ contained in P. Then,

$$|P \cap \Omega_k| = \sum_{R \in \mathcal{B}_k(P)} |R| \overset{(6.11)}{\leq} \frac{1}{\lambda_k} \sum_{R \in \mathcal{B}_k(P)} w(R) \leq \frac{1}{\lambda_k} w(P) \overset{(6.11)}{\leq} \frac{2^n \lambda_{k-1}}{\lambda_k} |P|.$$

Let $\lambda_k \coloneqq 2^{(n+1)k}\lambda_0 = 2^{(n+1)k} w(Q)/|Q|$. Then the estimate above gives

$$|P \cap \Omega_k| \leq \frac{|P|}{2}. \tag{6.13}$$

By Lemma 6.8 (ii) (applied with $\alpha = 1/2$) we get that there exists $\beta = \beta(p, [w]_{A_p}) \in (0,1)$ such that

$$w(P \cap \Omega_k) \leq \beta w(P)$$

Summing over all $P \in \mathcal{B}_{k-1}$ gives $w(\Omega_k) \leq \beta w(\Omega_{k-1})$, and iterating this inequality yields

$$w(\Omega_k) \leq \beta^k w(\Omega_0).$$

We may use (6.13) similarly to get $|\Omega_k| \leq 2^{-k}|\Omega_0|$, and so

$$\left| \bigcap_{k \geq 0} \Omega_k \right| = \lim_{k \to \infty} |\Omega_k| = 0.$$

Hence,

$$\begin{aligned}
\frac{1}{|Q|} \int_Q w^{1+\varepsilon} &= \frac{1}{|Q|} \int_{Q \setminus \Omega_0} w^{1+\varepsilon} + \frac{1}{|Q|} \sum_{k \geq 0} \int_{\Omega_k \setminus \Omega_{k+1}} w^{1+\varepsilon} \\
&\overset{(6.12)}{\leq} \lambda_0^{\varepsilon} \frac{w(Q \setminus \Omega_0)}{|Q|} + \frac{1}{|Q|} \sum_{k \geq 0} \lambda_k^{\varepsilon} w(\Omega_k \setminus \Omega_{k+1}) \leq \lambda_0^{\varepsilon} \frac{w(Q)}{|Q|} + \frac{1}{|Q|} \sum_{k \geq 0} \lambda_k^{\varepsilon} w(\Omega_k) \\
&\leq \lambda_0^{\varepsilon} \frac{w(Q)}{|Q|} + \lambda_0^{\varepsilon} \frac{w(Q)}{|Q|} \sum_{k \geq 0} 2^{(n+1)k\varepsilon} \beta^k.
\end{aligned}$$

Choosing $\varepsilon > 0$ so small that $2^{(n+1)\varepsilon}\beta < 1$, we get that the geometric series above converges, and so

$$\frac{1}{|Q|} \int_Q w^{1+\varepsilon} \leq C \lambda_0^{\varepsilon} \frac{w(Q)}{|Q|} = C \left(\frac{w(Q)}{|Q|} \right)^{1+\varepsilon}.$$

□

An easy corollary of the reverse Hölder inequality is the self-improving property of A_p weights.

Corollary 6.14. *For every $p > 1$ and $w \in A_p$ there exists $\varepsilon > 0$ such that $w \in A_{p-\varepsilon}$. In particular,*

$$A_p = \bigcup_{q \in [1,p)} A_q.$$

Proof. By Lemma 6.8 (iii) we have $w^{1-p'} \in A_{p'}$. The reverse Hölder inequality for $w^{1-p'}$ asserts that for some $\varepsilon > 0$

$$\left(\frac{1}{|Q|} \int_Q w^{(1-p')(1+\varepsilon)} \right)^{1/(1+\varepsilon)} \leq \frac{C}{|Q|} \int_Q w^{(1-p')}.$$

Let $q > 1$ be such that $1 - q' = (1-p')(1+\varepsilon)$. Then $q < p$, and the inequality above together with the A_p condition give $w \in A_q$. $\square$

Now we can easily prove the strong type (p,p) estimate with respect to A_p weights for the Hardy-Littlewood maximal operator, $p > 1$.

Proof of Theorem 6.12. Suppose that $w \in A_p$ with $p > 1$. Then, $w \in A_q$ for some $q < p$, and we already know that M is of strong type (r,r) with respect to w for all $q < r < \infty$ (see the discussion above Theorem 6.12). In particular, it is of strong type (p,p) with respect to w. $\square$

6.3 Characterization of A_1 weights

In this subsection we prove the following characterization of the A_1 weights.

Proposition 6.15. *Suppose that $f \in L^1_{loc}(\mathbb{R}^n)$ is such that $M_c f(x) < \infty$ for a.e. $x \in \mathbb{R}^n$. Then, for every $0 < s < 1$ the weight $w = (M_c f)^s$ is an A_1 weight, with $[w]_{A_1}$ depending only on s, and not on f.*

Conversely, for every $w \in A_1$ there exists $f \in L^1_{loc}(\mathbb{R}^n)$, $0 < s < 1$ and $C = C([w]_{A_1})$ such that

$$w(x) \leq M_c f(x)^s \leq C w(x) \quad \text{for a.e. } x \in \mathbb{R}^n. \tag{6.14}$$

Proof. Assume that $f \in L^1_{loc}(\mathbb{R}^n)$ and $M_c f(x) < \infty$ for a.e. $x \in \mathbb{R}^n$. We need to show that for every cube Q and a.e. $x \in Q$

$$\frac{1}{|Q|} \int_Q M_c f(y)^s \, dy \leq C M_c f(x)^s. \tag{6.15}$$

Fix Q, and observe that

$$M_c f(y)^s \leq M_c(f \mathbf{1}_{2Q})(y)^s + M_c(f \mathbf{1}_{2Q^c})(y)^s.$$

By Lemma 5.9

$$\frac{1}{|Q|}\int_Q M_c(f\mathbf{1}_{2Q})(y)^s\,dy \lesssim_s \frac{|Q|^{1-s}}{|Q|}\|f\mathbf{1}_{2Q}\|_{L^1}^s = \left(\frac{1}{|Q|}\int_{2Q}|f|\,dy\right)^s \lesssim M_cf(x)^s$$

for every $x \in Q$.

Now we want to estimate $M_c(f\mathbf{1}_{2Q^c})(y)$ for $y \in Q$. Observe that if R is a cube such that $y \in R$ and $\int_R |f\mathbf{1}_{2Q^c}| > 0$, then $R \cap Q \neq \varnothing$ and $R \setminus 2Q \neq \varnothing$. In particular, $\ell(R) \geq \ell(Q)/2$, and $Q \subset 5R$. It follows that

$$\frac{1}{|R|}\int_R |f\mathbf{1}_{2Q^c}(z)|dz \leq \frac{5^n}{|5R|}\int_{5R}|f|dz \lesssim M_cf(x).$$

Taking supremum over cubes R containing y, we get that $M_c(f\mathbf{1}_{2Q^c})(y) \lesssim M_cf(x)$ for every $y \in Q$, and so

$$\frac{1}{|Q|}\int_Q M_c(f\mathbf{1}_{2Q^c})(y)^s\,dy \leq CM_cf(x)^s.$$

This finishes the proof of (6.15) and the first half of the proposition.

Now suppose that $w \in A_1$. By Theorem 6.13, there exists $\varepsilon > 0$ such that

$$\left(\frac{1}{|Q|}\int_Q w^{1+\varepsilon}\right)^{1/(1+\varepsilon)} \leq C\frac{w(Q)}{|Q|}.$$

Together with the A_1 condition, this implies that $M(w^{1+\varepsilon})(x)^{1/(1+\varepsilon)} \leq Cw(x)$ for a.e. $x \in \mathbb{R}^n$. Since we also have $w^{1+\varepsilon}(x) \leq M(w^{1+\varepsilon})(x)$, taking $f = w^{1+\varepsilon}$ and $s = 1/(1+\varepsilon)$ we get (6.14). □

Proposition 6.15 is a useful tool for coming up with examples of A_1 weights. The following lemma allows us to construct A_p weights using A_1 weights.

Lemma 6.16. *Let $1 < p < \infty$. If $w_1, w_2 \in A_1$, then $w = w_1w_2^{1-p}$ is an A_p weight.*

Proof. We present the proof for $p = 2$, and leave the general case as an exercise. We need to show that for every cube Q

$$\left(\frac{1}{|Q|}\int w_1w_2^{-1}\right)\left(\frac{1}{|Q|}\int w_1^{-1}w_2\right) \leq C.$$

By the A_1 condition for w_1 and w_2, we have

$$\begin{aligned}&\left(\frac{1}{|Q|}\int w_1w_2^{-1}\right)\left(\frac{1}{|Q|}\int w_1^{-1}w_2\right)\\ &\qquad \leq (\operatorname*{ess\,inf}_{x\in Q} w_2(x))^{-1}\left(\frac{1}{|Q|}\int w_1\right)(\operatorname*{ess\,inf}_{x\in Q} w_1(x))^{-1}\left(\frac{1}{|Q|}\int w_2\right)\\ &\qquad \leq [w_1]_{A_1}[w_2]_{A_1}.\end{aligned}$$

□

Exercise 6.17. Modify the proof of Lemma 6.16 to cover all $1 < p < \infty$.

Remark 6.18. It turns out that the converse of Lemma 6.16 is also true: any A_p weight w can be written as $w = w_1 w_2^{1-p}$ for some $w_1, w_2 \in A_1$. This important result is known as the *factorization of A_p weights*, see Chapter V.5.3 in [Ste93] or Section 4 in [CU17] for a proof.

6.4 Extrapolation of weights

One of the key results in the theory of A_p weights is the *Rubio de Francia extrapolation theorem*, which says that a weighted inequality obtained for one exponent $1 < r < \infty$ implies the same for all $1 < p < \infty$.

Theorem 6.19. *Let $1 < p_0 < \infty$. Suppose that an operator T is of strong type (p_0, p_0) with respect to all weights $w \in A_{p_0}$, with operator norm depending only on $[w]_{A_{p_0}}$. Then, T is of strong type (p, p) with respect to all weights $w \in A_p$ and all $1 < p < \infty$.*

Proof. First, assume that $w \in A_1$. We will show that T is of strong type (p, p) with respect to w for all $1 < p < p_0$.

Let $f \in L^p(w)$. Note that $M_c f(x) < \infty$ for a.e. $x \in \mathbb{R}^n$ (because $M_c f \in L^p(w)$ by the strong (p, p) estimate with respect to w for M_c). Thus, by Proposition 6.15 we have $(M_c f)^{(p_0-p)/(p_0-1)} \in A_1$ (note that $p_0 - p < p_0 - 1$). Then, by Lemma 6.16 the weight $w \cdot (M_c f)^{p-p_0}$ is in A_{p_0}. Hence,

$$\begin{aligned}
\int |Tf|^p w &= \int |Tf|^p w (M_c f)^{-(p_0-p)p/p_0} (M_c f)^{(p_0-p)p/p_0} \\
&\leq \Big(\int |Tf|^{p_0} w (M_c f)^{p-p_0} \Big)^{p/p_0} \Big(\int (M_c f)^p w \Big)^{1-p/p_0} \\
&\lesssim \Big(\int |f|^{p_0} w (M_c f)^{p-p_0} \Big)^{p/p_0} \Big(\int |f|^p w \Big)^{1-p/p_0} \\
&\leq \Big(\int |f|^{p_0} w |f|^{p-p_0} \Big)^{p/p_0} \Big(\int |f|^p w \Big)^{1-p/p_0} = \int |f|^p w,
\end{aligned}$$

where the second inequality uses the strong (p_0, p_0) estimate for T with respect to $w \cdot (M_c f)^{p-p_0} \in A_{p_0}$ and the strong (p, p) estimate for M_c with respect to $w \in A_1 \subset A_p$, and the third inequality uses the fact that $f(x) \leq M_c f(x)$ a.e. and that $p - p_0 < 0$. This shows that T is of strong type (p, p) with respect to w.

Now assume that $w \in A_p$ for some $1 < p < \infty$. We will show that T is of strong type (p, p) with respect to w.

By the self-improving property of A_p weights (Corollary 6.14), there exists some $1 < q < p$ such that $w \in A_{p/q}$. Without loss of generality, assume that $1 < q < p_0$.

By duality, there exists $u \in L^{(p/q)'}(w)$ of norm 1 such that

$$\left(\int_{\mathbb{R}^n} |Tf|^p w\right)^{q/p} = \left(\int_{\mathbb{R}^n} (|Tf|^q)^{p/q} w\right)^{q/p} = \int_{\mathbb{R}^n} |Tf|^q uw. \tag{6.16}$$

We claim that for $a > 1$ small enough we have $M_c(|uw|^a) < \infty$ a.e. Indeed, since $w \in A_{p/q}$, we have $w^{1-(p/q)'} \in A_{(p/q)'}$ by Lemma 6.8 (iv). By the self-improving property of A_p weights, $w^{1-(p/q)'} \in A_{(p/q)'/a}$ for $a > 1$ small enough. But then by Theorem 6.12

$$\int M_c(|uw|^a)^{(p/q)'/a} w^{1-(p/q)'} \lesssim \int |uw|^{(p/q)'} w^{1-(p/q)'} = \int |u|^{(p/q)'} w = 1, \tag{6.17}$$

and so in particular $M_c(|uw|^a) < \infty$ a.e.

By Proposition 6.15, $M_c(|uw|^a)^{1/a}$ is an A_1 weight. Thus, we know by the first half of the proof that T is of strong type (q,q) with respect to $M_c(|uw|^a)^{1/a}$. Since $|uw| \leq M_c(|uw|^a)^{1/a}$, it follows that

$$\begin{aligned}
\int_{\mathbb{R}^n} |Tf|^q uw \leq \int_{\mathbb{R}^n} |Tf|^q M_c(|uw|^a)^{1/a} &\lesssim \int_{\mathbb{R}^n} |f|^q M_c(|uw|^a)^{1/a} \\
&= \int_{\mathbb{R}^n} |f|^q w^{q/p} M_c(|uw|^a)^{1/a} w^{-q/p} \\
\leq \left(\int |f|^p w\right)^{q/p} &\left(\int M_c(|uw|^a)^{(p/q)'/a} w^{1-(p/q)'}\right)^{1/(p/q)'} \overset{(6.17)}{\lesssim} \left(\int |f|^p w\right)^{q/p}.
\end{aligned}$$

Together with (6.16), this shows that T is of strong type (p,p) with respect to w. ☐

For some applications of the extrapolation theorem, and for more information about the theory of A_p weights, see e.g. the lecture notes [CU17].

7 Sparse domination and the A_2 theorem

In this course we have not paid much attention to the constants appearing in the estimates we proved. This is about to change.

It has been known for a long time that if $w \in A_p$ and T is a Calderón-Zygmund operator on $\mathbb{R}^n$, then T is of strong type (p,p) with respect to w, i.e.,

$$\|Tf\|_{L^p(w)} \leq C\|f\|_{L^p(w)},$$

with C depending on p, n, T, and $[w]_{A_p}$, see e.g. Section 7.4 in [Duo01]. Due to certain applications in PDEs, the precise dependence of C on $[w]_{A_p}$ was of interest, see [FKP91, AIS01, PV02]. It was conjectured that for $p = 2$ the dependence was linear, so that

$$\|Tf\|_{L^2(w)} \leq C[w]_{A_2}\|f\|_{L^2(w)}, \tag{7.1}$$

with $C = C(n,T)$. This estimate came to be known as the A_2 *conjecture*, and after many partial results it was finally confirmed in 2012 by Hytönen [Hyt12]. The proof of Hytönen was quite complicated, and we will follow a much simpler proof due to Lerner [Ler16], which uses the *sparse domination* technique.

The following exercise demonstrates that the estimate (7.1) is sharp, in the sense that it would be false if we replaced $[w]_{A_2}$ by $[w]_{A_2}^s$ for any $s < 1$.

Exercise 7.1. For any $s \in (0,1)$ let $w = |x|^{1-s}$ be a weight on $\mathbb{R}$.

(i) Show that $w \in A_2$, and $[w]_{A_2} \leq s^{-1}$.

(ii) Given $f_s(x) = x^{s-1}\mathbf{1}_{(0,1)}(x)$, show that $\|f_s\|_{L^2(w_s)} \leq s^{-1/2}$.

(iii) Prove that $\|Hf_s\|_{L^2(w_s)} \geq Cs^{-3/2}$, and conclude that (7.1) is sharp.

Remark 7.2. By the A_2 theorem, we get that all Calderón-Zygmund operators are of strong type $(2,2)$ with respect to all A_2 weights. By the extrapolation of A_p weights (Theorem 6.19), it follows that Calderón-Zygmund operators are of strong type (p,p) for all A_p weights, $1 < p < \infty$. Moreover, by a sharp version of the extrapolation theorem [DGPP05] one can get the sharp estimate

$$\|Tf\|_{L^p(w)} \leq C[w]_{A_p}^{\max(1,1/(p-1))}\|f\|_{L^p(w)}.$$

The same estimate can be obtained directly from Lerner's proof we will present, but for the sake of simplicity we will restrict attention to $p = 2$.

7.1 Sparse and Carleson families

Definition 7.3 (sparse family). Let $0 < \eta \le 1$. We will say that a family of cubes $\mathcal{S}$ is *η-sparse* if for every $Q \in \mathcal{S}$ there exists a measurable subset $E_Q \subset Q$ such that $|E_Q| \ge \eta|Q|$ and $\{E_Q\}_{Q\in\mathcal{S}}$ are pairwise disjoint.

Note that in the definition above we do not require the cubes in $\mathcal{S}$ to be dyadic, although this will often be the case.

Example 7.4. Any disjoint family of cubes is 1-sparse (just take $E_Q = Q$).

Example 7.5. If $k < l$, then $\mathcal{S} = \mathcal{D}_k \cup \mathcal{D}_l$ is 1/2-sparse. To see this, for each $Q \in \mathcal{D}_l$ let E_Q be the lower half of Q, and for each $P \in \mathcal{D}_k$ let

$$E_P := P \setminus \bigcup_{Q\in\mathcal{D}_l} E_Q.$$

More generally, $\mathcal{S} = \mathcal{D}_{k_1} \cup \cdots \cup \mathcal{D}_{k_m}$ is $1/m$-sparse, and we leave the proof as an exercise.

In the examples above the cubes from sparse families had only bounded intersection, in the sense that $\sum_{Q\in\mathcal{S}} \mathbf{1}_Q \in L^\infty(\mathbb{R}^n)$. The following exercise shows that this does not need to be the case.

Exercise 7.6. Prove that in $\mathbb{R}^n$ the family of all dyadic cubes containing 0 is $(1 - 2^{-n})$-sparse.

For dyadic cubes, the notion of sparseness is equivalent to the *Carleson packing condition*, which is also widely used in harmonic analysis. We will not need this fact for the proof of the A_2 conjecture, but it helps to gain intuition regarding sparse families.

Definition 7.7 (Carleson family). Let $\mathcal{S} \subset \mathcal{D}$. We say that $\mathcal{S}$ is Λ-Carlson if for every $R \in \mathcal{D}$

$$\sum_{Q\in\mathcal{S},\, Q\subset R} |Q| \le \Lambda|R|. \tag{7.2}$$

The following result is due to Lerner and Nazarov [LN19].

Proposition 7.8. *Let $\mathcal{S} \subset \mathcal{D}$. Then, $\mathcal{S}$ is η-sparse if and only if it is η^{-1}-Carleson.*

Proof. One of the implications is very easy. If $\mathcal{S}$ is η-sparse, then for any $R \in \mathcal{D}$

$$\sum_{Q\in\mathcal{S},\, Q\subset R} |Q| \le \eta^{-1} \sum_{Q\in\mathcal{S},\, Q\subset R} |E_Q| \le \eta^{-1}|R|,$$

where in the second inequality we used the fact that E_Q are pairwise disjoint.

Proving the converse inequality is more laboursome. Suppose that $\mathcal{S}$ is η^{-1}-Carleson. If we knew that $\mathcal{S} \subset \bigcup_{k\le K} \mathcal{D}_k$ for some $K \in \mathbb{Z}$, that is, there is a "bottom layer" of $\mathcal{S}$, then we could argue inductively as follows.

For every $Q \in \mathcal{S} \cap \mathcal{D}_K$ let E_Q be an arbitrary measurable subset of Q with $|E_Q| = \eta|Q|$. If E_Q has already been defined for $Q \in \mathcal{S}_K^N := \mathcal{S} \cap \bigcup_{N \le k \le K} \mathcal{D}_k$, then we define $E_Q \subset Q$ for $Q \in \mathcal{S} \cap \mathcal{D}_{N-1}$ as an arbitrary measurable subset of

$$Q \setminus \bigcup_{P \in \mathcal{S},\, P \subsetneq Q} E_P$$

such that $|E_Q| = \eta|Q|$. It is possible to find such a set because

$$\begin{aligned} \Bigg| Q \setminus \bigcup_{P \in \mathcal{S},\, P \subsetneq Q} E_P \Bigg| &= |Q| - \sum_{P \in \mathcal{S},\, P \subsetneq Q} |E_P| \ge |Q| - \eta \sum_{P \in \mathcal{S},\, P \subsetneq Q} |P| \\ &= (1+\eta)|Q| - \eta \sum_{P \in \mathcal{S},\, P \subset Q} |P| \ge \eta|Q|, \end{aligned} \tag{7.3}$$

where in the last inequality we used the η^{-1}-Carleson condition for $\mathcal{S}$. Using this "upwards" induction, we eventually define E_Q for every $Q \in \mathcal{S}$, and it is easy to see that $\{E_Q\}_{Q \in \mathcal{S}}$ are pairwise disjoint.

In the absence of the "bottom layer", we have to artificially introduce it. Given an integer $K \in \mathbb{Z}$, let $\mathcal{S}_K := \mathcal{S} \cap \bigcup_{k \le K} \mathcal{D}_k$. For all $Q \in \mathcal{S}_K$ we could define sets E_Q as above, but since we would like to take a limit $K \to \infty$, we have to be more careful than that.

For every $Q \in \mathcal{S} \cap \mathcal{D}_K$ let

$$E_Q^K := \eta^{1/n} Q,$$

so that $|E_Q| = \eta|Q|$ (recall that CQ denotes the cube with the same center as Q and with sidelength $C\ell(Q)$).

If E_Q^K has already been defined for $Q \in \mathcal{S}_K^N$, then for $Q \in \mathcal{S} \cap \mathcal{D}_{N-1}$ we define

$$F_Q^K := \bigcup_{P \in \mathcal{S}_K,\, P \subsetneq Q} E_P^K,$$

and

$$E_Q^K := tQ \setminus F_Q^K,$$

where $t \in (0,1]$ is the largest number such that $|E_Q^K| = \eta|Q|$. To see that such t exists, note that by (7.3) we have $|Q \setminus F_Q^K| \ge \eta|Q|$. The function $f(t) := |tQ \setminus F_Q^K|$ is continuous, monotone, and since $f(1) \ge \eta|Q|$ and $f(0) = 0$, we get $f(t) = \eta|Q|$ for some $t \in (0,1]$. We denote by t_Q^K the largest such t.

Thus, we have $E_Q^K \subset Q$ and $F_Q^K \subset Q$ defined for all $Q \in \mathcal{S}_K$. By definition, the sets $\{E_Q^K\}_{Q \in \mathcal{S}_K}$ are pairwise disjoint and $|E_Q^K| = \eta|Q|$.

Let $G_Q^K = E_Q^K \cup F_Q^K$. We claim that for any $Q \in \mathcal{S}_K$ we have

$$G_Q^K \subset G_Q^{K+1}. \tag{7.4}$$

First, assume $Q \in \mathcal{D}_K$. Then, we have $G_Q^K = E_Q^K = \eta^{1/n} Q$. On the other hand,

$$G_Q^{K+1} = tQ \cup F_Q^{K+1},$$

where $t = t_Q^{K+1}$ is such that $|tQ \setminus F_Q^{K+1}| = \eta |Q|$. In particular, $t \geq \eta^{1/n}$, so that (7.4) holds for $Q \in \mathcal{S} \cap \mathcal{D}_K$.

Now we proceed by induction. Suppose (7.4) holds for $P \in \mathcal{S}_K^N$, and let $Q \in \mathcal{S} \cap \mathcal{D}_{N-1}$. Observe that

$$F_Q^K = \bigcup_{P \in \mathcal{S}_K,\, P \subsetneq Q} E_P^K = \bigcup_{P \in \mathcal{S}_K,\, P \subsetneq Q} G_P^K,$$

and so by the inductive assumption (7.4) we have

$$F_Q^K \subset F_Q^{K+1} \tag{7.5}$$

Recall that $t_Q^K \in (0,1]$ is the largest number such that $|t_Q^K Q \setminus F_Q^K| = \eta |Q|$. By (7.5)

$$|tQ \setminus F_Q^{K+1}| \leq |tQ \setminus F_Q^K|,$$

and so

$$t_Q^{K+1} \geq t_Q^K. \tag{7.6}$$

Hence, (7.4) holds for Q, and this closes the induction.

Now, fix $Q \in \mathcal{S}$. By (7.6), $\{t_Q^K\}_K \subset (0,1]$ is a non-decreasing sequence, and so the limit

$$t_Q \coloneqq \lim_{K \to \infty} t_Q^K$$

exists, and $t_Q \in (0,1]$. At the same time, by (7.5) the sets F_Q^K are increasing in K. We define

$$F_Q \coloneqq \bigcup_{K=0}^{\infty} F_Q^K, \qquad E_Q \coloneqq t_Q Q \setminus F_Q.$$

Note that

$$|E_Q| = \left| \bigcap_{K=0}^{\infty} t_Q Q \setminus F_Q^K \right| = \lim_{K \to \infty} |t_Q Q \setminus F_Q^K| \geq \lim_{K \to \infty} |t_Q^K Q \setminus F_Q^K| = \eta |Q|,$$

where we used the definition of t_Q^K and the fact that $t_Q \geq t_Q^K$ for all K.

It remains to show that $\{E_Q\}_{Q \in \mathcal{S}}$ are pairwise disjoint. Let $Q, P \in \mathcal{S}$, and without loss of generality assume that $Q \subsetneq P$. Then, $E_Q \subset F_P$ because

$$E_Q \subset t_Q Q = \bigcup_{K=0}^{\infty} t_Q^K Q \subset \bigcup_{K=0}^{\infty} E_Q^K \cup F_Q^K \subset \bigcup_{K=0}^{\infty} F_P^K = F_P.$$

Recalling that $E_P \cap F_P = \varnothing$, we get $E_Q \cap E_P = \varnothing$. $\square$

Remark 7.9. The assumption "$\mathcal{S} \subset \mathcal{D}$" in Proposition 7.8 can be omitted if one defines the Carleson condition for families of non-dyadic cubes properly. See [Hän18] for an extension of Proposition 7.8 to general families of Borel sets.

Proposition 7.8 can be used to prove the following.

Exercise 7.10. Suppose that $\mathcal{S}_1, \dots, \mathcal{S}_k$ are sparse families of dyadic cubes, and that each $\mathcal{S}_j$ is η_j-sparse for some $\eta_j \in (0,1]$. Show that $\mathcal{S}_1 \cup \dots \cup \mathcal{S}_k$ is $1/(\sum_{j=1}^k \eta_j^{-1})$-sparse.

7.2 Sparse operators

For the sake of brevity, for $f \in L^1_{loc}(\mathbb{R}^n)$ and $Q \subset \mathbb{R}^n$ we will write

$$f_Q \coloneqq \frac{1}{|Q|} \int_Q f.$$

Definition 7.11 (sparse operator). If $\mathcal{S}$ is a sparse family of cubes, we define the associated *sparse operator* as

$$\mathcal{A}_{\mathcal{S}} f \coloneqq \sum_{Q \in \mathcal{S}} f_Q \mathbf{1}_Q.$$

In the next subsection we will prove the following *sparse domination* result.

Theorem 7.12. *Let T be a Calderón-Zygmund theorem. Then, for every compactly supported $f \in L^1(\mathbb{R}^n)$ there exists an η-sparse family of cubes $\mathcal{S}$ such that*

$$|Tf(x)| \leq C \mathcal{A}_{\mathcal{S}} |f|(x) \quad \text{for a.e. } x \in \mathbb{R}^n,$$

with $C = C(n,T)$ and $\eta = \eta(n)$.

It is not too difficult to show that sparse operators satisfy the weighted inequality postulated by the A_2 conjecture.

Proposition 7.13. *If $\mathcal{S}$ is an η-sparse family of cubes, then for any $w \in A_2$ and $f \in L^2(w)$*

$$\|\mathcal{A}_{\mathcal{S}} f\|_{L^2(w)} \leq C [w]_{A_2} \|f\|_{L^2(w)},$$

with $C = C(n, \eta)$

Together with Theorem 7.12, we easily get the A_2 conjecture:

Proof of the A_2 conjecture. By Theorem 7.12 and Proposition 7.13, for any compactly supported $f \in L^1(\mathbb{R}^n) \cap L^2(w)$ we have a sparse family $\mathcal{S}$ such that

$$\|Tf\|_{L^2(w)} \leq C \|\mathcal{A}_{\mathcal{S}} |f|\|_{L^2(w)} \leq C [w]_{A_2} \|f\|_{L^2(w)},$$

and so (7.1) holds for such functions. The case of general $f \in L^2(w)$ follows by the density of compactly supported functions from $L^1(\mathbb{R}^n) \cap L^2(w)$ in $L^2(w)$ (see Exercise 6.7). $\square$

Now we prove Proposition 7.13. In the proof we use the following variant of the Hardy-Littlewood maximal operator: for a Radon measure μ and $f \in L^1_{loc}(\mu)$ let

$$M^\mu f(x) = \sup_{r>0} \frac{1}{\mu(Q(x,r))} \int_{Q(x,r)} |f| \, d\mu,$$

where $Q(x,r)$ denotes the cube centered at x of sidelength r. It follows from the *Besicovitch covering theorem* that M^μ if of weak type $(1,1)$ on $(\mathbb{R}^n, \mu)$, and for $1 < p \leq \infty$ it is of strong type (p,p) on $(\mathbb{R}^n, \mu)$, with estimates depending only on n and p. See Theorem 2.19 in [Mat95], or Theorem 4.35 in [Par20].

Proof of Proposition 7.13. Let $E_Q \subset Q$ be the disjoint subsets from the definition of sparse families. By duality and Cauchy-Schwarz inequality, we have

$$\begin{aligned} \|\mathcal{A}_\mathcal{S} f\|_{L^2(w)} &= \sup_{\|g\|_{L^2(w)} \leq 1} \int \mathcal{A}_\mathcal{S} f(x) g(x) w(x) \, dx = \sup_{\|g\|_{L^2(w)} \leq 1} \sum_{Q \in \mathcal{S}} f_Q \int_Q gw \\ &\leq \sup_{\|g\|_{L^2(w)}} \left(\sum_{Q \in \mathcal{S}} |w(3Q) f_Q|^2 w(E_Q)^{-1} \right)^{1/2} \left(\sum_{Q \in \mathcal{S}} \left(\frac{1}{w(3Q)} \int_Q gw \right)^2 w(E_Q) \right)^{1/2} \end{aligned} \tag{7.7}$$

Let $Q \in \mathcal{S}$ and $x \in Q$. Observe that $Q \subset Q(x, 2\ell(Q)) \subset 3Q$. Hence,

$$\frac{1}{w(3Q)} \int_Q gw \leq \frac{1}{w(Q(x, 2\ell(Q)))} \int_{Q(x, 2\ell(Q))} gw \leq M^w g(x).$$

We have

$$\sum_{Q \in \mathcal{S}} \left(\frac{1}{w(3Q)} \int_Q gw \right)^2 w(E_Q) \leq \sum_{Q \in \mathcal{S}} \int_{E_Q} |M^w g|^2 w \leq \int_{\mathbb{R}^n} |M^w g|^2 w \leq C_n \|g\|^2_{L^2(w)}, \tag{7.8}$$

where in the second inequality we used that E_Q are disjoint, and in the last inequality we used the strong type $(2,2)$ estimates for M^w with respect to w.

Let $\sigma = w^{-1} \in A_2$. Then,

$$\sum_{Q \in \mathcal{S}} |w(3Q) f_Q|^2 w(E_Q)^{-1} = \sum_{Q \in \mathcal{S}} \left(\frac{1}{\sigma(3Q)} \int_Q f w \sigma \right)^2 \sigma(E_Q) \cdot \left(\frac{w(3Q)^2}{w(E_Q)} \frac{\sigma(3Q)^2}{\sigma(E_Q)} \frac{1}{|Q|^2} \right).$$

We claim that

$$\sup_{Q \in \mathcal{S}} \frac{w(3Q)^2}{w(E_Q)} \frac{\sigma(3Q)^2}{\sigma(E_Q)} \frac{1}{|Q|^2} \leq C_{n,\eta} [w]^2_{A_2}. \tag{7.9}$$

Assuming for the moment that this is the case, we can argue as in (7.8) (just swapping σ for w and fw for g) to get that

$$\sum_{Q \in \mathcal{S}} |w(3Q) f_Q|^2 w(E_Q)^{-1} \leq C_{n,\eta} [w]^2_{A_2} \|fw\|^2_{L^2(\sigma)} = C_{n,\eta} [w]^2_{A_2} \|f\|^2_{L^2(w)}.$$

Together with (7.8) and (7.7), this gives $\|\mathcal{A}_{\mathcal{S}} f\|_{L^2(w)} \leq C[w]_{A_2}\|f\|_{L^2(w)}$, as desired.

It remains to prove (7.9). By the A_2 condition for w,

$$\frac{w(3Q)}{|Q|}\frac{\sigma(3Q)}{|Q|} = 3^{2n}\frac{w(3Q)}{|3Q|}\frac{\sigma(3Q)}{|3Q|} \leq 3^{2n}[w]_{A_2},$$

so it suffices to show that for all $Q \in \mathcal{S}$

$$\frac{|Q|^2}{w(E_Q)\sigma(E_Q)} \leq C_{n,\eta}.$$

By Cauchy-Schwarz inequality,

$$|Q| \leq \eta^{-1}|E_Q| = \eta^{-1}\int_{E_Q} w^{1/2}\sigma^{1/2} \leq \eta^{-1}w(E_Q)^{1/2}\sigma(E_Q)^{1/2},$$

which finishes the proof. □

7.3 Auxiliary maximal operators

In the proof of Theorem 7.12 we will use a few auxiliary maximal operators. Define

$$\mathcal{M}_T f(x) = \sup_{Q \ni x} \operatorname*{ess\,sup}_{\xi \in Q} |T(f\mathbf{1}_{3Q^c})(\xi)|,$$

where the sup is taken over all cubes Q containing x. Compare $\mathcal{M}_T$ to the usual maximal operator associated to T, which we introduced in Section 5:

$$T_* f(x) = \sup_{\varepsilon>0}\left|\int_{|x-y|>\varepsilon} K(x,y)f(y)\,dy\right| = \sup_{\varepsilon>0}|T(f\mathbf{1}_{B(x,\varepsilon)^c})(x)|.$$

We define also a local version of $\mathcal{M}_T$. Given a cube Q_0, for $x \in Q_0$ we define

$$\mathcal{M}_{T,Q_0} f(x) = \sup_{x \in Q \subset Q_0} \operatorname*{ess\,sup}_{\xi \in Q} |T(f\mathbf{1}_{3Q_0 \setminus 3Q})(\xi)|.$$

Remark that

$$\mathcal{M}_{T,Q_0} f(x) = \sup_{x \in Q \subset Q_0} \operatorname*{ess\,sup}_{\xi \in Q} |T(f\mathbf{1}_{3Q^c})(\xi) - T(f\mathbf{1}_{3Q_0^c})(\xi)| \leq 2\mathcal{M}_T f(x). \quad (7.10)$$

In the remainder of this section the constant C may depend on the operator T and dimension n.

Lemma 7.14. *If $f \in L^1(\mathbb{R}^n)$, then for all $x \in \mathbb{R}^n$*

$$\mathcal{M}_T f(x) \leq CMf(x) + T_* f(x). \quad (7.11)$$

In particular, $\mathcal{M}_T$ is of weak type $(1,1)$.

Proof. Let $x \in \mathbb{R}^n$, $Q \ni x$ be a cube, and $\xi \in Q$. Set $B_x = B(x, 2\operatorname{diam} Q)$, so that $3Q \subset B_x$. Then, by linearity of T

$$|T(f\mathbf{1}_{3Q^c})(\xi)| \leq |T(f\mathbf{1}_{B_x^c})(\xi) - T(f\mathbf{1}_{B_x^c})(x)| + |T(f\mathbf{1}_{B_x\setminus 3Q})(\xi)| + |T(f\mathbf{1}_{B_x^c})(x)|. \tag{7.12}$$

The third term satisfies $|T(f\mathbf{1}_{B_x^c})(x)| \leq T_* f(x)$. Regarding the second term, by the size condition of the kernel

$$\begin{aligned}|T(f\mathbf{1}_{B_x\setminus 3Q})(\xi)| = \left|\int_{B_x\setminus 3Q} K(\xi,y)f(y)\,dy\right| &\leq \int_{B_x\setminus 3Q}\frac{1}{|\xi-y|^n}|f(y)|\,dy\\ &\leq C\frac{1}{|B_x|}\int_{B_x}|f| \leq CMf(x).\end{aligned}$$

Finally, we estimate the first term from the RHS of (7.12) using the smoothness of the kernel:

$$\begin{aligned}|T(f\mathbf{1}_{B_x^c})(\xi) - T(f\mathbf{1}_{B_x^c})(x)| &= \left|\int_{B_x^c}(K(\xi,y)-K(x,y))f(y)\,dy\right|\\ &\leq C\int_{B_x^c}\frac{|x-\xi|^\delta}{|x-y|^{n+\delta}}|f(y)|\,dy \leq C\int_{B_x^c}\frac{\ell(Q)^\delta}{|x-y|^{n+\delta}}|f(y)|\,dy\\ &= C\ell(Q)^\delta\sum_{k\geq 0}\int_{2^{k+1}B_x\setminus 2^k B_x}\frac{1}{|x-y|^{n+\delta}}|f(y)|\,dy\\ &\leq C\ell(Q)^\delta\sum_{k\geq 0}(2^k\ell(Q))^{-n-\delta}\int_{2^{k+1}B_x}|f(y)|\,dy \leq CMf(x).\end{aligned}$$

This shows (7.11). The weak $(1,1)$ estimate for $\mathcal{M}_T$ follows from (7.11) and weak $(1,1)$ estimates for M and T_*. ☐

Recall that for every Lebesgue measurable $f : \mathbb{R}^n \to \mathbb{R}$ a.e. $x \in \mathbb{R}^n$ is a point of *approximate continuity*, which means that for every $\varepsilon > 0$

$$\lim_{r\to 0}\frac{|\{y \in B(x,r) : |f(y)-f(x)| < \varepsilon\}|}{|B(x,r)|} = 1,$$

see Section 1.7.2 in [EG91].

Lemma 7.15. *If $f \in L^1(\mathbb{R}^n)$, then for a.e. $x \in Q_0$*

$$|T(f\mathbf{1}_{3Q_0})(x)| \leq C|f(x)| + \mathcal{M}_{T,Q_0}f(x). \tag{7.13}$$

Proof. Let $x \in \operatorname{int}(Q_0)$ be a Lebesgue point for f, and a point of approximate continuity for $T(f\mathbf{1}_{3Q_0})$. Fix $\varepsilon > 0$, so that

$$E(x,r) = \{y \in B(x,r) : |T(f\mathbf{1}_{3Q_0})(y) - T(f\mathbf{1}_{3Q_0})(x)| < \varepsilon\}$$

satisfies $\lim_{r\to 0}|E(x,r)|/|B(x,r)| = 1$.

Note that for every $r > 0$ we have $B(x,r) \subset Q(x,2r)$, where $Q(x,2r)$ is the cube centered at x of sidelength $2r$. Let $r > 0$ be so small that $Q(x,2r) \subset Q_0$. Then, for a.e. $y \in E(x,r)$

$$|T(f\mathbf{1}_{3Q_0})(x)| \le |T(f\mathbf{1}_{3Q_0})(y)| + \varepsilon = |T(f\mathbf{1}_{3Q_0\setminus 3Q(x,2r)})(y) + T(f\mathbf{1}_{3Q(x,2r)})(y)| + \varepsilon$$
$$\le \mathcal{M}_{T,Q_0}(x) + |T(f\mathbf{1}_{3Q(x,2r)})(y)| + \varepsilon,$$

and so

$$|T(f\mathbf{1}_{3Q_0})(x)| \le \mathcal{M}_{T,Q_0}(x) + \operatorname*{ess\,inf}_{\xi\in E(x,r)} |T(f\mathbf{1}_{3Q(x,2r)})(\xi)| + \varepsilon.$$

Note that by the weak $(1,1)$ estimate for T

$$|E(x,r)| \le \left|\{y\in\mathbb{R}^n : |T(f\mathbf{1}_{3Q(x,2r)})(y)| \ge \operatorname*{ess\,inf}_{\xi\in E(x,r)} |T(f\mathbf{1}_{3Q(x,2r)})(\xi)|\}\right|$$
$$\le \frac{C}{\operatorname{ess\,inf}_{\xi\in E(x,r)} |T(f\mathbf{1}_{3Q(x,2r)})(\xi)|}\int_{3Q(x,2r)} |f|.$$

Hence,

$$|T(f\mathbf{1}_{3Q_0})(x)| \le \frac{C}{|E(x,r)|}\int_{3Q(x,2r)} |f| + \mathcal{M}_{T,Q_0}(x) + \varepsilon$$
$$\xrightarrow{r\to 0} C'|f(x)| + \mathcal{M}_{T,Q_0}(x) + \varepsilon,$$

where we used $\lim_{r\to 0}|E(x,r)|/|B(x,r)| = 1$ and that x is a Lebesgue point of f. Taking $\varepsilon \to 0$ finishes the proof of (7.13). $\square$

7.4 Sparse domination of Calderón-Zygmund operators

In this subsection we prove Theorem 7.12, which we recall below.

Theorem. *Let T be a Calderón-Zygmund theorem. Then, for every compactly supported $f \in L^1(\mathbb{R}^n)$ there exists an η-sparse family of cubes $\mathcal{S}$ such that*

$$|Tf(x)| \le C\mathcal{A}_{\mathcal{S}}|f|(x) \quad \textit{for a.e. } x \in \mathbb{R}^n, \tag{7.14}$$

with $C = C(n,T)$ and $\eta = \eta(n)$.

We begin by proving the following local version of (7.14).

Lemma 7.16. *For any cube Q_0 there exists a $\frac{1}{2}$-sparse family $\mathcal{S}$ of cubes contained in Q_0 such that*

$$|T(f\mathbf{1}_{3Q_0})(x)| \le C \sum_{Q\in\mathcal{S}} |f|_{3Q}\mathbf{1}_Q(x) \quad \textit{for a.e. } x \in Q_0. \tag{7.15}$$

Proof. Without loss of generality assume that $Q_0 \in \mathcal{D}$. Denote by $\mathcal{D}(Q_0)$ the dyadic subcubes of Q_0. To prove (7.15), it suffices to prove the following recursive estimate: for any $Q_0 \in \mathcal{D}$ there exists a family $\mathcal{F} = \mathcal{F}(Q_0) \subset \mathcal{D}(Q_0)$ of pairwise disjoint cubes such that $\sum_{P\in\mathcal{F}(Q_0)} |P| \le \frac{1}{2}|Q_0|$ and

$$|T(f\mathbf{1}_{3Q_0})(x)|\mathbf{1}_{Q_0}(x) \le C|f|_{3Q_0}\mathbf{1}_{Q_0}(x) + \sum_{P\in\mathcal{F}(Q_0)} |T(f\mathbf{1}_{3P})(x)|\mathbf{1}_P(x) \tag{7.16}$$

for a.e. $x \in Q_0$. Indeed, to prove (7.15) it suffices to iterate the estimate (7.16). Set $\mathcal{S}_0 := \{Q_0\}$, and if $\mathcal{S}_k$ has already been defined, we set

$$\mathcal{S}_{k+1} := \bigcup_{P\in\mathcal{S}_k} \mathcal{F}(P)$$

and $\mathcal{S} := \bigcup_{k\ge 0}\mathcal{S}_k$. To see that $\mathcal{S}$ is $\frac{1}{2}$-sparse, just take

$$E_Q := Q \setminus \bigcup_{P\in\mathcal{S},\, P\subsetneq Q} P = Q \setminus \bigcup_{P\in\mathcal{F}(Q)} P$$

for any $Q \in \mathcal{S}$, so that $|E_Q| \ge \frac{1}{2}|Q|$. Iterating (7.16), we get for a.e. $x \in Q_0$

$$\begin{aligned}
|T(f\mathbf{1}_{3Q_0})|\mathbf{1}_{Q_0}(x) &\le C|f|_{3Q_0} + \sum_{P\in\mathcal{S}_1} |T(f\mathbf{1}_{3P})(x)|\mathbf{1}_P(x) \\
&\le C|f|_{3Q_0} + \sum_{P\in\mathcal{S}_1} C|f|_{3P} + \sum_{P\in\mathcal{S}_1}\sum_{Q\in\mathcal{F}(P)} |T(f\mathbf{1}_{3Q})(x)|\mathbf{1}_Q(x) \\
&= C|f|_{3Q_0} + \sum_{P\in\mathcal{S}_0} C|f|_{3P} + \sum_{P\in\mathcal{S}_2} |T(f\mathbf{1}_{3P})(x)|\mathbf{1}_P(x) \\
&\le \sum_{j=0}^{k}\sum_{P\in\mathcal{S}_j} C|f|_{3P} + \sum_{P\in\mathcal{S}_{k+1}} |T(f\mathbf{1}_{3P})(x)|\mathbf{1}_P(x)
\end{aligned}$$

for any $k \geq 0$. Note that $\sum_{P\in\mathcal{S}_k} |P| \leq \frac{1}{2}\sum_{P\in\mathcal{S}_{k-1}} |P| \leq 2^{-k}|Q_0|$, and so

$$\sum_{P\in\mathcal{S}_{k+1}} |T(f\mathbf{1}_{3P})(x)|\mathbf{1}_P(x) \xrightarrow{k\to 0} 0 \quad \text{for a.e. } x \in Q_0.$$

Thus, taking $k \to \infty$ in the previous estimate yields (7.15) for a.e. $x \in Q_0$.

Now our goal is to establish (7.16). First, observe that for any family $\mathcal{F} \subset \mathcal{D}(Q_0)$ of pairwise disjoint cubes we have

$$\begin{aligned}|T(f\mathbf{1}_{3Q_0})|\mathbf{1}_{Q_0} &= |T(f\mathbf{1}_{3Q_0})|\mathbf{1}_{Q_0\setminus\bigcup_{P\in\mathcal{F}} P} + \sum_{P\in\mathcal{F}} |T(f\mathbf{1}_{3Q_0})|\mathbf{1}_P \\ &\leq |T(f\mathbf{1}_{3Q_0})|\mathbf{1}_{Q_0\setminus\bigcup_{P\in\mathcal{F}} P} + \sum_{P\in\mathcal{F}} |T(f\mathbf{1}_{3Q_0\setminus 3P})|\mathbf{1}_P + \sum_{P\in\mathcal{F}} |T(f\mathbf{1}_{3P})|\mathbf{1}_P.\end{aligned}$$

Hence, to prove (7.16) we need to find a disjoint family $\mathcal{F} \subset \mathcal{D}(Q_0)$ such that $\sum_{P\in\mathcal{F}} |P| \leq \frac{1}{2}|Q_0|$ and

$$|T(f\mathbf{1}_{3Q_0})|\mathbf{1}_{Q_0\setminus\bigcup_{P\in\mathcal{F}} P} + \sum_{P\in\mathcal{F}} |T(f\mathbf{1}_{3Q_0\setminus 3P})|\mathbf{1}_P \leq C|f|_{3Q_0}. \tag{7.17}$$

Recall that by Lemma 7.14 the maximal operator $\mathcal{M}_T$ is weak $(1,1)$, with estimates depending only on n and T. Using also that $\mathcal{M}_{T,Q_0}f \leq 2\mathcal{M}_T f$ (7.10), we get that if $C_0 = C_0(n,T)$ is chosen large enough, then

$$E \coloneqq \{x \in Q_0 : |f(x)| > C_0|f|_{3Q_0}\} \cup \{x \in Q_0 : \mathcal{M}_{T,Q_0}f(x) > C_0|f|_{3Q_0}\}$$

satisfies

$$|E| \leq \frac{\int_{Q_0}|f|}{C_0|f|_{3Q_0}} + \frac{C\int_{Q_0}|f|}{C_0|f|_{3Q_0}} \leq \frac{1}{2^{n+2}}|Q_0|.$$

We apply the Calderón-Zygmund decomposition to function $\mathbf{1}_E$ at level $\lambda = 2^{-n-1}$ to get a collection of disjoint dyadic cubes $\mathcal{B} \subset \mathcal{D}(Q_0)$ such that for every $P \in \mathcal{B}$

$$\lambda \leq \frac{1}{|P|}\int_P \mathbf{1}_E \leq 2^n\lambda,$$

see Proposition 4.15. This is equivalent to

$$\frac{|P|}{2^{n+1}} \leq |E\cap P| \leq \frac{|P|}{2}. \tag{7.18}$$

At the same time, recall that for a.e. x outside of $\bigcup_{P\in\mathcal{B}} P$ we have $\mathbf{1}_E(x) \leq \lambda = 2^{-n-1}$. This means that $|E\setminus\bigcup_{P\in\mathcal{B}} P| = 0$. It follows that

$$\sum_{P\in\mathcal{B}} |P| \leq 2^{n+1}\sum_{P\in\mathcal{B}} |E\cap P| = 2^{n+1}|E| \leq \frac{2^{n+1}}{2^{n+2}}|Q_0| = \frac{1}{2}|Q_0|.$$

We set $\mathcal{F} = \mathcal{B}$.

Note that by (7.18) for every $P \in \mathcal{F}$ we have $P \cap E^c \neq \varnothing$. Thus, there exists $x \in P$ such that

$$\mathcal{M}_{T,Q_0} f(x) \leq C_0 |f|_{3Q_0},$$

and so

$$\operatorname*{ess\,sup}_{\xi \in P} |T(f \mathbf{1}_{3Q_0 \setminus 3P})(\xi)| \leq C_0 |f|_{3Q_0}.$$

This estimates the second term from the left hand side in (7.17). Regarding $|T(f\mathbf{1}_{3Q_0})(x)|\mathbf{1}_{Q_0 \setminus \bigcup_{P\in\mathcal{F}} P}$, we use (7.13) and the definition of E to get

$$\begin{aligned} |T(f\mathbf{1}_{3Q_0})(x)|\mathbf{1}_{Q_0 \setminus \bigcup_{P\in\mathcal{F}} P}(x) &\leq |T(f\mathbf{1}_{3Q_0})(x)|\mathbf{1}_{Q_0\setminus E}(x) \\ &\overset{(7.13)}{\leq} C|f(x)|\mathbf{1}_{Q_0\setminus E}(x) + \mathcal{M}_{T,Q_0} f(x)\mathbf{1}_{Q_0\setminus E}(x) \leq CC_0|f|_{3Q_0} + C_0|f|_{3Q_0}. \end{aligned}$$

Thus, we have (7.17), and the proof of (7.15) is complete. □

We are ready to finish the proof of sparse domination for Calderón-Zygmund operators.

Proof of Theorem 7.12. Suppose that $f \in L^1(\mathbb{R}^n)$ is compactly supported, and without loss of generality assume that $\operatorname{supp} f \subset R_0$ for some $R_0 \in \mathcal{D}$.

We construct a family $\mathcal{R} \subset \mathcal{D}$ which is a partition of $\mathbb{R}^n$ and such that for every $R \in \mathcal{R}$ we have $\operatorname{supp} f \subset 3R$. First, we add R_0 to $\mathcal{R}$. Then, we cover $3R_0 \setminus R_0$ by $3^n - 1$ cubes of sidelength $\ell(R_0)$, and we add them to $\mathcal{R}$. Next, we cover $9R_0 \setminus 3R_0$ by $3^n - 1$ cubes of sidelength $\ell(3R_0)$, and we add them to $\mathcal{R}$. Proceeding this way, we cover all of $\mathbb{R}^n$, and for every $R \in \mathcal{R}$ we have $\operatorname{supp} f \subset R_0 \subset 3R$.

Now, we apply Lemma 7.16 to each $R \in \mathcal{R}$. We get $\frac{1}{2}$-sparse families $\mathcal{S}(R)$ such that

$$|Tf(x)| = |T(f\mathbf{1}_{3R})(x)| \leq C \sum_{Q \in \mathcal{S}(R)} |f|_{3Q} \mathbf{1}_Q(x) \quad \text{for a.e. } x \in R.$$

Taking $\mathcal{S}' := \bigcup_{R\in\mathcal{R}} \mathcal{S}(R)$, we have

$$|Tf(x)| \leq C \sum_{Q\in\mathcal{S}'} |f|_{3Q} \mathbf{1}_Q(x) \quad \text{for a.e. } x \in \mathbb{R}^n.$$

Since $\mathcal{R}$ is a disjoint family, and cubes from $\mathcal{S}(R)$ are contained in R, we see that $\mathcal{S}'$ is $\frac{1}{2}$-sparse. Finally, we set $\mathcal{S} := \{3Q : Q \in \mathcal{S}'\}$, so that

$$|Tf(x)| \leq C \sum_{Q\in\mathcal{S}} |f|_Q \mathbf{1}_Q(x) = \mathcal{A}_{\mathcal{S}} |f|(x).$$

Note that $\mathcal{S}$ is $\frac{1}{3^n \cdot 2}$-sparse: if E_Q are the disjoint subsets associated to $Q \in \mathcal{S}'$, then

$$|E_Q| \geq \frac{1}{2}|Q| = \frac{1}{3^n \cdot 2}|3Q|.$$

This finishes the proof. $\square$

7.5 Necessity of the A_p condition

In the last few subsections we have shown that if $w \in A_p$ with $1 < p < \infty$, then all Calderón-Zygmund operators are of strong type (p,p) with respect to w (see Remark 7.2). Now we prove the converse.

Proposition 7.17. *Let $1 < p < \infty$. Suppose that w is a weight, and that one of the Riesz transforms R_j is of strong type (p,p) with respect to w. Then, $w \in A_p$.*

Hence, the A_p weights are truly the correct class of weights for which one can study singular integral operators. Note that in the proposition above we don't need to assume that all Calderón-Zygmund operators are bounded; it suffices to have one sufficiently non-degenerate operator, such as one of the Riesz transforms.

Proof. Denote by e_j the unit vector in the direction of the j-th coordinate axis. Fix a ball $B = B(x_B, r)$, and denote by B' a translate of B by $3re_j$, so that $B' = B + 3re_j$.

Let $f \in L^2(\mathbb{R}^n)$ be a non-negative function supported in B. Then, for $x \in B'$ we have

$$R_j f(x) = \int_B \frac{x_j - y_j}{|x-y|^{n+1}} f(y)\, dy.$$

Observe that if $y \in B$ and $x \in B' = B + 3re_j$, then if we write $y = x_B + ru$, $x = x_B + 3re_j + rv$, with $|u|, |v| \leq 1$, we have

$$\frac{x_j - y_j}{|x-y|^{n+1}} = \frac{3r + rv_j - ru_j}{|3re_j + rv - ru|^{n+1}} = \frac{1}{r^n} \cdot \frac{3 + v_j - u_j}{|3e_j + v - u|} \geq \frac{1}{5r^n},$$

and so

$$R_j f(x) \geq \frac{1}{5r^n} \int_B f(y) dy \sim \frac{1}{|B|} \int_B f = f_B.$$

By the strong (p,p) estimate for R_j we have

$$w(B')\,(f_B)^p \lesssim \int |R_j f|^p w \lesssim \int_B f^p w. \tag{7.19}$$

Similarly, if $x \in B$ and $y \in B'$ we have $\frac{x_j - y_j}{|x-y|^{n+1}} \leq -\frac{1}{5r^n}$. Hence, if non-negative f is supported in B', then for $x \in B$ we have $R_j f(x) \lesssim -f_{B'}$, and so

$$w(B)(f_{B'})^p \lesssim \int_{B'} f^p w.$$

Taking $f = \mathbf{1}_{B'}$ we get $w(B) \lesssim w(B')$. Plugging this into (7.19) yields

$$w(B)(f_B)^p \lesssim \int_B f^p w$$

for any non-negative $f \in L^2(\mathbb{R}^n)$ supported on B. But this inequality implies the A_p condition (as alread seen above (6.6)): if we take $f = w^{1-p'}\mathbf{1}_B$ (we may choose such f by approximating it with bounded function), then the inequality above becomes

$$w(B)\left(\frac{1}{|B|}\int_B w^{1-p'}\right)^p \lesssim \int_B w^{1-p'},$$

which is the ball variant of the A_p condition (see Exercise 6.9). □

Remark 7.18. Proposition 7.17 remains true if we replace the weight w by any Borel measure μ on $\mathbb{R}^n$. That is, if R_j is of strong type (p,p) on $(\mathbb{R}^n, \mu)$, then μ is absolutely continuous with respect to the Lebesgue measure, and its density is in A_p. See Section V.4.6 in [Ste93] for details.

8 Adjacent dyadic grids

Many operators commonly used in harmonic analysis have their dyadic counterparts. As a concrete example, consider the *dyadic Hardy Littlewood maximal operator* $M_{\mathcal{D}}$ given by

$$M_{\mathcal{D}} f(x) = \sup_{x \in Q \in \mathcal{D}} \frac{1}{|Q|} \int_Q |f|,$$

which is a dyadic counterpart of the operator M_c introduced in Subsection 6.1.

Due to the nice tree structure of the dyadic grid, the dyadic operators are often much easier to handle. However, at the end of the day one always needs to face the question: how does the dyadic operator relate to the original one?

Different variants of the Hardy-Littlewood maximal operator that we have seen before (centered, non-centered, balls, cubes...) were pointwise comparable to each other, see Exercise 6.2. It is immediate to see that for any $f \in L^1_{loc}(\mathbb{R}^n)$ we have the one-sided inequality

$$M_{\mathcal{D}} f(x) \leq M_c f(x),$$

simply because the supremum in the definition of M_c is taken over a larger family of cubes. On the other hand, the converse inequality

$$M_c f(x) \leq C M_{\mathcal{D}} f(x) \tag{8.1}$$

fails for all $C > 0$! To see this, observe that for any $f \neq 0$ we have $M_c f(x) > 0$ for all $x \in \mathbb{R}^n$, so in particular $\operatorname{supp}(M_c f) = \mathbb{R}^n$. On the other hand, for $f = \mathbf{1}_{[0,1]^n}$ we have

$$\operatorname{supp}(M_{\mathcal{D}}(\mathbf{1}_{[0,1]^n})) = \{x \in \mathbb{R}^n : x_j \geq 0 \text{ for all } j = 1, \dots, n\}$$

because all the dyadic cubes intersecting $[0,1]^n$ are contained in the set from the right hand side. This shows that the pointwise estimate (8.1) is false.[3].

What is the property of the dyadic lattice that makes it "incomparable" with the family of all cubes, or all balls? Recall that when proving the comparability $Mf(x) \sim M_c f(x)$, it was crucial that for every $x \in \mathbb{R}^n, r > 0$ we had a cube Q such that $B(x,r) \subset Q$ and $|Q| \sim r^n$, so that

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} |f| \lesssim \frac{1}{|Q|} \int_Q |f|.$$

We can no longer do the same when we restrict attention to dyadic cubes. It may happen that the smallest dyadic cube Q containing $B(x,r)$ satisfies $\ell(Q) \gg r$, or

[3]On the other hand, a weaker estimate in terms of the measure of level sets is true, see Lemma 2.12 in [Duo01]

even worse, that no such dyadic cube exists! This happens with any ball centered at 0.

To remedy this, we will consider some generalizations of the usual dyadic grid we worked with so far.

8.1 The one-third trick

The main idea is to replace the usual dyadic grid $\mathcal{D}$ by a finite number of grids $\{\mathcal{D}^e\}_{e\in\mathcal{E}}$, where each $\mathcal{D}^e$ has the same fundamental properties as $\mathcal{D}$, namely:

(D1) $\mathcal{D}^e = \bigcup_{k\in\mathbb{Z}} \mathcal{D}^e_k$, where each $\mathcal{D}^e_k$ is a partition of $\mathbb{R}^n$,

(D2) each $Q \in \mathcal{D}^e_k$ is a cube of sidelength $\ell(Q) = 2^{-k}$,

(D3) if $Q, P \in \mathcal{D}^e$ satisfy $Q \cap P \neq \varnothing$, then either $Q \subset P$ or $P \subset Q$. In particular, each $Q \in \mathcal{D}^e_k$ is contained in a unique parent cube $\widehat{Q} \in \mathcal{D}^e_{k-1}$, and contains exactly 2^n subcubes from $\mathcal{D}^e_{k+1}$.

The advantage of having multiple lattices is the following: they can be constructed in such a way that

(D4) for every $x \in \mathbb{R}^n$ and $r > 0$ there exists some $e \in \mathcal{E}$ and $Q \in \mathcal{D}^e$ such that $B(x, r) \subset Q$ and $\ell(Q) \lesssim r$.

Definition 8.1. A collection of dyadic grids satisfying (D1)–(D4) is called a collection of *adjacent dyadic grids*.

It has been known for a long time that one can construct a system of 3^n adjacent dyadic grids simply by setting for every $e \in \mathcal{E} := \{-\frac{1}{3}, 0, \frac{1}{3}\}^n$

$$\mathcal{D}^e(\mathbb{R}^n) = \{2^{-k}(m + [0,1)^n + (-1)^k e) : k \in \mathbb{Z},\ m \in \mathbb{Z}^n\}.$$

This is the famous "one-third trick", and it goes back to Garnett and Jones [GJ82] and independently Christ.

In dimension 1 an equivalent, and perhaps more illuminating, way of defining $\mathcal{D}^e(\mathbb{R})$ is the following. Define $\mathcal{D}^e_0(\mathbb{R})$ as the translation of $\mathcal{D}_0(\mathbb{R})$ by e, as above. This uniquely determines all the generations $\mathcal{D}^e_k(\mathbb{R})$ for $k \geq 0$, and we only have to define the dyadic ancestors of cubes from $\mathcal{D}^e_0(\mathbb{R})$. To do that, it suffices to define the dyadic ancestors of $I_0 := [0,1) + e$, and this will uniquely determine all the generations $\mathcal{D}^e_k(\mathbb{R})$ for $k < 0$. There are two possible dyadic parents for I_0, either $I^l_1 := [0,2) + e$ (so that I_0 is the left child of I^l_1) or $I^r_1 := [0,2) - 2e$ (so that I_0 is the right child of I^r_1). We choose I^r_1. Now we need to choose the dyadic parent to I^r_1, and this time we choose I^l_2 (so that I^r_1 is the left child of I^l_2). We proceed in this way indefinitely, always alternating between left and right possible parents .

Proposition 8.2. *The grids $\{\mathcal{D}^e\}_{e\in\mathcal{E}}$ form a collection of adjacent dyadic grids.*

Proof. The properties (D1) and (D2) are trivially true for $\mathcal{D}^e(\mathbb{R}^n)$, so it remains to check (D3) and (D4). First, we show that it suffices to check them for $n = 1$. Indeed, note that for any $Q \in \mathcal{D}^e(\mathbb{R}^n)$ we have

$$Q = I_1 \times \cdots \times I_n$$

for some dyadic intervals $I_i \in \mathcal{D}^{e_i}(\mathbb{R})$ with $|I_i| = \ell(Q)$.

To see (D3), let $Q, P \in \mathcal{D}^e(\mathbb{R}^n)$ be such that $Q \cap P \neq \varnothing$, and without loss of generality assume that $\ell(P) \geq \ell(Q)$. We have $Q = I_1 \times \cdots \times I_n$ and $P = J_1 \times \cdots \times J_n$, with $I_i, J_i \in \mathcal{D}^e(\mathbb{R})$ and $|I_i| = \ell(Q)$, $|J_i| = \ell(P)$. Since $P \cap Q \neq \varnothing$, we get $I_i \cap J_i \neq \varnothing$ for all $1 \leq i \leq n$. Recalling that $\ell(P) \geq \ell(Q)$, the property (D3) for $\mathcal{D}^{e_i}(\mathbb{R})$ gives $I_i \subset J_i$ for all i, and so $Q \subset P$. This gives (D3) for $\mathcal{D}^e(\mathbb{R}^n)$ assuming that it holds for $n = 1$.

Regarding (D4), if $x \in \mathbb{R}^n$ and $r > 0$, then $B(x, r) \subset \prod_{i=1}^n (x_i - r, x_i + r)$. Using (D4) for $n = 1$, we find $e_i \in \{-\frac{1}{3}, 0, \frac{1}{3}\}$ and $I_i \in \mathcal{D}^{e_i}(\mathbb{R})$ such that $(x_i - r, x_i + r) \subset I_i$ and $|I_i| \lesssim r$. Set $e = (e_1, \ldots, e_n)$. We may assume all the intervals I_i have equal length, and then $Q \coloneqq \prod_{i=1}^n I_i \in \mathcal{D}^e(\mathbb{R}^n)$, $B(x, r) \subset Q$ and $\ell(Q) \lesssim r$. This gives (D4) assuming that it is true in dimension 1.

Suppose now $n = 1$. It is easy to see that for any $x \in \mathbb{R}$ and $\frac{1}{12} \leq r < \frac{1}{6}$ the interval $(x - r, x + r)$ is contained in one of the unit intervals from $\bigcup_{e\in\mathcal{E}} \mathcal{D}^e_0(\mathbb{R})$. Then, by rescaling, it follows that for any $0 < r < \infty$ there exists $I \in \bigcup_{e\in\mathcal{E}} \mathcal{D}^e(\mathbb{R})$ such that $(x - r, x + r) \subset I$ and $6r \leq |I| \leq 12r$. So the property (D4) holds.

It remains to show that each $\mathcal{D}^e(\mathbb{R})$ satisfies the nestedness property (D3). Observe that it is equivalent to the following statement: for any $k \in \mathbb{Z}$ and $l \in \mathbb{Z}$ with $l < k$ the set of endpoints of intervals from $\mathcal{D}^e_k(\mathbb{R})$

$$V^e_k \coloneqq \{2^{-k}(m + (-1)^k e) : m \in \mathbb{Z}\}$$

satisfies $V^e_l \subset V^e_k$. It suffices to show this for $l = k - 1$, as the general case follows by induction.

For $e = 0$ this is trivial, so assume $e \in \{-\frac{1}{3}, \frac{1}{3}\}$. The crucial observation is that

$$2e = 3e - e \quad \text{and} \quad 3e \in \{-1, 1\}.$$

Thus,

$$\begin{aligned}
V^e_{k-1} &= \{2^{-k+1}(m + (-1)^{k-1} e) : m \in \mathbb{Z}\} = \{2^{-k}(2m + (-1)^{k-1} 2e) : m \in \mathbb{Z}\} \\
&= \{2^{-k}(2m + (-1)^{k-1} 3e - (-1)^{k-1} e) : m \in \mathbb{Z}\} \\
&= \{2^{-k}(2m + (-1)^{k-1} 3e + (-1)^k e) : m \in \mathbb{Z}\} \\
&\subset \{2^{-k}(m + (-1)^k e) : m \in \mathbb{Z}\} = V^e_k.
\end{aligned}$$

This finishes the proof of (D3). □

8.2 Conde-Alonso's grids

For most applications the 3^n adjacent dyadic grids from the previous subsection are perfectly enough. Nevertheless, in some cases it may be useful to have a system of adjacent grids consisting od fewer lattices (for example, if we care about the dimensional dependence of our estimates).

Recently, Conde-Alonso [Con13] proved that one only needs $(n+1)$ carefully chosen dyadic lattices $\{\mathcal{D}^j\}_{j=0}^n$ to form a family of adjacent dyadic grids in $\mathbb{R}^n$, and that this is the optimal number of lattices. The construction goes as follows.

Observe that a dyadic grid is uniquely determined by choosing a unit cube (which we will call "initial cube") and then choosing its dyadic ancestors[4]. Let $\mathcal{D}^0 := \mathcal{D}$ be the usual dyadic grid on $\mathbb{R}^n$. For any other $j \in \{1, \dots, n\}$ we define dyadic grids $\mathcal{D}^j$ using the following algorithm. Let $p_n > n$ be the smallest odd integer strictly larger than n, and let $\mathbf{v} := (1, 1, \dots, 1) \in \mathbb{R}^n$.

(i) We choose the initial cube of $\mathcal{D}^j$ to be

$$\mathbf{Q} := [0,1)^n + \frac{j}{p_n}\mathbf{v}.$$

This uniquely determines $\mathcal{D}^j_k$ for all generations $k \geq 0$.

(ii) There are 2^n possible dyadic parents of $\mathbf{Q}$. We choose $\mathbf{Q}^1$ to be the unique possible parent of $\mathbf{Q}$ satisfying

  (a) $\mathbf{Q}^1 = [0,2)^n + \lambda\mathbf{v}$

  (b) $\lambda \in \frac{2}{p_n}\mathbb{Z}$.

There are only two possible parents of $\mathbf{Q}$ satisfying (a), namely $[0,2)^n + \frac{j}{p_n}\mathbf{v}$ and $[0,2)^n + \frac{j-p_n}{p_n}\mathbf{v}$. Only one of them satisfies condition (b), and here we crucially use that p_n is odd. The choice of $\mathbf{Q}^1$ determines $\mathcal{D}^j_{-1}$.

(iii) By induction, the parent of $\mathbf{Q}^{k-1}$ is chosen as the unique possible dyadic parent satisfying

$$\mathbf{Q}^k = [0,2^k) + \lambda\mathbf{v} \quad \text{for some} \quad \lambda \in \frac{2^k}{p_n}\mathbb{Z}.$$

This choice determines $\mathcal{D}^j_{-k}$.

Proposition 8.3. *The grids $\{\mathcal{D}^j\}_{j=0}^n$ form a system of adjacent dyadic lattices.*

[4] We already saw this in the discussion above Proposition 8.2

Note that in the proof of Proposition 8.2 it was very easy to show that (D4) holds, and most work was dedicated into proving that $\mathcal{D}^e$ for $e \in \mathcal{E}$ are dyadic lattices. Now we are in the opposite situation, where the properties (D1)–(D3) are immediate for $\mathcal{D}^j$, but we need to work a bit to prove (D4). We begin by proving two auxiliary lemmas.

Lemma 8.4. *Fix $k \in \mathbb{Z}$. Let V_k^j denote the vertices of the cubes in $\mathcal{D}_k^j$. For all $j \in \{0, \ldots, n\}$*

$$V_k^j \subset V_k := \frac{2^{-k}}{p_n}\mathbb{Z}^n.$$

Proof. We prove first $V_k^j \subset V_k$. This is clear for $k = 0$. For general k, note first that if a single vertex v of a single cube $Q \in \mathcal{D}_k^j$ is in V_k, then the same is true for all the other vertices of all the other cubes in $\mathcal{D}_k^j$, simply because the other vertices are of the form

$$2^{-k}m + v = \frac{2^{-k}p_n}{p_n}m + v \in V_k$$

for some $m \in \mathbb{Z}^n$. Now, if $k \geq 1$, then $V_0^j \subset V_k^j$. Since $V_0^j \subset V_0 \subset V_k$, we get that some vertices in V_k^j are contained in V_k. Hence, $V_k^j \subset V_k$. On the other hand, if $k \leq -1$, then the cube $\mathbf{Q}^{-k}$ from the construction of $\mathcal{D}_k^j$ has a vertex at $\lambda\mathbf{v}$. Since $\lambda \in \frac{2^{-k}}{p_n}\mathbb{Z}$, we have $\lambda\mathbf{v} \in V_k$, and so $V_k^j \subset V_k$. $\square$

Lemma 8.5. *Denote by $\pi_i : \mathbb{R}^n \to \mathbb{R}$ the orthogonal projection to the x_i-axis. For any $i \in \{1, \ldots, n\}$ we have $\pi_i(V_k^j) \cap \pi_i(V_k^{j'}) = \varnothing$ whenever $j \neq j'$.*

Proof. First, observe that $V_k^j \cap V_k^{j'} = \varnothing$. This is simply because a single vertex determines the entire 2^{-k}-grid, and so $V_k^j \cap V_k^{j'} \neq \varnothing$ implies $V_k^j = V_k^{j'}$, which is only true if $j = j'$.

Now note that $\pi_i(V_k^j) = \pi_{i'}(V_k^j)$ for any $i, i' \in \{1, \ldots, n\}$. This is because for every i the set $\pi_i(V_k^j)$ is a 2^{-k} grid on $\mathbb{R}$, and they all contain the point $\frac{j}{p_n}$ (here we use the fact that $\mathbf{v} = (1, 1, \ldots, 1)$).

It follows that if we have $\pi_i(V_k^j) \cap \pi_i(V_k^{j'}) \neq \varnothing$ for some i, then $\pi_i(V_k^j) = \pi_i(V_k^{j'})$ for all i, and the grid structure of V_k^j then implies $V_k^j = V_k^{j'}$. $\square$

Proof of Proposition 8.3. To prove (D4), it suffices to show that for any axes-parallel cube $R \subset \mathbb{R}^n$ there exists some $j \in \{0, \ldots, n\}$ and $Q \in \mathcal{D}^j$ such that $R \subset Q$ and $\ell(Q) \lesssim \ell(R)$. Fix the unique $k \in \mathbb{Z}$ such that

$$\frac{2^{-k-1}}{p_n} \leq \ell(R) < \frac{2^{-k}}{p_n}.$$

Note that the upper bound implies

$$\#(\pi_i(R) \cap \pi_i(V_k)) \leq 1 \tag{8.2}$$

for all $i \in \{1, \ldots, n\}$, where $\#$ denotes cardinality.

We claim that for some $j \in \{0, \ldots, n\}$ and $Q \in \mathcal{D}_k^j$ we have $R \subset Q$. Suppose this is not the case. Then, for every j there exists $i_j \in \{1, \ldots, n\}$ such that $\pi_{i_j}(R) \cap \pi_{i_j}(V_k^j) \neq \varnothing$. Since there are $n+1$ families $\mathcal{D}_k^j$ and only n projections π_i, by the pigeonhole principle there exist $0 \leq j < j' \leq n$ such that $i_j = i_{j'} =: i$. Hence,

$$\pi_i(R) \cap \pi_i(V_k^j) \neq \varnothing \quad \text{and} \quad \pi_i(R) \cap \pi_i(V_k^{j'}) \neq \varnothing.$$

By Lemma 8.5 we have $\pi_i(V_k^j) \cap \pi_i(V_k^{j'}) = \varnothing$, and by Lemma 8.4 $\pi_i(V_k^j) \cup \pi_i(V_k^{j'}) \subset \pi_i(V_k)$. Hence, $\#(\pi_i(R) \cap \pi_i(V_k)) \geq 2$, but this is a contradiction with (8.2). □

The following exercise demonstrates that $n+1$ is the smallest possible number of adjacent dyadic grids.

Exercise 8.6. Suppose that $\mathcal{A}^1, \ldots, \mathcal{A}^n$ is a family of n dyadic lattices on $\mathbb{R}^n$. Show that there exists a point $q \in \mathbb{R}^n$ such that for any $\varepsilon > 0$ and $j \in \{1, \ldots, n\}$ if $B(x, \varepsilon) \subset Q \in \mathcal{A}^j$, then $\ell(Q) \geq 1$.

Hint: Find cubes $Q_j \in \mathcal{A}^j$, $j \in \{1, \ldots, n\}$ with $\ell(Q_j) = 1$ and such that $\bigcap_{j=1}^n \partial Q_j \neq \varnothing$. Then pick $q \in \bigcap_{j=1}^n \partial Q_j$.

8.3 An application

We give one simple application. Recall the dyadic variant of the Hardy-Littlewood maximal operator $M_{\mathcal{D}}$. As explained below (8.1), the pointwise estimate $M_c f(x) \leq C M_{\mathcal{D}} f(x)$ is false for all $C > 0$. However, if we replace $M_{\mathcal{D}} f$ by a sum of maximal operators associated to a collection of adjacent dyadic grids, things look better.

Lemma 8.7. *Suppose that $\{\mathcal{D}^e\}_{e \in \mathcal{E}}$ is a collection of adjacent dyadic grids. Then,*

$$M_c f(x) \lesssim \sum_{e \in \mathcal{E}} M_{\mathcal{D}^e} f(x). \tag{8.3}$$

Proof. Let Q be a cube containing x. By the definition of adjacent dyadic grids, there exists $e \in \mathcal{E}$ such that $Q \subset P \in \mathcal{D}^e$ and $\ell(P) \lesssim \ell(Q)$. Hence,

$$\frac{1}{|Q|} \int_Q |f| \lesssim \frac{1}{|P|} \int_P |f| \leq M_{\mathcal{D}^e} f(x).$$

Taking supremum over all cubes Q containing x finishes the proof. □

The estimate (8.3) can be used to give an alternative proof of the weak $(1, 1)$ estimate for M_c, avoiding the use of the $5r$ covering lemma. Indeed, establishing the weak $(1, 1)$ estimate for dyadic Hardy-Littlewood maximal operator is almost immediate.

Lemma 8.8. *The dyadic Hardy-Littlewood maximal operator $M_{\mathcal{D}}$ is weak* $(1, 1)$.

Proof. Let $f \in L^1(\mathbb{R}^n)$, and fix $\lambda > 0$. Let $\mathcal{B} \subset \mathcal{D}$ be the family of maximal dyadic cubes satisfying

$$\frac{1}{|Q|} \int_Q |f| > \lambda,$$

so that $\mathcal{B}$ is the family of cubes from the Calderón-Zygmund decomposition of f at level λ. Then,

$$\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\} \subset \bigcup_{Q \in \mathcal{B}} Q$$

and so

$$|\{x \in \mathbb{R}^n : M_{\mathcal{D}} f(x) > \lambda\}| \leq \sum_{Q \in \mathcal{B}} |Q| \leq \sum_{Q \in \mathcal{B}} \frac{\int_Q |f|}{\lambda} \leq \frac{\|f\|_{L^1}}{\lambda}.$$

□

Of course, in the lemma above it doesn't matter whether we take the usual dyadic grid $\mathcal{D}$ or some other grid $\mathcal{D}^e$. Together with (8.3), we immediately get the weak $(1, 1)$ estimate for M_c.

References

[AIS01] K. Astala, T. Iwaniec, and E. Saksman. Beltrami operators in the plane. *Duke Math. J.*, 107(1):27–56, 2001. doi:10.1215/S0012-7094-01-10713-8.

[Ast94] K. Astala. Area distortion of quasiconformal mappings. *Acta Math.*, 173(1):37–60, 1994. doi:10.1007/BF02392568.

[CMM82] R. R. Coifman, A. McIntosh, and Y. Meyer. L'integrale de Cauchy Definit un Operateur Borne sur L^2 Pour Les Courbes Lipschitziennes. *Ann. of Math.*, 116(2):361–387, 1982. doi:10.2307/2007065.

[Con13] J. M. Conde. A note on dyadic coverings and nondoubling Calderón–Zygmund theory. *J. Math. Anal. Appl.*, 397(2):785–790, 2013. doi:10.1016/j.jmaa.2012.08.015.

[CU17] D. Cruz-Uribe. Extrapolation and Factorization. *Lecture notes*, 2017. doi:10.48550/arXiv.1706.02620.

[CZ52] A. P. Calderon and A. Zygmund. On the existence of certain singular integrals. *Acta Math.*, 88(1):85–139, 1952. doi:10.1007/BF02392130.

[DGPP05] O. Dragičević, L. Grafakos, M. Pereyra, and S. Petermichl. Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces. *Publ. Mat.*, 49(1):73–91, 2005.

[Duo01] J. Duoandikoetxea. *Fourier analysis*, volume 29 of *Grad. Stud. Math.* Amer. Math. Soc., 2001.

[EG91] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL, USA, 1991.

[FKP91] R. A. Fefferman, C. E. Kenig, and J. Pipher. The theory of weights and the Dirichlet problem for elliptic equations. *Ann. Of Math.*, 134(1):65–124, 1991. doi:10.2307/2944333.

[GJ82] J. B. Garnett and P. W. Jones. BMO from dyadic BMO. *Pacific J. Math.*, 99(2):351–371, 1982.

[Gra14a] L. Grafakos. *Classical Fourier Analysis*, volume 249 of *Grad. Texts in Math.* Springer, New York, 3rd edition, 2014. doi:10.1007/978-1-4939-1194-3.

[Gra14b] L. Grafakos. *Modern Fourier analysis*, volume 250 of *Grad. Texts in Math.* Springer, New York, 3rd edition, 2014. doi:10.1007/978-1-4939-1230-8.

[Hän18] T. S. Hänninen. Equivalence of sparse and Carleson coefficients for general sets. *Ark. Mat.*, 56(2):333–339, 2018. doi:10.4310/ARKIV.2018.v56.n2.a8.

[Hyt12] T. P. Hytönen. The sharp weighted bound for general Calderón–Zygmund operators. *Ann. Of Math.*, 175(3):1473–1506, 2012. doi:10.4007/annals.2012.175.3.9.

[Ler16] A. K. Lerner. On pointwise estimates involving sparse operators. *New York J. Math*, 22:341–349, 2016.

[LN19] A. K. Lerner and F. Nazarov. Intuitive dyadic calculus: The basics. *Exposition. Math.*, 37(3):225–265, Sept. 2019. doi:10.1016/j.exmath.2018.01.001.

[Mat95] P. Mattila. *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, volume 44 of *Cambridge Stud. Adv. Math.* Cambridge Univ. Press, Cambridge, UK, 1995. doi:10.1017/CBO9780511623813.

[Par20] I. Parissis. Harmonic analysis. *Lecture notes*, 2020. URL `https://drive.google.com/file/d/0B7t_mQHDlsRsSWFFU0p0bEhPWFU/view?resourcekey=0-NLyWujr_-zJC4M5QrbbSGA`.

[PV02] S. Petermichl and A. Volberg. Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular. *Duke Math. J.*, 112(2):281–305, 2002. doi:10.1215/S0012-9074-02-11223-X.

[Ste70] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions*, volume 30 of *Princeton Math. Ser.* Princeton Univ. Press, Princeton, NJ, 1970.

[Ste93] E. M. Stein. *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, volume 43 of *Princeton Math. Ser.* Princeton Univ. Press, Princeton, NJ, 1993.

[Tol14] X. Tolsa. *Analytic capacity, the Cauchy transform, and non-homogeneous Calderón-Zygmund theory*, volume 307 of *Progr. Math.* Birkhäuser, Cham, 2014. doi:10.1007/978-3-319-00596-6.

[Ver21] J. Verdera. Birth and life of the L^2 boundedness of the Cauchy Integral on Lipschitz graphs. *arXiv*, 2021. doi:10.48550/arXiv.2109.06690.