ISSN 1851-149X

Fascículo 4

Cursos y seminarios de matemática

Serie B

Michael T. Lacey

Some Topics in Dyadic Harmonic Analysis

Departamento de Matemática

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

2009

Cursos y Seminarios de Matemática – Serie B

Fascículo 4

ISSN 1851-149X

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires
Ciudad Universitaria- Pabellón I
(1428) Ciudad de Buenos Aires
Argentina.
http://www.dm.uba.ar
e-mail. secre@dm.uba.ar
tel/fax: (+54-11)-4576-3335

Some Topics in Dyadic Harmonic Analysis

Michael T. Lacey

September 20, 2008

Preface

This manuscript is a course in Harmonic Anlysis, with an emphasis on dyadic Harmonic Analysis. Dyadic Harmonic Analysis can be traced back to the early years of the 20^{th} century, and A. Haar's basis of orthonormal functions, which have profound, and still useful connections to combinatorial and probabilistic reasoning. These themes have received a renewed attention in recent years, spurred on the one hand by the recognition, first by Stefanie Petermichl, that a notion of Haar shifts can be used to recover deep results about the Hilbert transform. On the other hand, the analysis of the Bilinear Hilbert Transform lead to a reëxamination of discrete, combinatorial decompositions of complicated operators.

We cover topics that on the one hand are familiar: Maximal Functions, Singular Integrals, Paraproducts, the T 1 Theorem of David and Journé and weighted inequalities. But, some proofs are recent. We prove the L^2 boundedness of the Hilbert transform H by using Haar shifts, an argument of Stefanie Petermichl from 2000. We give analogous proof of the boundedness of the commutator $[b, \mathrm{H}]$, where b is a funtion. We present the proof of a dyadic T 1 theorem, with an exceptionally transparent proof. And in the chapter on weighted inequalities, we have prove some basic facts with the shortest proofs available, in particular we follow Andrei Lerner's 2008 argument for the Muckenhoupt-Wheeden Theorem on one-weight inequalities for the Maximal function. As well, we prove the two-weight Maximal function Theorem of Eric Sawyer from 1082, and a recent beautiful two-weight result for Haar shifts due to Nazarov-Treil-Volberg, 2007.

The most striking absence is a proof of Carleson's Theorem on the pointwise convergence of Fourier series, which fits the themese of these notes. A future version should fix this omission. We also stress that the references in this manuscript are extremely limited, to just a handful of results. This is simply a matter of expediency, and a future version of this manuscript will include far more extensive references. We apologize in advance for our many omissions.

These notes reflect much of the material taught in a course in Harmonic Analysis, during the authors' stay at the Universidad of Buenos Aires in 2008, a stay funded in part by the Fulbright Foundation of Argentina. It is a pleasure to include this manuscript in the book series associated with the Mathematics Department at the UBA. I want to thank the students who attended the course, MagalíAnastasio, Alfredo González, Gustavo Massaccesi, Juan Medina, Maria del Carmen Moure, Carolina Mosquera, Victoria Paternostro, Alejandro Quintero, Ezequiel Rela, José Luis Romero, and Leandro Zuberman.

A very special thanks goes to Carlos Cabrelli and Ursula Molter, who encouraged the application to the Fulbright Foundation, and were warm wonderful hosts during the stay of myself and family in Argentina.

Michael T. Lacey, September 20, 2008

Contents

Chapter 1

Maximal Functions and Singular Integrals

1.1 Dilations and Translations

A dominant theme of this chapter is that the operators we consider exhibit invariance properties with respect to two groups of operators. The first group is the *translation* operators

$$\operatorname{Tr}_y f(x) := f(x-y)\,, \qquad y \in \mathbb{R}\,. \tag{1.1.1}$$

Formally, the adjoint operator is $(\operatorname{Tr}_y)^* = \operatorname{Tr}_{-y}$. The collection of operators $\{\operatorname{Tr}_y : y \in \mathbb{R}\}$ is a representation of the additive group $(\mathbb{R}, +)$.

It is an important and very general principle that a linear operator L acting on some vector space of functions, which is assumed to commute with all translation operators, is in fact given as convolution. In general, it is convolution with respect to a measure or distribution μ, thus,

$$\operatorname{L} f(x) = \int f(x-y)\; \mu(dy)\,.$$

For instance, with the identity operator, μ would be the Dirac pointmass at the origin.

The second group is the set of *dilations on L^p*, given by

$$\operatorname{Dil}^{(p)}_\lambda f(x) := \lambda^{-1/p} f(x/\lambda)\,, \qquad 0 < \lambda, p < \infty\,. \tag{1.1.2}$$

We make the definition so that $\|f\|_p = \|\operatorname{Dil}^{(p)}_\lambda f\|_p$, and deliberately permit $0 < p < 1$ in this definition. The *scale* of the dilation $\operatorname{Dil}^{(p)}_\lambda$ is said to be λ. These operators are a representation of the multiplicative group $(\mathbb{R}_+, *)$. It is a fact that we shall have reference to that the Haar measure of this group is dy/y.

It is a useful Theorem, one that we do not prove in these notes, that the set of operators L that are bounded from $L^2(\mathbb{R})$ to itself, and commute with both translations and dilations have a special form. They are linear combinations of the Identity operator, and the *Hilbert transform.*

The latter operator, fundamental to this study, is given by

$$\mathrm{H}\, f(x) := \text{p.v.} \int f(x-y)\, \frac{dy}{y}\,. \tag{1.1.3}$$

Here, we take the integral in the *principal value* sense, as the kernel $1/y$ is not integrable. Taking advantage of the fact that the kernel is odd, one can see that the limit below

$$\lim_{\epsilon \to 0} \int_{\epsilon < |y| < 1/\epsilon} f(x-y)\, \frac{dy}{y} \tag{1.1.4}$$

exists for all x, provided f is a Schwartz function, say. Thus, H has an unambiguous definition on a class of functions that is dense in all L^p. We shall take (1.1.4) as our general definition of principal value. The Hilbert transform is the canonical example of a *singular integral*, that is one that has to be defined in some principal value sense.

The operator H, being convolution, commutes with all translations. That it also commutes with all dilation operators follows from the observation that $1/y\mathbf{1}_{\mathbb{R}_+}$ is a multiplicative Haar measure, the Haar measure being unique up to a multiplicative constant.

The Hilbert transform has a profound connection to the theory of analytic functions. For f on $\mathbb{R}$, the function $f + i\,\mathrm{H}\, f$ admits analytic extension to the upper half plane $\mathbb{C}_+ = \{z \,:\, \mathrm{Im}(z) > 0\}$. (Thus, $\mathrm{H}\, f$ is sometimes refereed to as the *conjugate* to f.) Analytic functions are governed by the elliptic p.d.e. $\partial_{\bar z} F = 0$, an indication of the much deeper connection between singular integrals and a wide variety of p.d.e.s with an elliptic component.

1.2 Dyadic Grid and Haar Functions

Underlying this subject are the delicate interplay between local averages and differences. Some of this interplay can be encoded into the combinatorics of *grids*, especially the *dyadic grid.* Let us define

1.2.1 Definition. A collection of intervals $\mathcal{I}$ is a *grid* iff for all $I, J \in \mathcal{I}$, we have $I \cap J \in \{\emptyset, I, J\}$.

The primary example is that of the dyadic grid given by $\mathcal{D} := \{2^k(j, j+1) \,:\, j,k \in \mathbb{Z}\}$. There is a second possible choice of dyadic grids given by

$$\mathcal{D}' := \{[j2^k, (j+1)2^k) + (-1)^k \tfrac{1}{3} 2^k \,:\, j,k \in \mathbb{Z}\} \tag{1.2.2}$$

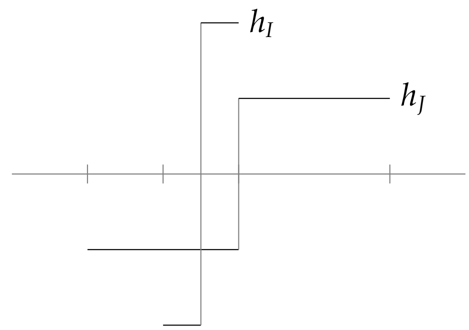


Figure 1.1: Two Haar functions.

It is a simple matter to check that this is a grid, using the identity $\frac{1}{2}(1-\frac{1}{3}) = \frac{1}{3}$. Despite the simplicity of this definition, there is a range of refinements of the notion that turn out to be helpful in different circumstances.

The Haar functions are a remarkable class of functions indexed by the dyadic grid $\mathcal{D}$. Set

$$h(x) = -\mathbf{1}_{(-1/2,0)} + \mathbf{1}_{(0,1/2)}\,,$$

a mean zero function supported on the interval $(-1/2, 1/2)$, taking two values, with L^2 norm equal to one. Define the *Haar function* (associated to interval I) to be

$$h_I := \operatorname{Dil}_I^2 h_I \tag{1.2.3}$$

$$\operatorname{Dil}_I^{(p)} := \operatorname{Tr}_{c(I)} \operatorname{Dil}_{|I|}^{(p)}\,, \qquad c(I) = \text{center of } I. \tag{1.2.4}$$

Here, we introduce the notion for the *Dilation associated with interval I.*

The Haar functions have profound properties, due to their connection to both analytical and probabilistic properties. At this point, we prove just their most elementary property, namely that they form a basis on $L^2(\mathbb{R})$.

1.2.5 Theorem. *The set of functions* $\{\mathbf{1}_{[0,1]}\} \cup \{h_I \,:\, I \in \mathcal{D},\, I \subset [0,1]\}$ *form an orthonormal basis for* $L^2([0,1])$. *The set of functions* $\{h_I \,:\, I \in \mathcal{D}\}$ *from an orthonormal basis for* $L^2(\mathbb{R})$.

Proof. The orthonormality follows immediately, and so the main point to prove is that the Haar functions form a basis . Consider the the first assertion, about the basis for $L^2([0,1))$, and the orthogonal projections

$$\mathrm{P}_n f = \langle f, \mathbf{1}_{[0,1]}\rangle + \sum_{\substack{I\in\mathcal{D}\\ |I|\ge 2^{-n}}} \langle f, h_I\rangle h_I\,.$$

Here, $n \ge 0$ is an integer. These projections are in fact a conditional expectation operator

$$\mathrm{P}_n f = \mathbb{E}(f \,:\, \mathcal{F}_n) = \sum_{\substack{I\in\mathcal{D}\\ |I|= 2^{-n-1}}} \frac{\mathbf{1}_I}{|I|} \int_I f(y)\, dy\,.$$

Here, $\mathcal{F}_n$ is the sigma-field generated by the dyadic subintervals of length 2^{-n-1}. Indeed, let V_n be the range of P_n and W_n the range of the conditional expectation operator. Then, each Haar function h_J with $|J| \geq 2^{-n}$ is in W_n so $V_n \subset W_n$. The dimension of W_n is 2^{n+1}, while by orthogonality of the Haar functions, the dimension of V_n is

$$1 + 1 + \cdots + 2^n = 2^{n+1}\,.$$

So, $V_n = W_n$. The union of the V_n is dense in $L^2([0,1])$, so the Haar functions form a basis.

Let us mention that the sequence of functions $\{\mathrm{P}_n f \;:\; n \in \mathbb{N}\}$ form a *martingale*, indicating the strong relationship between Haar functions and probabilistic reasoning. See (2.5.3) below.

Turning to the case of $L^2(\mathbb{R})$, let us take a dense class of functions, and argue that they are in the closure of the Haar functions. Our dense class will be bounded functions f supported on a closed interval $[-A, A]$. For an integer j, we have

$$L^2(\mathbb{R}) = \bigoplus_{\substack{J\in\mathcal{D}\\ |J|=2^j}} L^2(J)\,.$$

Moreover, we have a basis for each of the L^2 spaces on the right above. Namely, for dyadic interval J, an orthonormal basis for $L^2(J)$ is

$$\mathcal{H}(J) = \{\mathrm{Dil}_J^{(2)}\,\mathbf{1}_{[-1/2,1/2]}\} \cup \{h_I \,:\, I \in \mathcal{D},\ I \subset J\}\,.$$

Therefore, for each integer j, f is in the L^2-closure of the orthonormal basis for $L^2(\mathbb{R})$ given by $\bigcup_{k=-\infty}^{\infty} \mathcal{H}([k2^j, (k+1)2^j))$. But, the projection onto the functions in this basis which are *not* Haar functions has a small L^2 norm-squared

$$\begin{aligned}\Big\|\sum_{\substack{J\in\mathcal{D}\\ |J|=2^j}} \langle f, \mathrm{Dil}_J^{(2)}\,\mathbf{1}_{[-1/2,1/2]}\rangle\, \mathrm{Dil}_J^{(2)}\,\mathbf{1}_{[-1/2,1/2]}\Big\|_2^2 &= \sum_{\substack{J\in\mathcal{D}\\ |J|=2^j}} \langle f, \mathrm{Dil}_J^{(2)}\,\mathbf{1}_{[-1/2,1/2]}\rangle^2 \\ &\leq \langle f, \mathrm{Dil}_{(-2^j,0)}^{(2)}\,\mathbf{1}_{[-1/2,1/2]}\rangle^2 + \langle f, \mathrm{Dil}_{(0,2^j)}^{(2)}\,\mathbf{1}_{[-1/2,1/2]}\rangle^2 \\ &\leq 2A\|f\|_\infty^2 2^{-j/2}\,.\end{aligned}$$

This holds for all j such that $2^j \geq A$. Taking $j \to \infty$ then proves our Theorem.

□

1.2.6 Exercise. Show that for any interval I, there is an interval $J \in \mathcal{D} \cup \mathcal{D}'$ with $I \subset J$ and $|J| \leq 8|I|$. Here, $\mathcal{D}$ is the dyadic grid, and $\mathcal{D}'$ is the grid in (1.2.2).

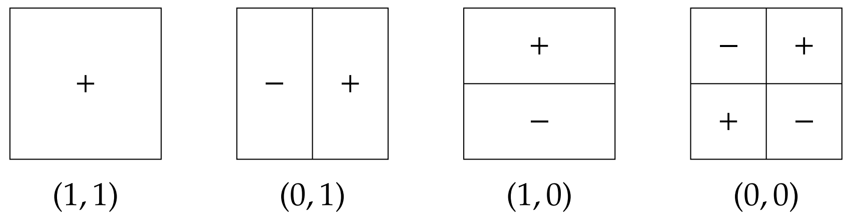


Figure 1.2: The four Haar functions in the plane.

1.3 Haar Functions in Higher Dimensions

We describe the Haar functions in higher dimensions. For integer d, let $\mathcal{D}^d$ denote the the dyadic cubes in dimension d. For $\sigma = (\sigma_1, \ldots, \sigma_d) \in \{0,1\}^d$, and $Q = \prod_{j=1}^d Q_j \in \mathcal{D}^d$, we define

(1.3.1) $$h_Q^\sigma(x_1, \ldots, x_d) = \prod_{j=1}^d h_{Q_j}^{\sigma_j}(x_j)\,.$$

Here, we refer to σ as the *signature* of the Haar function. The four possible functions in the plane are indicated in Figure 1.2.

1.3.2 Theorem. *In dimension d, considering first the unit cube, the functions below are an orthonormal basis for $L^2([0,1]^d)$.*

$$\{h^{\mathbf{1}}_{[0,1]^d}\} \cup \left\{h_Q^\sigma \,:\, Q \in \mathcal{D}^d\,,\, Q \subset [0,1]^d\,,\, \sigma \in \{0,1\}^d - \{\mathbf{1}\}\right\}.$$

And the functions below form a basis for $L^2(\mathbb{R}^d)$.

$$\{h_Q^\sigma \,:\, Q \in \mathcal{D}\,,\, \sigma \in \{0,1\}^d - \{\mathbf{1}\}\}\,.$$

Proof. Let us consider the first claim. For an integer $n \geq 0$ set V_n be the linear span of the Haar functions with volume at least 2^{-dn}. That is, V_n is the span of

$$\{h^{\mathbf{1}}_{[0,1]^d}\} \cup \left\{h_Q^\sigma \,:\, Q \in \mathcal{D}^d\,,\, Q \subset [0,1]^d\,,\, |Q| \geq 2^{-n}\,,\, \sigma \in \{0,1\}^d - \{\mathbf{1}\}\right\}.$$

And let W_n be the linear span of the dyadic cubes of volume at least $2^{-d(n+1)}$. It is clear that $V_n \subset W_n$. Also, the dimension of W_n is $2^{d(n+1)}$, as that is the number of dyadic cubes of volume equal to $2^{-d(n+1)}$. But, all the Haar functions are orthogonal, so that the dimension of V_n is

$$1 + (2^d - 1)\sum_{j=0}^n 2^{dj} = 1 + (2^d - 1)\frac{2^{d(n+1)} - 1}{2^d - 1} = 2^{d(n+1)}\,.$$

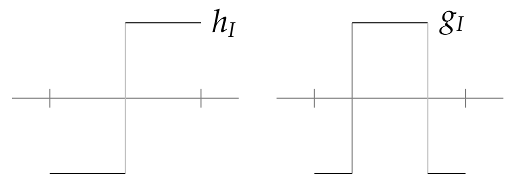


Figure 1.3: A Haar function h_I and its dual g_I.

Hence $V_n = W_n$. Thus, the Haar functions have dense span in $L^2([0,1]^d)$, so the proof of the first claim is finished.

The proof of the second claim can be completed in a manner similar to the proof of Theorem 1.2.5, and we omit the details.

□

1.4 The Hilbert Transform, and a Discrete Analog

The Hilbert transform, defined in (1.1.3) can be recovered in a remarkable way from Haar functions. Let us define

$$g = -\mathbf{1}_{(-1/4,-1/4)} + \mathbf{1}_{(-1/4,1/4)} - \mathbf{1}_{(1/4,1/2)} \tag{1.4.1}$$

$$= 2^{-1/2}\{h_{(-1/2,0)} + h_{(0,1/2)}\} \tag{1.4.2}$$

$$\mathfrak{H} f = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle g_I\,, \tag{1.4.3}$$

where as before, $g_I = \operatorname{Dil}^{(2)}_I g$. See Figure 1.3. It is clear that $\mathfrak{H}$ is a bounded operator on L^2. What is surprising is that that it can be used to recover the Hilbert transform exactly.

1.4.4 Proposition. *There is a non-zero constant c so that*

$$H = c \lim_{Y \to \infty} \int_0^Y \int_1^2 \operatorname{Tr}_y \operatorname{Dil}^{(2)}_\lambda \mathfrak{H} \operatorname{Dil}^{(2)}_{1/\lambda} \operatorname{Tr}_{-y} \frac{d\lambda}{\lambda} \frac{dy}{Y}\,. \tag{1.4.5}$$

This Proposition was discovered by Stefanie Petermichl.[1] A nice heuristic for the Proposition is that $\mathrm{H}\sin(x) = \cos x$, where h_I represents a local sine, and g_I a local cosine. As a Corollary, we have the estimate $\|H\|_2 \lesssim 1$, as $\mathfrak{H}$ is clearly bounded on L^2.

1

Petermichl, Stefanie. 2000. *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*, C. R. Acad. Sci. Paris Sér. I Math. **330**, 455–460. MR1756958 (2000m:42016) (English, with English and French summaries)

Proof. Consider the limit on the right in (1.4.5). Call the limit $\widetilde{\mathrm{H}}f$. This is seen to exist for each $x \in \mathbb{R}$ for Schwartz functions f. Define the auxiliary operators below, where we sum over small scales.

$$\mathrm{T}_j f := \sum_{\substack{I\in\mathcal{D}\\ |I|\le 2^j}} \langle f, h_j\rangle g_j .$$

The individual terms of this series are rapidly convergent. As $|I|$ becomes small, one uses the smoothness of the function f. As $|I|$ becomes large, one uses the fact that f is integrable, and decays rapidly.

Let us also note that the operator T_j is invariant under translations by an integer multiple of 2^j. Thus, the auxiliary operator

$$2^{-j}\int_0^{2^j} \mathrm{Tr}_{-t}\, f\, \mathrm{Tr}_t \; dt$$

will be translation invariant. Thus $\widetilde{\mathrm{H}}$ is convolution with respect to a linear functional on Schwartz functions, namely a distribution.

Concerning dilations, T is invariant under dilations by a power of 2. Now, dilations form a group under multiplication on $\mathbb{R}_+$, and this group has Haar measure $d\delta/\delta$ so that the operator below will commute with all dilations.

$$\int_0^1 \mathrm{Dil}^2_{1/\delta}\, \mathrm{T}\, \mathrm{Dil}^2_{\delta}\, \frac{d\delta}{\delta}$$

Thus, $\widetilde{\mathrm{H}}$ commutes with all dilations.

Let us set G_j to be the operator

$$\mathrm{G}_j f := \int_0^{2^j} \mathrm{Tran}_t \sum_{\substack{I\in\mathcal{D}\\ |I|=2^j}} \langle \mathrm{Tran}_{-t} f, h_I\rangle h_I .$$

This operator translates with translation and hence is convolution. We can write $\mathrm{G}_j f = \gamma_j * f$. By the dilation invariance of the Haar functions, we will have $\gamma_j = \mathrm{Dil}^1_{2^j}\,\gamma_0$. A short calculation shows that

$$\gamma_0(y) = \int_0^1 h_I(y+t)h_I(y)\; dt$$

This function is depicted in Figure 1.4. Certainly the operator $\sum_j G_j$ is convolution with $\sum_j \gamma_j(x)$. This kernel is odd and is strictly positive on $[0,\infty)$. This finishes our proof.

□

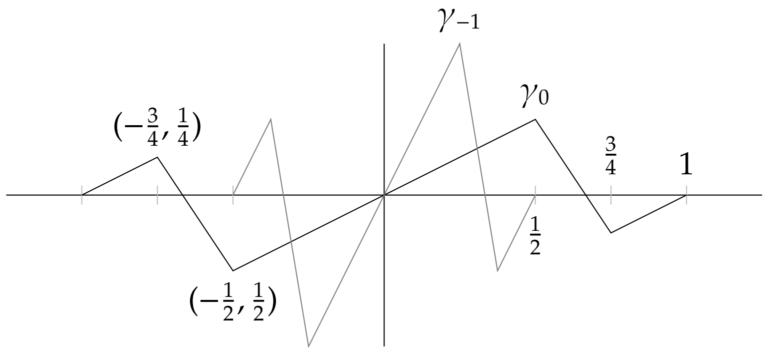


Figure 1.4: The graph of γ_0 and γ_{-1}.

1.4.1 Haar Shifts

Related to the definition of the operator $\mathfrak{H}$, there are other operators whose action on Haar functions is simple to describe. These operators, together with those of Paraproducts, individually capture a full range of Calderón Zygmund operators. Indeed, under composition, and summation, these operators can be used to derive a rich class of these operators.

Haar Multipliers The first 'shift' will be just a multiplicative change in signs of the Haar coefficients. For a map $\sigma : \mathcal{D} \longrightarrow \mathbb{R}$, set

$$\text{Ш}_{\text{sign},\sigma} f := \sum_{I \in \mathcal{D}} \sigma(I) \langle f, h_I \rangle h_I \,. \tag{1.4.6}$$

This completely explicit definition could have been given by

$$\text{Ш}_{\text{sign},\sigma} h_I := \sigma(I) h_I \,,$$

and then extending the definition of the operator linearly.

It is immediately clear that this operator can be bounded on L^2 iff σ takes bounded values. The remarkable thing is that this characterization continues to hold on all L^p.

Scale Shift. Let $\sigma : \mathcal{D} \longrightarrow \mathcal{D}$ be a map so that $\sigma(I) \subset I$, and $\sigma(I) = 2^s |I|$, where s is a fixed negative integer. The Scale Shift operator is

$$\text{Ш}_{\text{Scale},\sigma} f = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_{\sigma(I)} \,. \tag{1.4.7}$$

Location Shift. Let $n \in \mathbb{Z}$, and define the Location Shift operator by

$$\text{Ш}_{\text{Loc},n} f == \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_{I + n|I|} \tag{1.4.8}$$

In higher dimensions, the Haar functions have a *signature*, and one can define a corresponding Signature Shift operator.

1.4.2 The Conjugate Function

The Poisson kernel on the positive half-plane $\mathbb{R}^2_+ := \mathbb{R} \times [0, \infty)$ is given by

$$P(x, y) = \frac{1}{\pi} \cdot \frac{y}{x^2 + y^2}, \qquad x \in \mathbb{R}, \; y > 0 \tag{1.4.9}$$

This kernel is a solution in $\mathbb{R}^2_+$ of the Laplace equation with Dirichlet conditions

$$\Delta u = 0, \qquad u(x, 0) = \delta_0, \tag{1.4.10}$$

where δ_0 is the Dirac delta at $x = 0$. This is verified by direct computation. $\Delta P(x, y) \equiv 0$ for $(x, y) \in \mathbb{R}^2_+$, while $P(\cdot, y) \to \delta_0$ as $y \to 0$. The latter convergence is understood in the sense of distributions.

It follows that for Schwartz function f, that the equation

$$\Delta u = 0, \qquad u(x, 0) = f(x),$$

is solved by

$$u(x, y) = \int P(x, t) f(y - t) \, dt.$$

The condition $\Delta u = 0$ means that u is harmonic. It is a classical fact that there is a second function v on $\mathbb{R}^2$ for which $u + iv$ is an analytic function. v is referred to as the *harmonic conjugate* to u. The Cauchy Riemann equations relate u and v through the equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, , \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial y} \tag{1.4.11}$$

It turns out that v solves the equation

$$\Delta u = 0, \qquad u(x, 0) = \mathrm{H} f.$$

That is, the Hilbert transform carries the boundary values f into the *conjugate* boundary values $\mathrm{H} f$.

The Harmonic equation $\Delta u = 0$ is only the simplest instance of an elliptic partial differential equation. And what we point to here is the simplest instance of a common phenomena in partial differential equations which singular integrals play a key role in describing the solutions to these equations. The properties of singular integrals, and their non–local behaviors, determine the behavior of the solutions of these equations.

1.4.3 Analytic Decompositions

We permit the functions in $L^2(\mathbb{R})$ to take complex values. The operators

$$\mathrm{P}_\pm f(x) := \int_{\mathbb{R}_\pm} \widehat{f}(\xi)\, \mathrm{e}^{ix\xi}\; dx \tag{1.4.12}$$

are Fourier projections onto $\mathbb{R}_\pm$. By Plancherel's Theorem, they are bounded operators on $L^2(\mathbb{R})$. But, since the Hilbert transform has a particularly simple behavior with respect to the Fourier transform, namely $\widehat{\mathrm{H}\, f}(\xi) = -i\,\mathrm{sgn}(\xi)\widehat{f}(\xi)$, it follows that P is a linear combination of the identity and and the Hilbert transform. Hence, P actually maps $L^p(\mathbb{R})$ into $L^p(\mathbb{R})$ for all $1 < p < \infty$.

Specializing to $L^2(\mathbb{R})$, this space is the direct sum of the image of $L^2(\mathbb{R})$ under P, and its complement. We refer to $\mathrm{P}\, L^2(\mathbb{R})$ as the Hardy space $H^2(\mathbb{R})$ of functions in $L^2(\mathbb{R})$ with analytic extension to $\mathbb{R}^2$. Crudely, $\mathrm{P}\, f$ is the *analytic* part of f. The orthocomplement space $H^2_-(\mathbb{R}) = L^2(\mathbb{R}) \ominus H^2(\mathbb{R})$ is the *antianalytic* Hardy space.

1.5 The Hardy–Littlewood Maximal Function

A key maximal operator is

$$\mathrm{M}\, f(x) := \sup_{t>0} (\mathrm{Dil}^1_t\, \mathbf{1}_{[-1/2,1/2]}) * |f| = \sup_{t>0} \frac{1}{|I|} \int_{-t}^{t} |f(x-y)|\, dy.$$

At a given point x, this the supremum of all averages of f over an interval centered at x. The fundamental facts about this operator are

1.5.1 L^p Bound for the Hardy–Littlewood Maximal Function. *For integrable functions, we have the weak type estimate*

$$|\{\mathrm{M}\, f > \lambda\}| \lesssim \frac{\|f\|_1}{\lambda}, \qquad 0 < \lambda < \infty. \tag{1.5.2}$$

For $1 < p \le \infty$, M maps L^p into itself. In particular, we have the estimates

$$\|\mathrm{M}\|_p \lesssim \frac{p}{p-1}, \qquad 1 < p \le \infty. \tag{1.5.3}$$

A primary interest in maximal functions is indicated by the following corollary.

1.5.4 Lebesgue Differentiation Theorem. *For any function $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ we have*

$$\lim_{t\to 0} \frac{1}{2t} \int_{-t}^{t} f(x-y))\, dy = f(x) \qquad \text{a. e.} \tag{1.5.5}$$

By $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ we mean that for each $x \in \mathbb{R}$ there is a non-empty interval I containing x so that $f\mathbf{1}_I$ is integrable. Thus, f is locally in L^1.

Proof. The point of this proof is to illustrate the implications of maximal inequalities. Fix a finite interval I on which f is integrable. Consider the subspace $X \subset L^1(I)$ for which the convergence (1.5.5) holds. Clearly, X is a linear subspace of $L^1(I)$.

Also, every C^1 function f is also in X. Indeed, we can estimate

$$\Big|f(x) - \frac{1}{2t}\int_{-t}^{t} f(x-y))\,dy\Big| = \frac{1}{2t}\Big|\int_{x-y}^{x} f'(u)\,du\Big|\,dy \\ \le \|f'\|_\infty t\,.$$

Thus, X is dense in $L^1(I)$. And it remains to show that X is in fact a closed subspace of $L^1(I)$. This is in fact the consequence of the maximal function bound.

Let f be a point in the closure of X in the L^1 metric. Let $g \in X$ be a function which approximates f. Then, for x in the interior of I and $\epsilon > 0$ sufficiently small

$$\Big\{x \,:\, \limsup_{t\to 0}\Big|f(x) - \frac{1}{2t}\int_{-t}^{t} f(x-y))\,dy\Big| > \epsilon\Big\} \subset \{x \,:\, \mathrm{M}|f-g|(x) > \epsilon\}$$

The latter set can be estimated in measure, since we know the bounds for the Hardy Littlewood Maximal function, and so can be made arbitrarily small by choosing $g \in X$ sufficiently close to f in L^1 norm. It follows that the set below is a null set, which concludes the proof of the Theorem.

$$\Big\{x \in I \,:\, \limsup_{t\to 0}\Big|f(x) - \frac{1}{2t}\int_{-t}^{t} f(x-y))\,dy\Big| > 0\Big\}.$$

□

Let us first observe that the estimate at $p = 1$ is sharp. For if $f = \mathbf{1}_{[-1,1]}$, and one seeks a lower bound for $\mathrm{M}\,f(x)$ for $x > 2$, take the interval centered at x to have left hand endpoint being the origin. Then, $\mathrm{M}\,f(x) \ge 1/2x$. This long range effect prevents $\mathrm{M}\,f$ from being integrable. We leave it to the reader to verify that

$$\|\mathrm{M}\,f\|_p \gtrsim \frac{p}{p-1}, \qquad 1 < p < 2.$$

Hence the rate of growth of the norms in (1.5.3) is optimal.

The definition of the maximal function is quite flexible (a fact that we shall take advantage of in the proof of the Theorem). This is indicated in part by this

1.5.6 Proposition. *Let $\varphi : \mathbb{R} \to (0,\infty)$ be a symmetric decreasing function, with $\varphi(x) \to 0$ as $x \to \infty$. We then have the estimate*

$$\sup_{t>0}(\mathrm{Dil}^1_t\,\varphi) * |f| \le \mathrm{M}\,f \int_0^\infty -t\varphi'(t)\,dt.$$

Proof. One can write, for $x > 0$,

$$\varphi(x) = -\int_x^\infty \varphi'(s)\, ds = -\int_0^\infty \operatorname{Dil}_s^1 \mathbf{1}_{[-1,1]}(x)\, s\varphi'(s)\, ds.$$

Hence, it is the case that

$$\begin{aligned}(\operatorname{Dil}_t^1 \varphi) * f(x) &= \int_0^\infty (\operatorname{Dil}_{st}^1 \mathbf{1}_{[-1,1]}) * f(x)\, s\varphi'(s)\, ds \\ &\leq \mathrm{M} f(x) \int_0^\infty |s\varphi'(s)|\, ds.\end{aligned}$$

And the proposition follows. □

1.5.1 Grids and the Weak L^1 Inequality

The maximal function can never take a value greater than the sup norm of f. Hence, the L^∞ bound for the maximal function is immediate. We prove in this section the weak type bound at L^1. An interpolation argument will establish the bounds for all $1 < p < \infty$.

Given a collection of intervals $\mathcal{I}$, let us set

$$M^{\mathcal{I}} f = \sup_{I \in \mathcal{I}} \frac{\mathbf{1}_I}{|I|} \int_I f\, dy.$$

The usefulness of grids to us is made clear by the elementary

1.5.7 Proposition. *If $\mathcal{I}$ is a grid (See Definition 1.2.1.), then we have the inequality*

$$|\{M^{\mathcal{I}} f > \lambda\}| \leq \frac{\|f\|_1}{\lambda}.$$

Proof. The grid structure implies that the intervals in a grid are ordered by inclusion. Thus, for any subset of $\mathcal{I}$ for which the intervals in $\mathcal{I}$ have a bounded length contains maximal elements with respect to inclusion.

The set $\{M^{\mathcal{I}} f > \lambda\}$ is a union of intervals from the collection $\mathcal{I}$. Hence, there is a collection $\mathcal{J} \subset \mathcal{I}$ of pairwise disjoint intervals for which we have

$$\{M^{\mathcal{I}} f > \lambda\} = \bigcup_{J \in \mathcal{J}} J.$$

Obviously, for each interval $J \in \mathcal{J}$ we have $\int_J f\, dy \geq \lambda |J|$. Thus, it is the case that

$$\lambda |\{M^{\mathcal{I}} f > \lambda\}| \leq \sum_{J \in \mathcal{J}} \int_J f\, dy \leq \|f\|_1$$

□

Thus, for the dyadic intervals $\mathcal{D}$, we have $\mathrm{M}^{\mathcal{D}}$ satisfies the weak type inequality. This does not imply the same inequality for the Hardy–Littlewood maximal function. But observe that

$$\mathrm{M}\, f \le 16\, \mathrm{M}^{\mathcal{D}\cup\mathcal{D}'} f.$$

Here $\mathcal{D}'$ is as in (1.2.2). Indeed, this is a consequence of Exercise 1.2.6. For any interval I, we can select $J \in \mathcal{D} \cup \mathcal{D}'$ with $I \subset J$ and $|J| \le 16|I|$, so that

$$|I|^{-1} \int_I f(y)\, dy \le 16|J|^{-1} \int_J f(y)\, dy\,.$$

1.5.2 The Interpolation Argument

The deduction of the L^p inequalities for the maximal function are a consequence of the more general Marcinciewicz Interpolation Theorem. In this instance, the argument is simple enough that we give it directly. The weak L^1 bound is a distributional estimate. So we should relate L^p norms to distributional estimates. This is done by way of the formula

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

This is readily checked by using integration by parts.

Additionally for the maximal function, and $f \in L^1 \cap L^p$, observe that

$$\{\mathrm{M}\, f > \lambda\} \subset \{\mathrm{M}(f\mathbf{1}_{\{f>\lambda/2\}}) > \lambda/2\}.$$

Thus, we see that

$$\begin{aligned}
\|\mathrm{M}\, f\|_p^p &= p \int_0^\infty \lambda^{p-1} \{\mathrm{M}\, f > \lambda\}\, d\lambda \\
&\le p \int_0^\infty \lambda^{p-1} |\{M(f\mathbf{1}_{\{f>\lambda/2\}}\}|\, d\lambda \\
&\lesssim p \int_0^\infty \lambda^{p-2} \|f\mathbf{1}_{\{f>\lambda/2\}}\|_1 \\
&= p \int_0^\infty \int_{\lambda/2}^\infty |\{f > t\}|\, dt\, d\lambda \\
&= 2^p \frac{p}{p-1} \|f\|_p^p.
\end{aligned}$$

1.5.3 The TT^* Proof of the L^2 Bound

There is an instructive proof of L^2 bound for the Maximal Function, as it introduces two useful principles. The first, is the $\mathrm{T}\,\mathrm{T}^*$ identity, for the norm of an operator T acting on a Hilbert space $\mathcal{H}$:

$$\|\mathrm{T}\|^2 = \|\mathrm{T}\,\mathrm{T}^*\|\,. \tag{1.5.8}$$

Indeed, let $x \in \mathcal{H}$ be a norm one vector for which

$$\begin{aligned} \|\mathrm{T}\|^2 &= \|\mathrm{T}^*\|^2 \\ &= \langle \mathrm{T}^* x, \mathrm{T}^* x\rangle \\ &= \langle \mathrm{T}\,\mathrm{T}^* x, x\rangle \\ &\le \|\mathrm{T}\,\mathrm{T}^*\|\,. \end{aligned}$$

This proves half of (1.5.8), with the other half being obvious. To use (1.5.8) to prove the boundedness of some operator, one proves that for a positive operator T, that one has $\mathrm{T}\,\mathrm{T}^* \le K(\mathrm{T} + \mathrm{T}^*)$, whence $\|\mathrm{T}\| \le K$.

The second is that in considering Maximal Functions, it is frequently useful to pass to a *linearization*. Namely, the boundedness of the Maximal Function is equivalent to the boundedness of a family of linear operators, defined as follows. To each $I \in \mathcal{D}$, associate $E(I) \subset I$, so that the sets $\{E(I) \,:\, I \in \mathcal{D}\}$ are disjoint subsets. Define

$$\mathrm{T} f = \sum_{I\in\mathcal{D}} \frac{\mathbf{1}_{E(I)}}{|I|} \int_I f(y)\,dy\,. \tag{1.5.9}$$

It is clear that $\mathrm{T} f \le \mathrm{M} f$ pointwise, regardless how the sets $E(I)$ are selected. On the other hand, If we take $E(I)$ to be the set of those $x \in I$ where the supremum in the definition of $\mathrm{M} f(x)$ is achieved, up to a multiplicative factor of 2 say, by the average over I, then we reverse the inequality. $\mathrm{M} f \le 2\,\mathrm{T} f$. The notion of a linearization is useful in the analysis of many maximal operators.

To bound the Maximal Function, it therefore suffices to find an absolute bound on any linearization T as in (1.5.9). This we do in the case of $p = 2$ using the $\mathrm{T}\,\mathrm{T}^*$ approach. We calculate

$$\begin{aligned} \mathrm{T}\,\mathrm{T}^* f &= \sum_{I\in\mathcal{D}} \sum_{J\in\mathcal{D}} \mathbf{1}_{E(I)} \frac{\langle \mathbf{1}_I, \mathbf{1}_J\rangle}{|I|} \frac{\langle \mathbf{1}_{E(J)}, f\rangle}{|J|} = I + I\!I \\ I &= \sum_{I,J\,:\, I\subset J} \mathbf{1}_{E(I)} \frac{\langle \mathbf{1}_{E(J)}, f\rangle}{|J|} \\ I\!I &= \sum_{I,J\,:\, I\supsetneq J} \mathbf{1}_{E(I)} \frac{\langle \mathbf{1}_{E(J)}, f\rangle}{|I|} \end{aligned}$$

In I, sum over I to see that $I \le \mathrm{T}^* f$. And, in $I\!I$, sum over J to see that $I\!I \le \mathrm{T} f$. Thus, $\mathrm{T}\,\mathrm{T}^* f \le \mathrm{T} f + \mathrm{T}^* f$ pointwise, for positive f. This completes the proof of the L^2 estimate for the Maximal Function.

1.5.4 Exercises

1.5.10 Exercise. For a function $f : [0,1] \longrightarrow \mathbb{R}$, show that

$$\int_0^1 |f| \log_+ |f|\,dx \simeq \int_0^\infty \log_+ \lambda |\{f > \lambda\}|\,d\lambda$$

where $\log_+ x = \max(1, \log x)$.

1.5.11 Exercise. Prove this Theorem of E. M. Stein. Show that if $f \in L \log L([0,1])$, then $\mathrm{M} f$ is integrable on the same interval.

1.5.12 Exercise. Show that for a function f on $[0,1]$, with $\mathrm{M} f$ integrable on $[0,1]$, then $f \in L(\log L)$.

1.5.13 Exercise. Prove the Vitali Covering Lemma: For any finite collection $\mathcal{I}$ of intervals, there is a sub collection $\mathcal{I}' \subset \mathcal{I}$, of pairwise disjoint intervals, for which

$$\bigcup_{I \in \mathcal{I}} I \subset \bigcup_{I' \in \mathcal{I}'} 2I.$$

The construction of $\mathcal{I}'$ is an application of the greedy heuristic.

1.5.14 Exercise. Use the Vitali Covering Lemma to prove that the maximal function maps L^1 into weak L^1.[2]

1.5.15 Exercise. Let ϕ be a symmetric decreasing function on $\mathbb{R}$, with $\phi(x) \to 0$ as $x \to \infty$, and assume in addition that

$$-\int_0^\infty t\phi'(t)\, dt < \infty.$$

Show that the maximal function below maps L^1 into weak L^1.

$$\mathrm{M} f(x) := \sup_{t>0} (\mathrm{Dil}_t^1 \phi) * f(x).$$

The main point to observe is that for $x > 0$ we have $\phi(x) = -\int_0^\infty \mathbf{1}_{[-t,t]} \phi'(t)\, dt$.

1.6 Fractional Integral Operators

The object of study here are the so-called *fractional integral operators* defined by

$$\mathrm{I}_\alpha f := \int f(x-y) \frac{dy}{|y|^{1-\alpha}}, \qquad 0 < \alpha < 1 . \tag{1.6.1}$$

Let us begin with a simple dimension-counting exercise to understand what the mapping properties of this transform should be. We agree that dx is measured in units of length ℓ. Then the units of $\|f\|_p$ are $\ell^{1/p}$. The units of $\mathrm{I}_\alpha f$ are ℓ^α, and of $\|\mathrm{I}_\alpha f\|_q$ are $\ell^{\alpha+1/q}$. Therefore, in order for I_α to map L^p to L^q, we need to have $\alpha + 1/q = 1/p$. This is the task that we turn to next.

1.6.2 Theorem. *Suppose* $0 < \alpha < 1$, *and* $1 < p, q < \infty$ *are related by* $1/p = 1/q + \alpha$. *Then we have*

$$\|\mathrm{I}_\alpha f\|_q \lesssim \|f\|_p . \tag{1.6.3}$$

[2] This method of proof is more commonly known than the one we adopted.

Our proof illustrates a method to pass from a discrete model of a continuous operator, like I_α. But what is the discrete model of I_α? Consider

$$\mathrm{J}_\alpha f := \sum_{I\in\mathcal{D}} \frac{\langle f, \mathbf{1}_I\rangle}{|I|^{1-\alpha}} \mathbf{1}_I \,. \tag{1.6.4}$$

This operator is positive like I_α, and has the same dimensions. So the basic facts should be as in the Theorem above.

1.6.5 Proposition. *Suppose* $0 < \alpha < 1$, *and* $1 < p, q < \infty$ *are related by* $1p = 1/q + \alpha$. *Then we have*

$$\|\mathrm{J}_\alpha f\|_q \lesssim \|f\|_p \,. \tag{1.6.6}$$

Let us see how to recover the continuous I_α from J_α.

1.6.7 Proposition. *For* $0 < \alpha < 1$, *there is a non-zero constant* $c = c(\alpha)$ *so that*

$$\mathrm{I}_\alpha = \lim_{Y\to\infty} \int_0^Y \mathrm{Tr}_{-y}\, \mathrm{J}_\alpha\, \mathrm{Tr}_y \,\frac{dy}{Y}$$

Proof. The operator I_α is characterized as being a convolution operator with a positive even kernel, and satisfies

$$\mathrm{I}_\alpha \circ \mathrm{Dil}^p_\lambda = \mathrm{Dil}^q_\lambda \circ \mathrm{I}_\alpha \,, \qquad \lambda > 0 \,.$$

Here p, q are any choices of indices as in Theorem 1.6.2.

Note that our discrete operator J_α satisfies an approximate to this last property

$$\mathrm{I}_\alpha \circ \mathrm{Dil}^p_{2^j} = \mathrm{Dil}^q_{2^j} \circ \mathrm{I}_\alpha \,, \qquad j \in \mathbb{Z} \,. \tag{1.6.8}$$

But it is not translation invariant, and it need not pointwise dominant I_α due to a standard issue with respect to the dyadic grid, namely that the dyadic grid $\mathcal{D}$ distinguishes points. For instance, the points ± 1 are not both contained in a dyadic interval. This is remedied by the averaging over translations made above.

If f is a Schwarz function, it is straight forward to see that

$$\lim_{Y\to\infty} \int_0^Y \mathrm{Tr}_{-y}\, \mathrm{J}_\alpha\, \mathrm{Tr}_y\, f(x) \,\frac{dy}{Y} \qquad \text{exists for all } x.$$

Let us call the limiting operator $\widetilde{J}_\alpha f$. It follows that $\widetilde{J}_\alpha$ commutes with all translations, and that it satisfies the dilation conditions (1.6.8), and has a positive even kernel. Hence it must be a multiple of I_α. That it is a non-zero multiple is easy to verify.

□

And so we turn to the proof of the basic inequalities for fractional integration.

Proof of Theorem 1.6.2. We will prove that the operators J_α satisfy the estimates of the Theorem, and so by Proposition 1.6.7 we deduce the Theorem.

The combinatorics of the dyadic grid will be helpful to us in making a direct comparison of the operator J_α to the Maximal Function. Fix a positive $f \in L^p$ and set

$$\mathcal{D}_k = \Big\{ I \in \mathcal{D} \,:\, \frac{\langle f, \mathbf{1}_I \rangle}{|I|} \le 3^k \Big\} .$$

Let set $\mathcal{D}_k^*$ to be the maximal intervals in $\mathcal{D}_k$.

We are not concerned with the fact that the collections $\mathcal{D}_k$ are not disjoint. The estimate we need is that

$$\sum_{I \in \mathcal{D}_k} \frac{\langle f, \mathbf{1}_I \rangle}{|I|^{1-\alpha}} \mathbf{1}_I \lesssim 3^k \sum_{I \in \mathcal{D}_k^*} |I^*|^\alpha \mathbf{1}_{I^*} .$$

This is straight forward to see, as we only appeal to the inequality $\sum_{j \le j_0} 2^{\alpha j} \lesssim 2^{j_0 \alpha}$.

Set $D_k = \bigcup \{ I^* \,:\, I^* \in \mathcal{D}_k^* \}$. These sets are increasing in k, as the collections $\mathcal{D}_k$ increase in k. Thus, our estimate for $\mathrm{J}_\alpha f$ is as follows.

$$\begin{aligned} \mathrm{J}_\alpha f &\lesssim \sum_{k=-\infty}^{\infty} 3^k \sum_{I^* \in \mathcal{D}_k^*} |I^*|^\alpha \mathbf{1}_{I^*} \\ &\lesssim \sum_{k=-\infty}^{\infty} 3^k \mathbf{1}_{D_k \backslash D_{k-1}} \sum_{I^* \in \mathcal{D}_k^*} |I^*|^\alpha \mathbf{1}_{I^*} \end{aligned}$$

But the sets $D_k \backslash D_{k-1}$ are themselves disjoint sets, hence

$$\begin{aligned} \|J_\alpha f\|_q^q &\lesssim \sum_{k=-\infty}^{\infty} 3^{qk} \Big\| \mathbf{1}_{D_k \backslash D_{k-1}} \sum_{I^* \in \mathcal{D}_k^*} |I^*|^\alpha \mathbf{1}_{I^*} \Big\|_q^q \\ &\lesssim \sum_{k=-\infty}^{\infty} 3^{qk} \sum_{I^* \in \mathcal{D}_k^*} |I^*|^{1+q\alpha} \\ &\lesssim \|f\|_p^{q-p} \sum_{k=-\infty}^{\infty} 3^{kp} \sum_{I^* \in \mathcal{D}_k^*} |I^*| \\ &\lesssim \|f\|_p^q . \end{aligned}$$

Here, we have relied upon the observation that

$$\sup_I 3^k |I|^{1/p} \lesssim \|f\|_p ,$$

as well as appealing to the Maximal Function estimate in the last line. □

1.7 The Calderón Zygmund Decomposition and the Weak Type Bound

The weak-type estimate for the Maximal Function, see Proposition 1.5.7, is an easy consequence of a Covering Lemma. But there is no corresponding proof for the Hilbert transform. Deeper methods are necessary, and the most powerful and flexible method is the Calderón Zygmund Decomposition.

1.7.1 Theorem: Calderón Zygmund Decomposition. *For $f \in L^1(\mathbb{R})$ and $\lambda > 0$ we can write $f = g + b$, where g is a 'good' function, and b is a 'bad' function. These two functions satisfy*

$$\|g\|_2 \le 2\lambda \|f\|_1 , \tag{1.7.2}$$

and b is a sum of b_j, for $j \ge 1$, where the functions b_j are supported on disjoint dyadic intervals I_j, where

$$\Big|\bigcup_j I_j\Big| \lesssim \lambda^{-1} \|f\|_1 , \tag{1.7.3}$$

$$\int_{I_j} b_j(y)\, dy = 0 . \tag{1.7.4}$$

Proof. Take the intervals I_j, $j \ge 1$, to be the maximal dyadic intervals in the set $\{\mathrm{M}^{\mathcal{D}}|f| > \lambda\}$. Define g as follows:

$$g(x) = \begin{cases} f(x) & x \notin \bigcup_j I_j \\ \frac{1}{|I_j|}\int_{I_j} f(y)\, dy & x \in I_j\,,\, j \ge 1 \end{cases} \tag{1.7.5}$$

It follows that $\|g\|_\infty \le 2\lambda$, so that (1.7.2) holds.

To continue, we set

$$b_j(x) := \mathbf{1}_{I_j}(x)\Big[f(x) - |I_j|^{-1} \int_{I_j} f(y)\, dy \Big] . \tag{1.7.6}$$

It is clear that the conclusions of the Decomposition holds. □

1.7.7 Theorem. *For both operators* $\mathrm{T} = \mathrm{H}$*, the Hilbert transform, or* $\mathrm{T} = \mathfrak{H}$ *as defined in* (1.4.3)*, map L^1 into $L^{1,\infty}$, that is we have the inequality*

$$\sup_{\lambda > 0} \lambda |\{\mathrm{T} f > \lambda\}| \lesssim \|f\|_1 . \tag{1.7.8}$$

Proof. Consider first $\mathrm{T} = \mathfrak{H}$. Take $\lambda > 0$, and $f \in L^1$, and apply the Calderón Zygmund Decomposition. Then, write $\mathfrak{H} = \mathfrak{H}_+ + \mathfrak{H}_-$, where

$$\mathfrak{H}_- \phi = \sum_{j=1}^{\infty} \sum_{I \,:\, I \subset I_j} \langle \phi, h_I \rangle g_I .$$

Then, note that $\mathfrak{H}_{-}g = 0$, since the good function is constant on each interval I_j. And $\mathfrak{H}b$ is supported on $E = \bigcup_j I_j$. Therefore, for $x \notin E$, we have $\mathfrak{H}f(x) = \mathfrak{H}_{+}f(x) = \mathfrak{H}_{+}g(x)$. And, we can prove (1.7.8) by estimating as follows.

$$\begin{aligned} |\{\mathfrak{H}f > \lambda\}| &\le |E| + |\{\mathfrak{H}f > \lambda\}| \\ &\lesssim \|f\|_1 + |\{x \in \mathbb{R} - E \ :\ \mathfrak{H}f > \lambda\}| \\ &\lesssim \|f\|_1 + |\{x \in \mathbb{R} - E \ :\ \mathfrak{H}g > \lambda\}| \\ &\lesssim \|f\|_1 + \lambda^{-2}\|g\|_2^2 \\ &\lesssim \|f\|_1 \,. \end{aligned}$$

The point is that one does not attempt to estimate $\mathfrak{H}f$ on the set E, whence one can substitute f for the square-integrable g above.

We should emphasize that despite the truth of Proposition 1.4.4, we cannot use the weak-type inequality for $\mathfrak{H}$ to deduce the same inequality for the Hilbert transform. The reason is that not only is the $L^{1,\infty}$-norm a quasi-norm (that is, the triangle inequality holds with a constant larger than one), but it also does not admit an equivalent norm. See Exercises 1.7.9 and 1.7.11.

Thus, we have to argue directly to see that the same inequality holds for the Hilbert transform. The only point is that the smoothness condition on the kernel must be used to control the 'bad' function. Fix $f \in L^1$, a finite sum of Haar functions. Apply the decomposition above, thus $f = g + b$. In the first step, we address the fact that the weak L^1 norm is in fact only a quasi norm.

$$\{|\mathrm{H}\, f| > 2\lambda\} \subset \{|\mathrm{H}\, g| > \lambda\} \cup \{|\mathrm{H}\, b| > \lambda\}\,.$$

Using the L^2 estimate on the good function, we have

$$|\{|\mathrm{H}\, g| > \lambda\}| \le \lambda^{-2}\|g\|_2^2 \le 2\lambda^{-1}\|f\|_1\,.$$

For the bad function, recall that the function $b = \sum_j b_j$, and b_j is supported on the disjoint interval I_j. We do not attempt to make any estimate of $\mathrm{H}\, b$ on the set

$$\Big|\bigcup_j 2I_j\Big| \lesssim \lambda^{-1}|\{\mathrm{M}\, f > \lambda\}| \lesssim \lambda^{-1}\|f\|_1\,.$$

Now, $\mathrm{H}\, b_j(x)$ for $x \notin 2I_j$ admits a good pointwise estimate, due to the mean zero property of the b_j, and the smoothness of the kernel $1/y$:

$$\begin{aligned} |\mathrm{H}\, b_j(x)| &= \Big|\int_{I_j} b_j(y)\frac{dy}{x-y}\Big| \\ &= \Big|\int_{I_j} b_j(y)\Big\{\frac{1}{x-y} - \frac{1}{x-c(I_j)}\Big\}\, dy\Big| \\ &\lesssim \frac{|I_j|}{|x-c(I_j)|^2}\|b_j\|_1 \end{aligned}$$

$$\lesssim \lambda \frac{|I_j|^2}{|x - c(I_j)|^2}$$

So we see that

$$\|\mathrm{H}\, b_j\|_{L^1(\mathbb{R}-2I_j)} \lesssim \lambda |I_j| \,.$$

Setting $E = \bigcup_j 2I_j$, it follows that

$$\|\mathrm{H}\, b\|_{L^1(\mathbb{R}-E)} \lesssim \lambda \sum_j |I_j| \lesssim \|f\|_1 \,.$$

This permits us to conclude that

$$|\{|\mathrm{H}\, b| > \lambda\}| \leq |E| + |\{x \notin E \,:\, |\mathrm{H}\, b| > \lambda\}| \lesssim \lambda^{-1} \|f\|_1 \,.$$

Our proof of the weak type bound is complete.

□

Exercises

1.7.9 Exercise. Defining

$$\|f\|_{1,\infty} = \sup_{\lambda>0} \lambda |\{f > \lambda\}| \,, \tag{1.7.10}$$

show that

$$\|f + g\|_{1,\infty} \leq 2\{\|f\|_{1,\infty} + \|g\|_{1,\infty}\}$$

but, we need not have $\|f + g\|_{1,\infty} \leq \|f\|_{1,\infty} + \|g\|_{1,\infty}$. In other words, $\|\cdot\|_{1,\infty}$ is a quasi-norm, but not a norm. It is a deeper fact that there is no other norm $[\![\cdot]\!]$ and finite positive constant K with $K^{-1}[\![f]\!] \|f\|_{1,\infty} \leq K[\![f]\!]$ for all functions f.

1.7.11 Exercise. Defining

$$\|f\|_{p,\infty} = \sup_{\lambda>0} \lambda |\{f > \lambda\}|^{1/p} \,, \qquad 1 \leq p < \infty \,, \tag{1.7.12}$$

show that

$$\|f + g\|_{p,\infty} \leq 2^{1/p}\{\|f\|_{1,\infty} + \|g\|_{1,\infty}\}$$

For $1 < p < \infty$, there is a norm $[\![\cdot]\!]_p$ and finite positive constant K_p with $K_p^{-1}[\![f]\!]_p \|f\|_{1,\infty} \leq K_p[\![f]\!]_p$ for all functions f.

1.8 The Sharp Function

The following variant of the maximal function focuses not on the values of the function, but its deviations from local mean values.

$$f^\sharp(x) = \sup_{x\in J} \frac{1}{|I|} \int_J \Big| f - |J|^{-1} \int_J f \, dt \Big| \, dy\,. \tag{1.8.1}$$

We refer to it as the *sharp function.* Of great importance is that one has the equivalence of norms $\|f\|_p \simeq \|f^\sharp\|_p$, for all $1 < p < \infty$.

1.8.2 Theorem. *For all $1 < p < \infty$, we have $\|f\|_p \simeq \|f^\sharp\|_p$.*

Now, it is clear that $f^\sharp(x) \le 2\,\mathrm{M} f(x)$, so that we clearly have $\|f^\sharp\|_p \lesssim \|f\|_p$ for $1 < p < \infty$. And so the content of the Theorem is that we have the reverse inequality.

But is also the case that the pointwise estimate $\mathrm{M} f \lesssim f^\sharp$ must fail completely. What we should show however, is that this estimate must in some sense hold most of the time. We shall do so by way of an important technique, a distributional estimate of somewhat sophisticated formulation. Such estimates are referred to as *good-λ inequalities.*

1.8.3 Lemma. *For all $\lambda > 0$, all $0 < \eta < 1$, and all f in some L^p class, for finite $p > 1$, we have*

$$|\{\mathrm{M} f > 2\lambda,\ f^\sharp < \eta\lambda\}| \le \eta |\{\mathrm{M} f > \lambda\}|. \tag{1.8.4}$$

Proof. The set $\{\mathrm{M} f > \lambda\}$ is a union of maximal dyadic intervals J such that $|J|^{-1} \int_J |f| dx \ge \lambda^2$. Fix such an interval J. We should show that

$$|\{x \in J \ :\ \mathrm{M} f > 2\lambda,\ f^\sharp < \eta\lambda\}| \le \eta|J|.$$

This estimate is then summed over J to conclude the Lemma.

We may suppose that there is some $x \in J$ for which $\mathrm{M} f(x) > 2\lambda$, and yet $f^\sharp(x) < \eta\lambda$. Let

$$g = \Big(f - |J|^{-1} \int_J f \, dy\Big)\mathbf{1}_J$$

Then, it is the case that $\|g\|_1 \le \eta\lambda|J|$, and $\mathrm{M} g(x) > \lambda$. Thus, using the weak type estimate at L^1 for the maximal function,

$$\begin{aligned} |\{x \in J \ :\ \mathrm{M} f(x) > 2\lambda,\ f^\sharp(x) < \eta\lambda\}| &\le |\{\mathrm{M} g(x) > \lambda\}| \\ &\le \lambda^{-1}\|g\|_1 \le \eta|J|. \end{aligned}$$

□

Proof of Theorem 1.8.2. We only need to prove that $\|f\|_p \lesssim \|f^\sharp\|_p$. In fact, we will show that

$$\|f\|_p \le \|\mathrm{M} f\|_p \lesssim \|f^\sharp\|_p\,, \qquad 1 < p < \infty\,.$$

Now, for $\lambda > 0$,

$$\begin{aligned}|\{\mathrm{M} f > 2\lambda\}| &\le |\{f^\sharp > \eta\lambda\}| + |\{\mathrm{M} f > 2\lambda,\ f^\sharp < \eta\lambda\}| \\ &\le |\{f^\sharp > \eta\lambda\}| + \eta|\{\mathrm{M} f > \lambda\}|.\end{aligned}$$

We use the identity $\|g\|_p^p = p\int_0^\infty \lambda^{p-1}|\{g > \lambda\}|\,d\lambda$. With the inequality above, we see that

$$\begin{aligned}\|\mathrm{M} f\|_p^p &\le \int_0^\infty \lambda^{p-1}|\{f^\sharp > \eta\lambda/2\}|\,d\lambda + \eta\int_0^\infty \lambda^{p-1}|\{\mathrm{M}|f| > \lambda/2\}|\,d\lambda \\ &\le (2/\eta)^p\|f^\sharp\|_p^p + 2^p\eta\|\mathrm{M} f\|_p^p\end{aligned}$$

This inequality holds for all $0 < \eta < 1$, and if we take $\eta = 2^{-p-1}$, we will have the inequality

$$\|\mathrm{M} f\|_p^p \le 2^{p(p+2)}\|f^\sharp\|_p + \tfrac{1}{2}\|\mathrm{M} f\|_p^p\,,$$

which clearly proves the Theorem. □

1.8.5 Exercise. Define the following p-variant of the sharp function.

$$f^{\sharp,p}(x) = \sup_{I\in\mathcal{D}}\left[\frac{\mathbf{1}_I(x)}{|I|}\int_I\left|f(y) - |I|^{-1}\int_I f(z)\,dz\right|^p dy\right]^{1/p} \tag{1.8.6}$$

Show that $\|f\|_q \simeq \|f^{\sharp,p}\|_q$ for $p < q < \infty$.

1.8.7 Exercise. Let $f^{\sharp\sharp}$ be the sharp function without restriction on the the intervals being dyadic.

$$f^{\sharp\sharp}(x) = \sup_I \frac{\mathbf{1}_I(x)}{|I|}\int_I\left|f(y) - |I|^{-1}\int_I f(z)\,dz\right|dy \tag{1.8.8}$$

Show that $\|f\|_p \simeq \|f^{\sharp\sharp}\|_p$ for $1 < p < \infty$.

1.9 The Haar Littlewood-Paley Square Function

The Littlewood-Paley Square Function is a principle which manifests itself in many different forms. Of all of these, the most elementary is the Haar version, defined as

$$\mathrm{S}(f) := \left[\sum_{I\in\mathcal{D}}\frac{\langle f, h_I\rangle^2}{|I|}\mathbf{1}_I\right]^{1/2}. \tag{1.9.1}$$

Note that we are taking the difference between the average value of f on the left and right halfs of I, and summing up over all I. In flavor this is rather close to the sharp function, but simpler in its expression. Thus, the sharp function and the Square Function share the property of being alternate ways of computing L^p norms.

1.9.2 Theorem. *We have the equivalence of norms*

$$\|\mathrm{S}(f)\|_p \simeq \|f\|_p , \qquad 1 < p < \infty .$$

This is a profound result, with many consequences and ramifications. One of these is

1.9.3 Corollary. *For $1 < p < \infty$, the Haar basis is an unconditional basis of L^p. Namely, for all subsets $\mathcal{E} \subset \mathcal{D}$ the orthogonal projections*

$$\mathrm{P}_{\mathcal{E}} f = \sum_{I \in \mathcal{E}} \langle f, h_I \rangle h_I$$

extend to uniformly bounded operators on L^p to L^p.

While this seems like an abstract, infinitary property, it is in fact quite useful. If some operator has a reasonable expansion in the Haar basis, then this Corollary assures you that there are a wide variety of techniques at your disposal to bound the operator on L^p spaces.

Proof of Corollary 1.9.3. Estimate

$$\begin{aligned} \|\mathrm{P}_{\mathcal{E}} f\|_p &\simeq \|\mathrm{S}(\mathrm{P}_{\mathcal{E}} f)\|_p \\ &\leq \|\mathrm{S}(f)\|_p \\ &\simeq \|f\|_p , \end{aligned}$$

where the inequality is obvious. □

This proof illustrates another aspect of the Littlewood-Paley inequalities: Distinct scales of the Haar basis essentially decouple.

Proof of Theorem 1.9.2. At $p = 2$, we have $\|f\|_2 = \|\mathrm{S}(f)\|_2$, which follows from the fact that the Haar basis is an orthonormal basis for L^2. This fact we will appeal to in the the first stage proof, where we prove one-half of the claimed inequalities, namely

$$\|\mathrm{S}(f)\|_p \simeq \|f\|_p , \qquad 2 < p < \infty . \tag{1.9.4}$$

In this range of p's, functions are locally square integrable, which leads to an appeal of local Hilbertian methods. A basic fact for this proof is then

$$\|f\|_p \simeq \|f^{\sharp,2}\|_p , \qquad 2 < p < \infty . \tag{1.9.5}$$

This variant of Theorem 1.8.2 is indicated in Exercise 1.8.5.

Now, for any dyadic interval I we have

$$\int_I \Bigl| f - |I|^{-1} \int_I f dy \Bigr|^2 dx = \sum_{J \,:\, J \subset I} \langle f, h_J \rangle^2 . \tag{1.9.6}$$

Thus, it follows that

$$f^{\sharp,2} \le [\mathrm{M}\,\mathrm{S}(f)^2]^{1/2} .$$

So we conclude that

$$\|f\|_p \simeq \|f^{\sharp,2}\|_p \lesssim \|[\mathrm{M}\,\mathrm{S}(f)^2]^{1/2}\|_p \lesssim \|\mathrm{S}(f)\|_p , \qquad 2 < p < \infty .$$

This is one-half of the inequalities in (1.9.4).

For the other half, we dominate $\mathrm{S}(f)^{\sharp,2}$. Recall that for any function g,

$$\inf_c \int_0^1 |g - c|^2 \, dx = \int_0^1 \Bigl| g - \int_0^1 g \, dy \Bigr|^2 . \tag{1.9.7}$$

The import is this: The definition of the sharp function calls for subtracting off the mean value of the function in question on an interval. But, in seeking an upper bound on sharp function, we need not calculate the mean exactly, but merely make an informed guess of the mean. With this observation, it is straight forward to select c so that

$$\frac{1}{|I|} \int \Bigl| \mathrm{S}(f) - c \Bigr|^2 \le |I|^{-1} \sum_{J \,:\, J \subsetneq I} \langle f, h_J \rangle^2$$

Indeed, we simply take $c^2 = \sum_{K \,:\, \supset I} \frac{\langle f, h_K \rangle^2}{|K|}$, which permits us to subtract off all the terms that are constant on I. And so by (1.9.6), it follows that $\mathrm{S}(f)^{\sharp,2} \le f^{\sharp,2}$. Thus, we have completed the proof of (1.9.6).

We come to the second stage of the proof, in which we seek to extend the inequalities of (1.9.4) to $1 < p < 2$. Arguments of these types are, in the general case, harder as the locally square integrable property is lost. In the current context, we can prove the inequalities

$$\|\mathrm{S}(f)\|_p \simeq \|f\|_p , \qquad 1 < p < 2 ,$$

by methods of duality. The duality is the most straight forward in the following argument. For $f \in L^p$, with $1 < p < 2$, and of Schwartz class, let us take $g \in L^{p'}$ which is dual in the sense that $\|f\|_p = \langle f, g \rangle$, and $\|g\|_{p'} = 1$. Then, g is also of Schwartz class, hence both f and g are square integrable, and we can write

$$\|f\|_p = \langle f, g \rangle = \sum_I \langle f, h_I \rangle \langle h_I, g \rangle$$

$$\begin{aligned}
&\leq \sum_{I\in\mathcal{D}} \int \frac{|\langle f, h_I\rangle\langle h_I, g\rangle|}{|I|}\mathbf{1}_I\, dx \\
&\leq \langle \mathrm{S}(f), \mathrm{S}(g)\rangle \\
&\leq \|\mathrm{S}(f)\|_p \|\mathrm{S}(g)\|_{p'} \\
&\lesssim \|S(f)\|_p \|g\|_{p'} \leq \|S(f)\|_p\,.
\end{aligned}$$

Here, we have used Cauchy-Schwartz to pass to the Square Functions, and then used Hölder's Inequality. Thus, $\|f\|_p \lesssim \|\mathrm{S}(f)\|_p$.

To prove the reverse inequality, we also use duality, but it is expressed in a more complicated way, as we must 'dualize' $\mathrm{S}(f)$. In order to do this, we have to consider $\mathrm{S}(f)$ as an element of $L^p_{\ell^2(\mathcal{D})}$, that is we view

$$\varphi = \Big\{ \frac{\langle f, h_I\rangle}{\sqrt{|I|}}\mathbf{1}_I \,:\, I \in \mathcal{D}\Big\}$$

as a function taking values in $\ell^2(\mathcal{D})$, which is p-integrable. Then, duality is expressed as a function $\gamma = \{\gamma_I(x) \,:\, I \in \mathcal{D}\} \in L^{p'}_{\ell^2(\mathcal{D})}$, which is of norm one and

$$\|\varphi\|_{L^p_{\ell^2}} = \langle \varphi, \gamma\rangle = \sum_{I\in\mathcal{D}} \frac{\langle f, h_I\rangle}{\sqrt{|I|}} \int_I \gamma_I \, dy\,. \tag{1.9.8}$$

It is clear that we can assume that γ_I is supported on I. In general, we need to permit that $\gamma_I(x)$ varies over the interval I. But the pairing above shows that only the mean-value of γ_I over I is important. We will suppose that γ_I is in fact constant on I, and leave it as an exercise to justify this supposition. (See Exercise 1.9.17 and 1.9.18.) Therefore, we can define function

$$g = \sum_{I\in\mathcal{D}} \Big(\gamma_I \sqrt{|I|}\Big)\cdot h_I\,.$$

And with this definition, $\mathrm{S}(g) = \|\gamma\|_{\ell^2(\mathcal{D})}$.

Then, we estimate

$$\begin{aligned}
\|\mathrm{S}(f)\|_p = \|\varphi\|_{L^p_{\ell^2}} &= \langle \varphi, \gamma\rangle \\
&= \sum_{I\in\mathcal{D}} \langle f, h_I\rangle \gamma_I \sqrt{|I|} \\
&= \langle f, g\rangle \\
&\leq \|f\|_p \|g\|_{p'} \\
&\lesssim \|f\|_p \Big\|\mathrm{S}(g)\Big\|_{p'} \\
&\lesssim \|f\|_p\,.
\end{aligned}$$

And so our proof is finished.

□

1.9.1 Exercises

1.9.9 Exercise. The Littlewood Paley inequalities can be proved directly in the case that p is an even integer. For $p = 2j$, j an integer, show that

$$\|f\|_p^p = \frac{p!}{2^j} \sum_{I_1 \subset I_2 \subset \cdots \subset I_j} |I_1| \prod_{k=1}^{j} \frac{\langle f, h_{I_k}\rangle^2}{|I_k|} \tag{1.9.10}$$

$$\|\mathrm{S} f\|_p^p = j! \sum_{I_1 \subset I_2 \subset \cdots \subset I_j} |I_1| \prod_{k=1}^{j} \frac{\langle f, h_{I_k}\rangle^2}{|I_k|} \tag{1.9.11}$$

Note that as $p \to \infty$, this shows that $\|f\|_p \lesssim \sqrt{p}\|\mathrm{S}(f)\|_p$

1.9.12 Exercise. We have the inequality

$$\sup_{\lambda>0} \lambda|\{\mathrm{S} f > \lambda\}| \le 3\|f\|_1 \tag{1.9.13}$$

And by interpolation, this proves

$$\|\mathrm{S} f\|_p \lesssim \|f\|_p\,, \qquad 1 < p < 2\,. \tag{1.9.14}$$

Prove this by using the obvious L^2 inequality and the Calderón Zygmund Decomposition.

1.9.15 Exercise. Consider a Square function with a 'shift' in location, a point that will be formalized in the next chapter. For an integer n, let

$$\mathrm{S}_n f := \Big[\sum_{I\in\mathcal{D}} \frac{\langle f, h_I\rangle^2}{|I|} \mathbf{1}_{I+n|I|}\Big]^{1/2} \tag{1.9.16}$$

So the influence of the Ith Haar coefficient is felt at a proportionally distant from I. Investigate the L^p properties of the this square function by showing that

$$(\mathrm{S}_n f)^{\sharp,1+\epsilon} \lesssim n^{\epsilon}\, \mathrm{M} f\,.$$

1.9.17 Exercise. Let φ be a non-negative Schwarz function with $\int \varphi \; dx = 1$. For $t_j > 0$, show that

$$\Big\| \Big[\sum_j \big|\mathrm{Dil}^1_{t_j} \varphi * f_j\big|^2\Big]^{1/2} \Big\|_p \lesssim \Big\| \Big[\sum_j |f_j|^2\Big]^{1/2} \Big\|_p\,, \qquad 1 < p < \infty\,.$$

(Consider the case of $2 \le p < \infty$ first, and appeal to the Maximal function bound.)

1.9.18 Exercise. Use the previous exercise to justify replacing $\gamma_I(x)$ by its average value on I in (1.9.8).

1.9.19 Exercise. Show that for the Scale Shift operator in (1.4.7) are bounded on all L^p, $1 < p < \infty$. How does the value of the scale shift parameter s affect the L^p norm, and a function of s and p?

1.9.20 Exercise. Show that for the Location Shift operator in (1.4.7) are bounded on all L^p, $1 < p < \infty$. How does the value of the location shift parameter n affect the L^p norm, and a function of n and p?

Chapter 2

Calderón-Zygmund Operators

2.1 Paraproducts

Products, and certain kind of renormalized products are common objects. Let us explain the renormalized products in a very simple situation. We begin with the definition of a *paraproduct*, as a bilinear operator. Define

$$h_I^0 = h_I\,, \qquad h_I^1 = |h_I^0| = \mathrm{Dil}_I^2\,\mathbf{1}_{[-1/2,1/2]}\,. \tag{2.1.1}$$

The superscript ^{0} indicates a mean-zero function, while the superscript ^{1} indicates a non-zero integral. Now define

$$\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(f_1, f_2) := \sum_{I\in\mathcal{D}} \frac{\langle f_1, h_I^{\epsilon_1}\rangle}{\sqrt{|I|}} \langle f_2, h_2^{\epsilon_2}\rangle h_I^{\epsilon_3}\,, \qquad \epsilon_j \in \{0,1\}. \tag{2.1.2}$$

For the most part, we consider cases where there is one choice of ϵ_j which is equal to one, but in considering fractional integrals, one considers examples where there two ϵ_j equal to one.

Why the name paraproduct? This is probably best explained by the identity

$$f_1 \cdot f_2 = \mathrm{P}^{1,0,0}(f_1, f_2) + \mathrm{P}^{0,0,1}(f_1, f_2) + \mathrm{P}^{0,1,0}(f_1, f_2)\,. \tag{2.1.3}$$

Thus, a product of two functions is a sum of three paraproducts. The three individual paraproducts in many respects behave like products, for instance we will see that there is a Hölder Inequality. And, very importantly, in certain instances they are *better* than a product.

To verify (2.1.3), let us first make the self-evident observation that

$$\frac{1}{|J|}\int_J g(y)\,dy = \frac{\langle g, h_I^1\rangle}{\sqrt{|I|}} = \sum_{J\,:\,J\supsetneq I} \langle g, h_J\rangle h_J(I)\,, \tag{2.1.4}$$

where $h_J(I)$ is the (unique) value h_J takes on I. In (2.1.3), expand both f_1 and f_2 in the Haar basis,

$$f_1 \cdot f_2 = \left\{ \sum_{I \in \mathcal{D}} \langle f_1, h_I \rangle h_I \right\} \cdot \left\{ \sum_{J \in \mathcal{D}} \langle f_2, h_J \rangle h_J \right\}.$$

Split the resulting product into three sums, (1) $I = J$, (2) $I \subsetneq J$ (3) $J \subsetneq I$. In the first case,

$$\sum_{I,J \,:\, I=J} \langle f_1, h_I \rangle \langle f_2, h_J \rangle (h_I)^2 = \mathrm{P}^{0,0,1}(f_1, f_2)\,.$$

In the second case, use (2.1.4).

$$\begin{aligned} \sum_{I,J \,:\, I \subsetneq J} \langle f_1, h_I \rangle \langle f_2, h_J \rangle h_I \cdot \frac{1}{|I|} \int_I h_J(y)\, dy &= \sum_I \langle f_1, h_I \rangle \frac{\langle f_2, h_I^1 \rangle}{\sqrt{|I|}} h_I \\ &= \mathrm{P}^{0,1,0}(f_1, f_2)\,. \end{aligned}$$

And the third case is as in the second case, with the role of f_1 and f_2 switched.

A rudimentary property is that Paraproducts should respect Hölder's inequality, a matter that we turn to next.

2.1.5 Theorem. *Suppose at most one of $\epsilon_1, \epsilon_2, \epsilon_3$ are equal to one. We have the inequalities*

$$\|\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(f_1, f_2)\|_q \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2}\,, \qquad 1 < p_1, p_2 < \infty\,,\ 1/q = 1/p_1 + 1/p_2\,. \tag{2.1.6}$$

Proof. As a matter of convenience, we assume that $\epsilon_1 = 1$.

The proof splits into two cases. The case where $q > 1$ permits a natural appeal to duality, and nicely illustrates the main step. Thus, take $p_3 = q/(q-1)$ to be the index dual to q. We then have $1/p_1 + 1/p_2 + 1/p_3 = 1$, and we should bound $|\langle \mathrm{P}^{1,0,0}(f_1, f_2), f_3 \rangle|$. In fact, we will bound

$$I := \sum_{I \in \mathcal{D}} \frac{|\langle f_1, h_I^1 \rangle|}{\sqrt{|I|}} |\langle f_2, h_I \rangle \langle f_3, h_I \rangle| \lesssim \prod_{j=1}^{3} \|f_j\|_{p_j}\,. \tag{2.1.7}$$

Let us write I above as

$$I = \int \sum_{I \in \mathcal{D}} \prod_{j=1}^{3} \frac{|\langle f_j, h_I^{\epsilon_j} \rangle|}{\sqrt{|I|}} \mathbf{1}_I(x)\, dx$$

We have no orthogonality in f_1, but do in f_2 and f_3. Then pointwise in x, we take the supremum of the terms associated with f_1, which is the Maximal Function, and use Cauchy-Schwartz in the other two to get the Square Function. Hence,

$$I \le \int \mathrm{M} f_1 \cdot \mathrm{S}(f_2) \cdot \mathrm{S}(f_3)\, dx$$

$$\le \|\mathrm{M}\, f_1\|_{p_1} \|\mathrm{S}(f_2)\|_{p_2} \|\mathrm{S}(f_3)\|_{p_3}$$
$$\lesssim \|f_1\|_{p_1} \|f_2\|_{p_2} \|f_3\|_{p_3} \,.$$

To pass to the second line, we have used Hölder's inequality, and the third line follows from the bounds we have established for the Maximal and Square Functions. This completes the proof of (2.1.7).

For the proof in the case where we cannot appeal to duality, begin by observing that it suffices to show the weak-type inequality

$$\|\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(f_1, f_2)\|_{q,\infty} \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2} \,, 1 < p_1, p_2 < \infty \,,\ \tfrac{1}{q} = \tfrac{1}{p_1} + \tfrac{1}{p_2} \ge 1 \,. \tag{2.1.8}$$

For then, one can use the inequalities already established, and Marcinceiwcz interpolation to complete the proof.

It is a matter of convenience to note that we need not establish the full range of inequalities, but only a single instance. Namely, it suffices to establish

$$|\{\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(f_1, f_2) > 1\}| \lesssim 1 \,, \qquad \|f_1\|_{p_1} = \|f_2\|_{p_2} = 1 \tag{2.1.9}$$

where the index $q \le 1$. This is so, as we are proving the inequality for a class of operators which are invariant under dilations by factors of 2, thus any single instance of the weak-type inequality proves the full range of inequalities.

Fix f_1, f_2 as in (2.1.9). We will invoke a simplified version of the Calderón Zygmund Decomposition, to reduce the inequality (2.1.9) to those we have already proved. Take

$$E = \{\mathrm{M}\, f_1 > 1\} \cup \{\mathrm{M}\, f_2 > 1\} \,,$$

which satisfies the estimate $|E| \lesssim 1$, where the implied constant depends upon p_1 and p_2. Let $\mathcal{I}$ be the maximal dyadic intervals such that either

$$|I|^{-1} \int_I |f_j| \, dx \ge 1 \,, \qquad j = 1 \text{ or } j = 2.$$

Then set

$$\phi_j(x) = \begin{cases} f_j(x) & x \notin E \\ |I|^{-1} \int_I f_j \, dx & x \in I \,,\ I \in \mathcal{I} \,. \end{cases}$$

We must have $\|\phi_j\|_\infty \le 2$, whence $\|\phi_j\|_{2p_j} \lesssim 1$. Moreover, it is essential to note that if we are *off* the set E, that we have the equality

$$\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(f_1, f_2)(x) = \mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(\phi_1, \phi_2)(x) \,, \qquad x \notin E \,.$$

Now $1/r = \frac{1}{2p_1} + \frac{1}{2p_2} < 1$, so by the first case we can prove (2.1.9) as follows.

$$\begin{aligned} |\{\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(f_1, f_2) > 1\}| &\le |E| + |\{\mathrm{P}^{\epsilon_1,\epsilon_2,\epsilon_3}(\phi_1, \phi_2) > 1\}| \\ &\lesssim 1 + \Big[\|\phi_1\|_{2p_1} \|\phi_2\|_{2p_2}\Big]^r \\ &\lesssim 1 \,. \end{aligned}$$

□

Exercises

2.1.10 Exercise. Prove Theorem 2.1.5 for the paraproduct $\mathrm{P}^{0,0,1}$. (The significant case are those inequalities where duality does not apply.)

2.2 Paraproducts and Carleson Embedding

We have indicated that Paraproducts are better than products in one way. These fundamental inequalities are the subject of this section. Let us define the notion of *(dyadic) Bounded Mean Oscillation*, BMO for short, by

$$\|f\|_{\mathrm{BMO}} = \sup_{J\in\mathcal{D}}\Big[|J|^{-1}\sum_{I\subset J}\langle f,h_I\rangle^2\Big]^{1/2}. \tag{2.2.1}$$

This norm has an expression in terms of the sharp function, $\|f\|_{\mathrm{BMO}} = \|f^{\sharp,2}\|_\infty$, so it is not a complete stranger to us.

2.2.2 Theorem. *Suppose that at exactly one of ϵ_2 and ϵ_3 are equal to* 1*.*

$$\Big\|\mathrm{P}^{0,\epsilon_2,\epsilon_3}(f_1,\cdot)\Big\|_{p\to p} \simeq \|f_1\|_{\mathrm{BMO}}\,, \qquad 1<p<\infty\,. \tag{2.2.3}$$

Indeed, we have

$$\Big\|\mathrm{P}^{0,1,0}(f_1,\cdot)\Big\|_{p\to p} \simeq \sup_J\|\mathrm{P}^{0,1,0}(f_1,|J|^{-1/p}\mathbf{1}_J)\|_p \simeq \|f_1\|_{\mathrm{BMO}}\,. \tag{2.2.4}$$

Here, we are treating the paraproduct as a linear operator on f_2, and showing that the operator norm is characterized by $\|f_1\|_{\mathrm{BMO}}$. Obviously, $\|f\|_{\mathrm{BMO}} \le 2\|f\|_\infty$, and again this a crucial point, there are unbounded functions with bounded mean oscillation, with the canonical example being $\ln x$. Thus, paraproducts are, in a specific sense, better than pointwise products of functions.

Proof. The case $p=2$ is essential, and let us discuss the case of $\mathrm{P}^{0,1,0}$ in detail. Note that the dual of the operator

$$f_2 \longrightarrow \mathrm{P}^{0,1,0}(f_1,f_2)\,,$$

that is we keep f_1 fixed, is the operator $\mathrm{P}^{0,0,1}(f_1,\cdot)$, so it is enough to consider $\mathrm{P}^{0,1,0}$ in the L^2 case.

One direction of the inequalities is as follows.

$$\begin{aligned}\|\mathrm{P}^{0,\epsilon_2,\epsilon_3}(f_1,\cdot)\|_{2\to2} &\ge \sup_J\|\mathrm{P}^{0,\epsilon_2,\epsilon_3}(f_1,h_J^1)\|_p\\ &\ge \|f_1\|_{\mathrm{BMO}}\end{aligned}$$

as is easy to see from inspection. Thus, the BMO lower bound on the operator norm arises solely from testing against normalized indicator sets.

For the reverse inequality, we compare to the Maximal Function. Fix f_1, f_2, and let

$$\mathcal{D}_k = \left\{ I \in \mathcal{D} \,:\, \frac{|\langle f_2, h_I \rangle|}{\sqrt{|I|}} \simeq 2^k \right\}$$

Let $\mathcal{D}_k^*$ be the maximal intervals in $\mathcal{D}_k$. The L^2-bound for the Maximal Function gives us

$$\sum_k 2^{2k} \sum_{I^* \in \mathcal{D}_k^*} |I^*| \lesssim \|\mathrm{M} f_2\|_2^2 \lesssim \|f\|_2^2 \,. \tag{2.2.5}$$

Then, for $I^* \in \mathcal{D}_k^*$ we have

$$\begin{aligned} \Big\| \sum_{\substack{I \in \mathcal{D}_k \\ I \subset I^*}} \langle f_1, h_I \rangle 2^k h_I \Big\|_2^2 &= 2^{2k} \sum_{\substack{I \in \mathcal{D}_k \\ I \subset I^*}} \langle f_1, h_I \rangle^2 \\ &\le 2^{2k} \|f_1\|_{\mathrm{BMO}}^2 |I^*| \end{aligned}$$

And so we are done by (2.2.5).

For the case of $2 < p < \infty$, let us provide a lower bound on the operator norm of the paraproduct by

$$\begin{aligned} \left\| \mathrm{P}^{0,1,0}(f_1, \cdot) \right\|_{p \to p} &\ge \sup_J \left\| \mathrm{P}^{0,1,0}(f_1, |J|^{-1/p} \mathbf{1}_J) \right\|_p \\ &\ge \sup_J |J|^{-1/p} \Big\| \sum_{I \subset J} \langle f_1, h_I \rangle h_I \Big\|_p \\ &\ge \sup_J |J|^{-1/2} \Big\| \sum_{I \subset J} \langle f_1, h_I \rangle h_I \Big\|_2 = \|f_1\|_{\mathrm{BMO}} \,. \end{aligned}$$

The case of $1 < p < 2$ requires a significant additional feature of the BMO norm, the John-Nirenberg inequalities that we will turn to below.

Turning to the upper bound, there are several proofs. For instance, with the L^2 bound established, one can appeal to the Calderón Zygmund Decomposition to conclude the weak-type inequality at L^1. That in turn can be interpolated to the L^p bound, for $1 < p < 2$. Duality then provides the L^p bounds for $2 < p < \infty$. We will leave this proof to the reader.

□

Exercises

2.2.6 Exercise. Show that $\|\log x\|_{BMO} < \infty$.

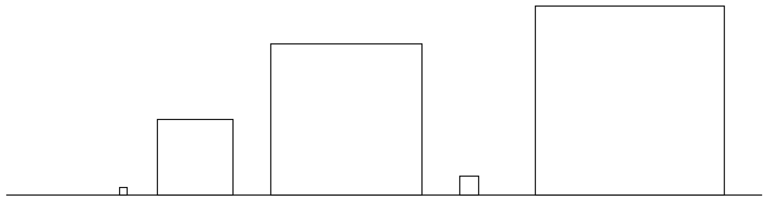


Figure 2.1: Some Carleson Boxes

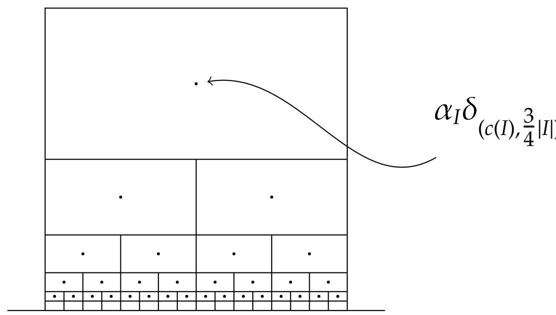


Figure 2.2: The measure μ_α of (2.3.4). The circles are uniformly separated in the hyperbolic metric on the plane.

2.2.7 Exercise. Prove that $\mathrm{P}^{0,0,0}$ maps $BMO \times BMO \longrightarrow BMO$.

2.2.8 Exercise. Use the L^2 bound for the paraproducts, and the Calderón Zygmund Decomposition to show that $\mathrm{P}^{0,\epsilon_2,\epsilon_3}$ maps $BMO \times L^1 \longrightarrow L^{1,\infty}$, provided at most one of ϵ_2, ϵ_3 are equal to one.

2.3 Carleson Measures, John-Nirenberg Inequality

The BMO norm, with its self-refining definition, has remarkable properties, of which the most versatile is the *John-Nirenberg Inequality*. Indeed, some of these properties can be phrased in the absence of function theory, which is a clue to the utility of these notions in a wide variety of settings.

2.3.1 Definition. Let $\alpha = \{\alpha_I : I \in \mathcal{D}\}$ be a sequence of positive numbers. Define

$$\|\alpha\|_{CM} := \sup_J |J|^{-1} \sum_{I \subset J} \alpha(I). \tag{2.3.2}$$

The supremum is taken over all intervals J, not just dyadic intervals. In this definition, 'CM' stands for Carleson measures.

The mention of 'measure' is of course historical, and merits discussion. We work on the upper-half plane $\mathbb{R}^2_+$, with boundary $\mathbb{R}$. Given interval $I \subset \mathbb{R}$, we set

$\mathrm{Box}(I) = I \times (0, |I|) \subset \mathbb{R}^2_+$. See Figure 2.1 Then, by a classical definition, a positive measure μ on $\mathbb{R}^2_+$ is a *Carleson measure* iff

$$\|\mu\|_{CM(\mathbb{R}^2_+)} = \sup_I |I|^{-1} \mu(\mathrm{Box}(I)) < \infty \,. \tag{2.3.3}$$

Notice that μ is in essence a two-dimensional object, controlled in a one-dimensional fashion. To make the connection with the definition above, given $\alpha = \{\alpha_I \,:\, I \in \mathcal{D}\}$, define measure

$$\mu_\alpha = \sum_{I \in \mathcal{D}} \alpha_I \delta_{(c(I)\,,\,2^{-1}|I|)} \,. \tag{2.3.4}$$

Here, δ_z denotes a Dirac point mass at z, and $c(I)$ is the center of I. Thus, for dyadic interval I we have

$$\mu_\alpha(\mathrm{Box}(J)) = \sum_{I \subset J} \alpha_I \,.$$

There is a close connection to that of *BMO*. Namely, we have

$$\|\{\langle f, h_I \rangle^2\}\|_{CM} = \|f\|^2_{BMO} \,. \tag{2.3.5}$$

Thus, the results of this section are closely related to the analysis of *BMO* functions.

The significance of the definition of Carleson Measure is the following version of the Carleson Embedding Theorem.

2.3.6 A Carleson Embedding Theorem. *Let $1 < p < \infty$, and $f \in L^p(\mathbb{R})$. Consider the Poisson extension of f to $\mathbb{R}^2_+$ given by*

$$\mathrm{P}_t f(x) = \int f(y) \frac{t\,dy}{|x-y|^2 + t^2} \,.$$

The operator $f \longrightarrow \mathrm{P}_t f$ is bounded from $L^p(dx)$ to $L^p(\mathbb{R}^2_+, \mu)$ iff μ is a Carleson measure, as defined in (2.3.3). *Namely, we have*

$$\|\mathrm{P}_t\|_{L^p(dx) \to L^p(\mathbb{R}^2_+, \mu)} \simeq \|\mu\|_{CM(\mathbb{R}^2_+)}$$

We will prove a discrete analog of this Theorem below. Define

$$\mathrm{T}(\alpha, f) := \sum_{I \in \mathcal{D}} \alpha_I \frac{\langle f, \mathbf{1}_I \rangle}{|I|} \mathbf{1}_I \,.$$

Continuing our observation (2.3.5), we make a connection between this definition and the Square Function of a Paraproduct operator.

$$\mathrm{S}(\mathrm{P}^{0,1,0}(f_1, f_2))^2 = \sum_I \langle f_1, h_I \rangle^2 \frac{\langle f_2, \mathbf{1}_I \rangle^2}{|I|} \mathbf{1}_I \le \mathrm{T}(\{\langle f_2, \mathbf{1}_I \rangle^2\}, f_2^2) \,.$$

The main inequalities for the operators T are:

2.3.7 Theorem. *For $1 < p < \infty$, we have the estimate*

$$\|\mathrm{T}(\alpha,\cdot)\|_{p\to p} \simeq \|\alpha\|_{CM}$$

Note that applying the operator $\mathrm{T}(\alpha,\cdot)$ to indicator sets, we see that the upper bound implies the inequalities below, which are formally *stronger* than (2.3.2),

$$\sup_J |J|^{-1/p} \Big\| \sum_{I\subset J} \frac{\alpha_I}{|I|} \mathbf{1}_I \Big\|_p \lesssim \|\alpha\|_{CM}\,, \qquad 1 < p < \infty\,.$$

As is typical, these are fundamental inequalities, and we turn to their proof first.

2.3.8 John Nirenberg Inequality. *For any interval J, and $1 < p < \infty$, we have the estimate*

$$\Big\| \sum_{I\subset J} \frac{\alpha(I)}{|I|} \mathbf{1}_I \Big\|_p \lesssim p\|\alpha\|_{CM}\,|J|^{1/p}. \tag{2.3.9}$$

Proof. This is a critical result, and one that admits different proofs. We present here only the simplest proof, and explore other approaches in the exercises.

It suffices to prove the Lemma for integers p. For an interval J, set

$$A_J := \sum_{I\subset J} \frac{\alpha(I)}{|I|} \mathbf{1}_I.$$

Then, we prove by induction that

$$\int |A_J|^p \, dx \le p!\|\alpha\|_{CM}^p |J|^p, \qquad p = 2, 3, \ldots$$

By the trivial estimation $p! \le p^p$, the Lemma follows.

The case $p = 1$ following from the definition of Carleson measure norm. For the inductive case. By the basic dyadic grid property that if two intervals intersect, then one contains the other, we have

$$\begin{aligned} \int |A_J|^{p+1} \, dx &\le (p+1) \sum_{I\subset J} \frac{\alpha(I)}{|I|} \int_I A_I^p \, dx \\ &\le (p+1)!\|\alpha\|_{CM}^p \sum_{I\subset J} \alpha(I) \\ &\le (p+1)!\|\alpha\|_{CM}^{p+1} |J|. \end{aligned}$$

□

The connection between the growth of L^p norms and exponential Orlicz classes is relevant here, recall in particular Appendix A, and Proposition A.0.2. Note that the growth of the L^p-norms shows that we have the inequality below, for some $0 < c$ absolute.

$$\frac{1}{|J|}\int_J \exp\Big(c\sum_{I\subset J}\frac{\alpha_I}{|I|}\mathbf{1}_I\Big)\,dx \lesssim \|\alpha\|_{CM}\,. \tag{2.3.10}$$

Thus, the sums are *exponentially* integrable.

Proof of Theorem 2.3.7. For the lower bound on the operator norm of $\mathrm{T}(\alpha,\cdot)$, note that for any $1 < p < \infty$,

$$\begin{aligned}\|\mathrm{T}(\alpha,\cdot)\|_{p\to p} &\geq \sup_J \|\mathrm{T}(\alpha,|J|^{-1/p}\mathbf{1}_J)\|_p \\ &\geq \sup_J |J|^{-1/p}\Big\|\sum_{I\subset J}\frac{\alpha_I}{|I|}\mathbf{1}_I\Big\|_p \\ &\geq \sup_J |J|^{-1}\Big\|\sum_{I\subset J}\frac{\alpha_I}{|I|}\mathbf{1}_I\Big\|_1 \\ &= \|\alpha\|_{CM}\,,\end{aligned}$$

as the L^p-norm dominates the L^1-norm.

And so we should prove the upper bound on the operator norm of $\mathrm{T}(\alpha,\cdot)$, assuming that $\|\alpha\|_{CM} = 1$. To do this, for $1 \leq p < \infty$, we show that there is a constant K so that for all $\|f\|_p = 1$, we have

$$|\{\mathrm{T}(\alpha,f) \geq 1\}| \leq K\,. \tag{2.3.11}$$

This is a single instance of the weak-type inequality, with dilation invariance supplying the full weak-type inequality, and interpolation the strong-type inequality for $1 < p < \infty$.

We rely upon the John-Nirenberg Inequalities, and the following elementary fact. For positive functions f_j, and constants $\lambda_j > 0$, we have

$$\Big\{\sum_j f_j > \sum_j \lambda_j\Big\} \subset \bigcup_j \{f_j > \lambda_j\}\,.$$

Indeed, if $\sum_j f_j > \sum_j \lambda_j$, then we must have $f_j > \lambda_j$ for at least one j.

Now, set $E = \{\mathrm{M}f > 1\}$, which is a set of measure bounded by a constant. We do not estimate $\mathrm{T}(\alpha,f)$ on this set. For $j > 0$,

$$\mathcal{D}_j = \{I \in \mathcal{D} \,:\, 4^{-j} \leq \frac{\langle f,\mathbf{1}_I\rangle}{|I|} < 4^{-j}\}.$$

We use these collections of intervals to decompose the sum $\mathrm{T}(\alpha, f)$, defining

$$T_j = \sum_{I \in \mathcal{D}_j} \alpha_I \mathbf{1}_I \,.$$

Notice that for $x \notin E$, we have

$$\mathrm{T}(\alpha, f)(x) \le \sum_j 4^{-j} T_j \,.$$

And so we can estimate

$$|\{x \in \mathbb{R} - E \;:\; \mathrm{T}(\alpha, f)(x) > 1\}| \le \sum_{j \ge 0} |\{T_j > 2^{-j-1}\}|. \tag{2.3.12}$$

$$\le \sum_{j \ge 0} 2^{-rj} \|T_j\|_r^r \,. \tag{2.3.13}$$

In the last line, we use Chebyscheff inequality. And we argue that for e. g. $r = 2p$, the last term sums to a constant.

To estimate the norm of T_j, we use a straight forward extension of the estimate in Lemma 2.3.8. Denoting the maximal elements of $\mathcal{D}_j$ by $\mathcal{D}_j^*$, we can estimate

$$\|T_j\|_r^r \le \sum_{I^* \in \mathcal{D}_j^*} \Big\| \sum_{I \,:\, I \subset I^*} \frac{\alpha_I}{|I|} \mathbf{1}_I \Big\|_r^r \tag{2.3.14}$$

$$\lesssim (Cr)^r \sum_{I^* \in \mathcal{D}_j^*} |I^*| \,. \tag{2.3.15}$$

But, the boundedness of the Maximal Function implies that

$$\sum_{j=0}^{\infty} 4^{-pj} \sum_{I^* \in \mathcal{D}_j^*} |I^*| \lesssim 1 \,.$$

Thus, for $r = 2p$, we can combine (2.3.13) and (2.3.15) to complete the proof.

□

There is a further property of the Carleson Measure norm that is another reflection of the John-Nirenberg inequalities. To state it, let us define an apparently weaker Carleson Measure norm by

$$\|\alpha\|_{CM,0} := \inf\Big\{ \Lambda > 0 \;:\; \sup_J |J|^{-1} \Big| \Big\{ x \;:\; \sum_{I \subset J} \frac{\alpha_I}{|I|} \mathbf{1}_I > \Lambda \Big\} \Big| < \tfrac{1}{2} \Big\} . \tag{2.3.16}$$

With this definition, we are replacing an implicit L^1 norm in (2.3.2) in an 'L^0' norm.

2.3.17 Proposition. *We have the inequality*

$$\|\alpha\|_{CM} \lesssim \|\alpha\|_{CM,0} \,. \tag{2.3.18}$$

Proof. Fix a sequence $\alpha = \{\alpha_I \,:\, I \in \mathcal{D}\}$ with $\|\alpha\|_{CM,0} = 1$. Consider the function

$$G(\lambda) = \sup_J |J|^{-1} \Big| \Big\{ x \in J \,:\, \sum_{I \subset J} \frac{\alpha_I}{|I|} \mathbf{1}_I > \lambda \Big\} \Big| \,.$$

We will argue that

$$G(\lambda + 2) \le \tfrac{1}{2} G(\lambda) \,, \qquad \lambda \ge 1 \,. \tag{2.3.19}$$

It follows that we have $G(\lambda) \lesssim 2^{-\lambda/2}$, which inequality should be compared with (2.3.10). That is, this proof will provide another proof of the John-Nirenberg inequalities.

Fix $\lambda \ge 1$, and an interval J so that for some $x \in J$ we have

$$\sum_{I \subset J} \frac{\alpha_I}{|I|} \mathbf{1}_I(x) > \lambda + 2 \,.$$

Let $\mathcal{K}^*$ be the maximal dyadic intervals in $K \subset J$ for which we have

$$\lambda < \sum_{K \subset I \subset J} \frac{\alpha_I}{|I|} \,.$$

It follows from the definition of $\|\alpha\|_{CM,0} = 1$ that we then necessarily have

$$\lambda \le \sum_{K \subset I \subset J} \frac{\alpha_I}{|I|} \le \lambda + 1 \,.$$

Hence, our estimate is

$$\begin{aligned} |J|^{-1} \Big| \Big\{ x \in J \,:\, \sum_{I \subset J} \frac{\alpha_I}{|I|} \mathbf{1}_I > \lambda + 2 \Big\} \Big| &\le |J|^{-1} \sum_{K^* \in \mathcal{K}} \Big| \Big\{ x \in J \,:\, \sum_{I \subset K^*} \frac{\alpha_I}{|I|} \mathbf{1}_I > 1 \Big\} \Big| \\ &\le \tfrac{1}{2} G(\lambda) \,. \end{aligned}$$

So our proof is complete. □

2.3.1 Exercises

2.3.20 Exercise. We explore other proofs of the John Nirenberg inequality. At the first juncture, let $\|\alpha\|_{CM} = 1$, and for a set U set

$$F_U := \sum_{I \subset U} \frac{\alpha(I)}{|I|} \mathbf{1}_I$$

Consider this condition: For a fixed $1 < p < \infty$, and every U there is a set $V \subset U$ with $|V| \le \frac{1}{2}|U|$ for which

$$\|F_U\|_p \lesssim |U| + \|F_V\|_p \tag{2.3.21}$$

Show that this condition is equivalent to (2.3.9).

2.3.22 Exercise. Prove (2.3.21) by duality. Thus, fix $g \in L^{p'}$ of norm one, where p' is conjugate to p, for which $\|F_U\|_p = \langle F_U, g\rangle$. Then, for an appropriate constant K, set $V := \{Mg > K|U|^{-1/p'}\}$. Conclude (2.3.21).

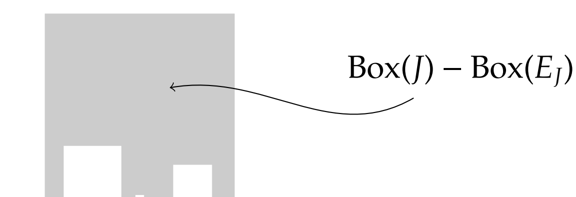
2.3.23 Exercise. Extend the definition of Box(I) in (2.3.3) in the following way. Define

$$\mathrm{Box}\Big(\bigcup_{I\in\mathcal{I}} I\Big) = \bigcup_{I\in\mathcal{I}} \mathrm{Box}(I)$$

where $\mathcal{I}$ is a collection of disjoint intervals. Suppose that μ is a positive measure on $\mathbb{R}^2_+$ which satisfies this property: For each interval J there is a set $E_J \subset J$ which is a union of disjoint intervals, so that

$$\mu(\mathrm{Box}(J) - \mathrm{Box}(E_J)) \le |J|\,.$$

The set $\mathrm{Box}(J) - \mathrm{Box}(E_J)$ is shown in gray below. Show that $\|\mu\|_{CM(\mathbb{R}^2_+)} < \infty$.



2.3.24 Exercise. For a function $f \in L^1(\mathbb{R})$, set $E = \{\mathrm{M} f > 1\}$, and set

$$\phi = \sum_{I\,:\,I\not\subset E} \langle f, h_I\rangle h_I\,.$$

That is, we only reconstruct f with those Haar coefficients not associated with a large value of f. Show that $\|\phi\|_{\mathrm{BMO}} \lesssim 1$. Use the boundedness of the Haar Square Function from L^1 to weak-L^1, see (1.9.13), and Proposition 2.3.17.

2.4 John-Nirenberg Inequality for *BMO* Functions

Having detailed the John-Nirenberg Inequality for Carleson Measures, and we will use these estimates to prove the analogous estimates for *BMO* functions in this section. There are two components. The first is a sharp estimate on the distribution of *BMO* functions.

2.4.1 Theorem: John Nirenberg Inequality for *BMO* **Functions, First Version.** *We have this distributional estimate for $f \in BMO$. For any interval I,*

$$\Big\| f - |I|^{-1} \int_I f(y)\, dy \Big\|_{\exp(L)} \lesssim \|f\|_{BMO} .$$

Proof. Using Exercise 1.9.9, which we will reprove in the § 2.5, we can estimate

$$\begin{aligned} \Big\| f - |I|^{-1} \int_I f(y)\, dy \Big\|_{L^p(I)} &= \Big\| \sum_{J \subset I} \langle f, h_J \rangle h_J \Big\|_p \\ &\lesssim \sqrt{p} \Big\| \Big[\sum_{J \subset I} \frac{\langle f, h_J \rangle^2}{|J|} \mathbf{1}_J \Big]^{1/2} \Big\|_p \\ &\lesssim p \|f\|_{BMO} |J|^{1/p} . \end{aligned}$$

The first inequality follows from Exercise 1.9.9, and the second Lemma 2.3.8, applied, keeping in mind that a square intercedes in passing from a *BMO* function, see (2.3.5).

□

The second version of the John-Nirenberg estimates is that we can make a weaker form of the norm.

2.4.2 Theorem: John Nirenberg Inequality for *BMO* **Functions, Second Version.** *We have this estimate*

$$\|f\|_{BMO} \lesssim \sup_J \frac{1}{|J|} \int_J \Big| f - \frac{1}{|J|} \int_J f(y)\, dy \Big|\, dx . \tag{2.4.3}$$

Recall that in (2.2.1)*, we have defined the BMO-norm in terms of square-integrability.*

Proof. Assume the right-hand side of (2.4.3) is equal to one. By the weak-L^1 bound for the Square Function, see (1.9.13), we have

$$\sup_J |J|^{-1} \Big\| \Big[\sum_{I \subset J} \frac{\langle f, h_I \rangle^2}{|I|} \mathbf{1}_I \Big]^{1/2} \Big\|_{1,\infty} \lesssim 1 .$$

By the definition in (2.3.16), we see that

$$\|\{\langle f, h_I \rangle^2\}\|_{CM,0} \lesssim 1 .$$

Therefore, by Proposition 2.3.17, we have $\|\{\langle f, h_I \rangle^2\}\|_{CM} \lesssim 1$, and this completes our proof. □

Exercises

2.4.4 Exercise. Use $\|\log x\|_{BMO} \lesssim 1$ to show that Theorem 2.4.1 is sharp.

2.5 Chang-Wilson-Wolff Inequality

A function f is in *BMO* iff its Square Function is in *BMO*. What if one knows more about the square function, can more be said of the function? The Chang-Wilson-Wolff inequality addresses this in the instance when the Square Function is actually a bounded function.

2.5.1 Theorem: Chang-Wilson-Wolff Inequality. *We have the estimate below.*

$$\|f\|_{\exp(L^2)} \lesssim \|\mathrm{S}(f)\|_\infty \,. \tag{2.5.2}$$

Proof. We give the proof of Chang Wilson and Wolff, in the real valued case, which they learned from Herman Rubin. Indeed, this proof can be regarded as the conditional version of a proof of the Khintchine inequalities given in Appendix B.

Let us recall that a sequence of functions $g_1, \dots,$ form a *martingale* iff for all sequences

$$\mathbb{E}(g_{n+1} \,|\, g_1, \dots, g_n) = g_n \,. \tag{2.5.3}$$

Here, we are taking the conditional expectation of g_{n+1} with respect to the sigma field generated by $g_1, \dots, g_n$.

Let $\mathcal{F}_n$ be the sigma field generated by the dyadic intervals of length 2^{-n}, so that

$$f_n := \mathbb{E}(f \,|\mathcal{F}_n) = \mathbb{E}f + \sum_{|I| \ge 2^{-n}} \frac{\langle f, h_I\rangle}{|I|} h_I$$

is a dyadic martingale. We assume that $\mathbb{E}f = 0$.

For $t > 0$ we define a new martingale by the formula

$$q_n := \mathrm{e}^{tf_n} \Big[\prod_{j=1}^{n-1} \mathbb{E}(\mathrm{e}^{t(f_{j+1}-f_j)} \,|\mathcal{F}_j)\Big]^{-1} .$$

Of course, it is hardly obvious that q_n is a martingale, and so we check this now. Clearly, q_n is $\mathcal{F}_n$ measurable. We should then check that $\mathbb{E}(q_{n+1} \,|\, \mathcal{F}_n) = q_n$.

$$\begin{aligned}
\mathbb{E}(q_{n+1} \,|\, \mathcal{F}_n) &= \mathbb{E}\Big(\mathrm{e}^{tf_{n+1}} \Big[\prod_{j=1}^{n} \mathbb{E}(\mathrm{e}^{t(f_{j+1}-f_j)} \,|\mathcal{F}_j)\Big]^{-1} \Big|\, \mathcal{F}_n\Big) \\
&= \mathbb{E}(\mathrm{e}^{tf_{n+1}} \,|\mathcal{F}_n) \cdot \Big[\prod_{j=1}^{n} \mathbb{E}(\mathrm{e}^{t(f_{j+1}-f_j)} \,|\mathcal{F}_j)\Big]^{-1}
\end{aligned}$$

$$= \mathbb{E}(\mathrm{e}^{t(f_{n+1}-f_n)} \mid \mathcal{F}_n) \cdot \mathrm{e}^{tf_n} \cdot \Big[\prod_{j=1}^{n} \mathbb{E}(\mathrm{e}^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\Big]^{-1}$$

$$= \mathrm{e}^{tf_n} \cdot \Big[\prod_{j=1}^{n-1} \mathbb{E}(\mathrm{e}^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j)\Big]^{-1} = q_n \,.$$

And therefore, $\mathbb{E}q_n = 1$ for all n.

The fact that we work with a dyadic martingale enters. For we can appeal to (B.0.3) to see that

$$\prod_{j=1}^{n-1} \mathbb{E}(\mathrm{e}^{t(f_{j+1}-f_j)} \mid \mathcal{F}_j) \le \prod_{j=1}^{n-1} \mathbb{E}(\mathrm{e}^{t^2(f_{j+1}-f_j)^2} \mid \mathcal{F}_j) = \prod_{j=1}^{n-1} \mathrm{e}^{t^2(f_{j+1}-f_j)^2} = \mathrm{e}^{t^2\,\mathrm{S}(f)^2} \,.$$

Therefore, under the assumption that $\|\mathrm{S}(f)\|_\infty \le 1$, we see that

$$\mathbb{E}\,\mathrm{e}^{tf_n - t^2} \le \mathbb{E}q_n = 1\,.$$

As this holds for all n, we can take $n \to \infty$. Therefore, we have for $\lambda > 0$,

$$\mathbb{P}(f > \lambda) \le \mathrm{e}^{-t\lambda}\, \mathbb{E}e^{tf} \le \mathrm{e}^{-t\lambda + t^2} \,.$$

Taking $t = \lambda/2$ proves the Chang Wilson Wolff inequality in the case that f is real valued. □

Let us use this result to prove an inequality for the Square Function that we have already discussed in Exercise 1.9.9.

2.5.4 Theorem. *We have the estimate*

$$\|\mathrm{M} f\|_p \lesssim (1 + \sqrt{p})\|\mathrm{S}(f)\|_p\,, \qquad 1 < p < \infty\,. \tag{2.5.5}$$

The first step is to derive a good-λ inequality, as in Lemma 1.8.3. This time, we will use it to derive a very efficient estimate involving the Square Function.

2.5.6 Proposition. *For $\lambda > 0$ we have the inequality*

$$\mathbb{P}(\mathrm{M} f > 2\lambda\,;\, \mathrm{S}(f) < \epsilon\lambda) \lesssim \mathrm{e}^{-c\epsilon^{-2}}\, \mathbb{P}(\mathrm{M} f > \lambda)\,, \qquad 0 < \epsilon < \tfrac{1}{2}\,. \tag{2.5.7}$$

Here $\mathrm{M} f$ is the dyadic maximal function, and $0 < c < 1$ is an absolute constant. The point of the estimate is that it holds for all $0 < \epsilon < \frac{1}{2}$, with the constant on the right tending to zero as $\epsilon \downarrow 0$.

Proof. Define a stopping time by

$$\tau = \min\Big\{n \,:\, \sum_{j=1}^{n} (f_j - f_{j-1})^2 \ge \epsilon\lambda\Big\}\,.$$

As is usual, the minimum of the empty set will be taken to be $+\infty$.

Let $f_I = \mathbb{P}(I)^{-1}\mathbb{E}f\mathbf{1}_I$ be the average value of f on I.

Let $\mathcal{Q}$ be the maximal dyadic intervals with $f_I \geq \lambda\mathbb{P}(I)$, so that

$$\{\mathrm{M}\, f > \lambda\} = \bigcup_{I \in \mathcal{Q}} I\,.$$

On each I, define $E_I := I \cap \{\mathrm{M}\, f > 2\lambda\,;\, \mathrm{S}(f) < \epsilon\lambda\}$. This is the main point: If E_I is non-empty then $\mathbb{E}f\mathbf{1}_I \leq (1+\epsilon)\lambda\mathbb{P}(I)$. Indeed, let I' denote the dyadic interval which contains it and is twice as long. So the average value of f on I' is less than λ. If our claim is not true, then

$$|\langle f, h_{I'}\rangle| \geq \epsilon\lambda\mathbb{P}(I)\,,$$

contradicting E_I being non-empty.

Now observe that

$$\mathbb{P}(E_I) = \mathbb{P}(\mathrm{M}\, f > 2\lambda\,;\, \tau = \infty) \leq \mathbb{P}(\mathrm{M}(f_\tau - f_I) > (1-\epsilon)\lambda)\,.$$

Moreover, $\|\mathrm{S}(f_\tau)\|_\infty \leq \epsilon\lambda$. Therefore, by the Chang Wilson Wolff inequality applied to the renormalized martingale $f_\tau - f_I$,

$$\mathbb{P}(E_I) \lesssim \mathrm{e}^{-c\epsilon^{-2}}\,\mathbb{P}(I)\,.$$

By summing over $I \in \mathcal{Q}$ we complete the proof. □

Proof of Theorem 2.5.4. There is a standard way to pass from the good-λ Inequalities to norm inequalities, which we have already seen in the proof of Theorem 1.8.2. The difference in this setting is that we have a much sharper version of the good-λ inequality to work with. Since $|f| \leq \mathrm{M}\, f$, it suffices to prove the estimate $\|\mathrm{M}\, f\|_p \lesssim B_p\|\mathrm{S}(f)\|_p$. First observe that

$$\begin{aligned}\mathbb{P}(\mathrm{M}\, f > 2\lambda) &\leq \mathbb{P}(\mathrm{S}(f) \leq \epsilon\lambda) + \mathbb{P}(\mathrm{M}\, f > 2\lambda\,;\, \mathrm{S}(f) < \epsilon\lambda)\\ &\leq \mathbb{P}(\mathrm{S}(f) \leq \epsilon\lambda) + C\,\mathrm{e}^{-c\epsilon^{-2}}\,\mathbb{P}(\mathrm{M}\, f > \lambda)\,.\end{aligned}$$

Then, we can compute

$$\begin{aligned}\|\mathrm{M}\, f\|_p^p &= p2^p \int_0^\infty \lambda^{p-1}\mathbb{P}(\mathrm{M}\, f > 2\lambda)\, d\lambda\\ &\leq p2^p \int_0^\infty \mathbb{P}(\mathrm{S}(f) \leq \epsilon\lambda)\, d\lambda + p2^p C\,\mathrm{e}^{-c\epsilon^{-2}} \int_0^\infty \lambda^{p-1}\mathbb{P}(\mathrm{M}\, f > \lambda)\, d\lambda\\ &\leq (2/\epsilon)^p\|\mathrm{S}(f)\|_p^p + p2^p C\,\mathrm{e}^{-c\epsilon^{-2}}\|\mathrm{M}\, f\|_p^p\,.\end{aligned}$$

Observe that if we take $\epsilon \simeq p^{-1/2}$, we can conclude

$$\|\mathrm{M}\, f\|_p^p \lesssim (C\sqrt{p})^p\|\mathrm{S}(f)\|_p^p$$

which proves the desired inequality. □

2.6 Commutator Bound

We would like to explain a classical result on commutators.

2.6.1 Theorem. *For a function b, and $1 < p < \infty$ we have the equivalence*

$$\|[b, \mathrm{H}]\|_{p\to p} \simeq \|b\|_{\mathrm{BMO}}\,,$$

where this is the non-dyadic BMO *given by*

$$\sup_{I \text{ interval}} \left[|I|^{-1} \int_I \left| f - |I|^{-1} \int_I f(y)\, dy \right| dx \right]^{1/2}.$$

We refer to this as a classical result, as it can be derived from the Nehari theorem, as we will explain below. Indeed, the upper bound is clear, and the lower bound on the operator norm is found by applying the commutator to normalized indicators of integrals, and we suppress the proof.

But, in many circumstances, different proofs admit different modifications, and so we present a 'real-variable' proof, deriving the upper bound from the Haar shift, and the Paraproduct bound in a transparent way.

Replacing the Hilbert transform by the Haar Shift as defined in (1.4.3), we prove

$$\|[b, \mathfrak{H}]\|_{p\to p} \lesssim \|b\|_{\mathrm{BMO}} \tag{2.6.2}$$

The last norm is dyadic-BMO, which is strictly smaller than non-dyadic BMO. But Proposition 1.4.4 requires that we use all translates and dilates to recover the Hilbert transform, and so the non-dyadic BMO norm will be invariant under these translations and dilations.

The Proposition is that $[b, \mathfrak{H}]$ can be explicitly computed as a sum of Paraproducts which are bounded.

2.6.3 Proposition. *We have*

$$[b, \mathfrak{H}]f = \mathrm{P}^{0,1,0}(b, \mathfrak{H}f) - \mathfrak{H} \circ \mathrm{P}^{0,1,0}(b, f) \tag{2.6.4}$$

$$+ \mathrm{P}^{0,0,1}(b, \mathfrak{H}f) - \mathfrak{H} \circ \mathrm{P}^{0,0,1}(b, f) \tag{2.6.5}$$

$$+ \widetilde{\mathrm{P}}^{0,0,0}(b, f)\,. \tag{2.6.6}$$

In the last line, $\widetilde{\mathrm{P}}^{0,0,0}(b, f)$ is defined to be

$$\widetilde{\mathrm{P}}^{0,0,0}(b, f) = \sum_{I \in \mathcal{D}} \frac{\langle b, h_I^0 \rangle}{\sqrt{I}} \langle f, h_I^0 \rangle (h^0_{I_{\mathrm{left}}} + h^0_{I_{\mathrm{right}}})\,.$$

Each of the five terms on the right are L^p-bounded operators on f, provided $b \in \mathrm{BMO}$, so that the upper bound on the commutator norm in Theorem 2.6.1 follows as an easy corollary. The paraproduct in (2.6.6) does not hew to our narrow definition of a Paraproduct, but it is degenerate in that it is of signature $(0, 0, 0)$, and thus even easier to control than the other terms.

Proof. Now, $[b, \mathfrak{H}]f = b\mathfrak{H}f - \mathfrak{H}(b \cdot f)$. Apply (2.1.3) to both of these products. We see that

$$[b, \mathfrak{H}]f = \sum_{\vec{\epsilon}=(1,0,0),(0,1,0),(0,0,1)} \mathrm{P}^{\vec{\epsilon}}(b, \mathfrak{H}f) - \mathfrak{H}\,\mathrm{P}^{\vec{\epsilon}}(b, f)\,.$$

The choices of $\vec{\epsilon} = (0, 1, 0), (0, 0, 1)$ lead to the first four terms on the right in (2.6.4).

The terms that require more care are the difference of the two terms in which a 1 falls on a b. In fact, we will have

$$\mathrm{P}^{(1,0,0)}(b, \mathfrak{H}f) - \mathfrak{H}\,\mathrm{P}^{(1,0,0)}(b, f) = \widetilde{\mathrm{P}}^{0,0,0}(b, f)\,.$$

To analyze this difference quickly, let us write

$$\langle \mathfrak{H}f, h_I\rangle = \operatorname{sgn}(I)\langle f, h_{\mathrm{Par}(I)}\rangle$$

where $\mathrm{Par}(I)$ is the 'parent' of I, and $\operatorname{sgn}(I) = 1$ if I is the left-half of $\mathrm{Par}(I)$, and is otherwise -1. This definition follows immediately from the definition of g_I in (1.4.1). Now observe that

$$\begin{aligned}\langle \mathrm{P}^{\vec{\epsilon}}(b, \mathfrak{H}f), h_I^0\rangle &= \langle \mathfrak{H}f, \mathrm{P}^{\vec{\epsilon}}(b, h_I^0)\rangle \\ &= \frac{\langle b, h_I^1\rangle}{\sqrt{|I|}} \cdot \langle \mathfrak{H}f, h_I^0\rangle \\ &= \langle f, h^0_{\mathrm{Par}(I),}\rangle \operatorname{sgn}(I) \frac{\langle b, h_I^1\rangle}{\sqrt{|I|}}\end{aligned}$$

And on the other hand, we have

$$\langle \mathfrak{H}P^{1,0,0}(b, f), h_I\rangle = \frac{\langle b, h^1_{\mathrm{Par}(I)}\rangle}{\sqrt{|\mathrm{Par}(I)|}} \operatorname{sgn}(I)\langle f, h^0_{\mathrm{Par}(I)}\rangle$$

Comparing these two terms, we see that we should examine the term that falls on b. But a calculation shows that

$$\sqrt{2}h_I^1 - h^1_{\mathrm{Par}(I)} = -\operatorname{sgn}(I)h^0_{\mathrm{Par}(I)}.$$

Thus, we see that this difference is of the claimed form.

□

Exercises

2.6.7 Exercise. Consider the discrete model J_α of the fractional integral operator defined in (1.6.4). Consider the commutator

$$[b, \mathrm{J}_\alpha]\,, \qquad 0 < \alpha < 1\,.$$

Find an analog of Proposition 2.6.3 for this commutator. Conclude that

$$\|[b, \mathrm{J}_\alpha]\|_{p\to q} \simeq \|b\|_{BMO}\,,$$

where $1 < p, q < \infty$ and $1p = 1/q + \alpha$, and the *BMO* norm is dyadic *BMO*.

2.6.8 Exercise. Prove this result of Chanillo. For $0 < \alpha < 1$, let I_α be the fractional integral operator defined by (1.6.1). Show that

$$\|[b, \mathrm{I}_\alpha]\|_{p\to q} \simeq \|b\|_{BMO}\,,$$

where $1 < p, q < \infty$ are as in Theorem 1.6.2.

2.7 Dyadic H^1 and *BMO*.

We have already defined the space *BMO*, in its dyadic and non-dyadic versions. We turn to the *predual* of *BMO*, which is the Hardy space H^1, in its dyadic version. Of course it is the case that each bounded function is in *BMO*. What is essential is that certain unbounded functions are in *BMO*. And this has the consequence that the the Hardy space H^1 is a proper subset of L^1.

Let us begin with a dyadic *atom.* Say that a is an H^1 *atom* if it is mean zero, supported on an dyadic interval I, and satisfies

$$\int_J |a|^2\, dx \le |J|^{-1}.$$

This last condition is equivalent to $\sum_{I\subset J} |\langle a, h_I\rangle|^2 \le |J|^{-1}$.

Note that a acts on *BMO* as a linear functional, by the duality

$$\langle a, b\rangle = \int_J ab\, dx = \sum_{I\subset J} \langle a, h_I\rangle\langle b, h_I\rangle$$

What is essential here is that on the interval J, a has mean zero, so we can subtract off the mean of b. Then, the pairing of a and b reduces to the usual L^2 pairing.

Conversely, for all $b \in BMO$, we can choose an atom a for which $\langle a, b\rangle = \|b\|_{BMO}$ This is done by taking an interval J for which the right hand side of (2.2.1) is close to the supremum. Then we can take an atom a supported on J, for which the pairing $\langle a, b\rangle$ is equal to the $L^2(J)$ norm of $b - |J|^{-1}\int_J b\, dy$.

But the set of atoms does not form a linear space. To remedy this situation we define

$$\|f\|_{H^1} := \inf\Big\{\sum_j |c_j| \;:\; f = \sum_j c_j a_j,\ \ a_j \text{ is an atom}\Big\} \tag{2.7.1}$$

It is then the case that the dual of H^1 is *BMO*.

2.7.2 Theorem. *We have*

$$(H^1)^* = BMO\,.$$

The Hardy space is a distinguished subspace of L^1, one on which a variety of operators are bounded, whereas they are not bounded on L^1.

2.7.3 Theorem. *We have these inequalities. First, H^1 embeds into L^1, namely*

$$\|f\|_{H^1} \le \|f\|_1$$

Moreover, we have the following equivalences

$$\|f\|_{H^1} \simeq \|\mathrm{S}\, f\|_1 \simeq \|\mathrm{M}\, f\|_1\,. \tag{2.7.4}$$

It is to be stressed that the maximal function does not have absolute values inside the intergral , namely

$$\mathrm{M}\, f(x) = \sup_{I\in\mathcal{D}} \mathbf{1}_I(x)\Big|\frac{1}{|I|}\int_I f(y)\;dy\Big|\,.$$

A critical part of this duality statement is that H^1 is a subset of L^1 which has a separable dual. In addition, it is the case that a range of operators admit natural mapping properties on these spaces, whereas they do not on L^∞, nor L^1. We carry out half of the proof of this Theorem in this section, and leave the second half of the proof to the following section.

Proof of Upper Bounds on Square Function and Maximal Function. Let us first see that $\|\mathrm{S}\, f\|_1 \lesssim \|f\|_{H^1}$. The principle property of the Square Function we use is that if a is an atom supported on dyadic interval I, then $\mathrm{S}\, a$ is also supported on I, a property that follows immediately from the fact that I is dyadic and a has mean zero. It follows that

$$\|\mathrm{S}\, a\|_1 \le |I|^{1/2}\|\mathrm{S}\, a\|_2 \le 1\,.$$

Consider $f \in H^1$ of norm one. Thus, there are atoms a_j supported on dyadic intervals I_j, for which

$$f = \sum_j c_j\, a_j\,, \qquad \sum_j |c_j| = 1\,.$$

and $\sum_j |c_j| \le 2\|f\|_{H^1}$, say. Then, by the triangle inequality

$$\begin{aligned}\|\mathrm{S}\, f\|_1 &\le \sum_j |c_j|\|\mathrm{S}\, a_j\|_1 \\ &\sum_j |c_j| \le 2\|f\|_{H^1}\,.\end{aligned}$$

The proof that $\|\mathrm{M}\, f\|_{H^1} \lesssim \|f\|_{H^1}$ is based upon the observation that for an atom a supported on dyadic interval I, then $\mathrm{M}\, a(x)$ is zero for $x \notin I$. Details are omitted. □

2.7.1 Hardy Space Form of the Calderón Zygmund Decomposition

The proof of the inequality $\|f\|_{H^1} \lesssim \|\mathrm{S}\, f\|_1$ is essentially contained in the following Lemma.

2.7.5 Lemma. *Let f be a finite linear combination of Haar functions. Then, we can write $f = \sum_k f_k$, where each f_k is also a finite linear combination of Haar functions with*

$$|\mathrm{supp}(f_k)| \lesssim |\{\mathrm{S}\, f > 2^k\}|\,, \tag{2.7.6}$$

$$\|f_k\|_2 \lesssim 2^k |\{\mathrm{S}\, f > 2^k\}|^{1/2}\,. \tag{2.7.7}$$

In particular,

$$\|f_k\|_{H^1} \lesssim 2^k |\{\mathrm{S}\, f > 2^k\}| \tag{2.7.8}$$

If $\|\mathrm{S}\, f\|_1 < \infty$, it follows that $f \in H^1$.

Proof. Let $\Omega_k = \{\mathrm{S}\, f > 2^k\}$. Take $\mathcal{D}_k$ to be those dyadic intervals I for which $\langle f, h_I\rangle \neq 0$, and k is the largest integer such that $|I \cap \widetilde{\Omega}_k| \geq \frac{1}{2}$. We then take

$$f_k = \sum_{I \in \mathcal{D}_k} \langle f, h_I\rangle h_I\,.$$

Observe that

$$\mathrm{supp}(f_k) \subset \widetilde{\Omega}_k := \{\mathrm{M}^{\mathcal{D}}\, \mathbf{1}_{\Omega_k} > \tfrac{1}{2}\}\,,$$

so by the weak type inequality for the maximal function, (2.7.6) holds. But also, to estimate the L^2 norm of the f_k, let us first observe that for $I \in \mathcal{D}_k$, we necessarily have

$$|I \cap (\widetilde{\Omega}_k - \Omega_{k+1})| \geq \tfrac{1}{2}|I|.$$

We can use this to estimate the L^2 norm of f_k as follows.

$$\begin{aligned}
\|f_k\|_2^2 &= \sum_{I \in \mathcal{D}_k} \langle f, h_I\rangle^2 \\
&= \int \sum_{I \in \mathcal{D}_k} \frac{\langle f, h_I\rangle^2}{|I|} \mathbf{1}_I \, dx \\
&\leq 2 \int \sum_{I \in \mathcal{D}_k} \frac{\langle f, h_I\rangle^2}{|I|} \mathbf{1}_{I \cap \widetilde{\Omega}_k - \Omega_{k+1}} \, dx \\
&\lesssim 2^{2k} |\Omega_k|\,.
\end{aligned}$$

Our proof is complete. □

The proof that the maximal function can also be used to define the H^1 norm is contained in the following Lemma, whose proof is left as an exercise.

2.7.9 Lemma. *Let f be a finite linear combination of Haar functions. Then, $f = \sum_k f_k$ where f_k is also a finite linear combination of Haar functions with*

$$|\mathrm{supp}(f_k)| \lesssim |\{\mathrm{M} f > 2^k\}| \,, \tag{2.7.10}$$

$$\|f_k\|_2 \lesssim 2^k |\{\mathrm{M} f > 2^k\}|^{1/2} \,. \tag{2.7.11}$$

Finally, we can write this version of the Calderón Zygmund Decomposition.

2.7.12 The Calderón Zygmund Decomposition using Hardy Space. *Let $1 < p < 2$, $f \in L^p$ of norm one and $0 < \alpha < \infty$. Then we can write $f = g + b$ where*

$$\|g\|_2 \lesssim \alpha^{1-p/2} \,,$$
$$\|b\|_{H^1} \lesssim \alpha^{1-p} \,.$$

Proof. We can assume that $\alpha = 2^a$ for some integer a. We apply Lemma 2.7.5, and using this notation define g by

$$g = \sum_{k \le a} f_k \,.$$

This also defines b. The Lemma then follows from the assumption that $f\ inL^p$ of norm one, and that therefore $\|\mathrm{S} f\|_p \lesssim 1$. Namely, using (2.7.8) we see that

$$\|b\|_{H^1} \lesssim \sum_{k \ge a} 2^k |\{\mathrm{S} f > 2^k\}|$$
$$\lesssim \sum_{k \ge 1} 2^{(1-p)k} \lesssim \alpha^{1-p} \,.$$

□

2.7.2 *BMO* and the Boundedness of Operators

2.7.13 Theorem. *Suppose that T is a linear operator which is bounded on L^2, and in addition is bounded as a map from L^∞ to BMO. Then, T extends to a bounded linear operator from L^p to itself for all $2 < p < \infty$. If T is bounded from H^1 to L^1, then it extends to a bounded linear operator on L^p for $1 < p < 2$.*

The proof of the Theorem is easily available. Consider the sub-linear map $f \longrightarrow (Tf)^{\sharp,2}$. By hypothesis, this is bounded as a map from L^2 to weak L^2. Again by assumption, it is bounded as a map from L^∞ to itself:

$$\|(Tf)^{\sharp,2}\|_\infty = \|Tf\|_{BMO} \lesssim \|f\|_\infty$$

Hence, by the usual Marcinciewcz interpolation, we have $\|(Tf)^{\sharp,2}\|_p \lesssim \|f\|_p$ for the range $2 < p < \infty$. But, then for the same range of p, using Theorem 1.8.2,

$$\|Tf\|_p \simeq \|(Tf)^{\sharp,2}\|_p \lesssim \|f\|_p$$

The last conclusion of the Theorem follows by duality. Assuming that T is bounded from H^1 to L^1, it follows by duality that the adjoint T^* maps L^∞ to BMO. The adjoint also maps L^2 to itself, and so T^* maps L^p to itself for $2 < p < \infty$. This is the desired conclusion for T.

We remark that the Theorem remains true, with obvious changes, if we assume that T is bounded from L^p to itself for any $1 < p < \infty$. But, the case we have stated is by far the most important one.

2.8 The $T1$ Theorem

The T 1 Theorem is a profound fact about Calderón Zygmund Operators, providing a set of testing conditions which characterize the boundedness of these operators.

We will recall the Theorem, and then prove a dyadic version of the result. An operator T is associated with a kernel $K(x, y)$ if

$$\langle \mathrm{T} f, g\rangle = \iint f(y)K(x,y)g(x)\, dydx \tag{2.8.1}$$

for all Schwartz functions f, g with disjoint supports. Notice that with this definition, the adjoint T^* is associated with kernel $L(x, y) = K(y, x)$.

2.8.2 Definition. An operator T is a *standard Calderón Zygmund operator* if T has kernel $K(x, y)$ which satisfies for some $C > 0$

$$|K(x,y)| \le C|x-y|^{-1}\,, \qquad \text{(Size Condition)} \tag{2.8.3}$$

$$|\nabla K(x,y)| \le C|x-y|^{-2}\,, \qquad \text{(Smoothness Condition)} \tag{2.8.4}$$

Let us set $\|K\|_{CZ}$ to be the least constant C for which the inequalities above are true.

Standard examples of such kernels are (1) the Hilbert transform, (2) in dimensions $d \ge 2$, one has the Riesz transforms R_j, for $1 \le j \le d$, where the kernel of R_j is $y_j/|y|^d$, (3) in dimension $d = 2$ one has the Beurling transform, given in complex variables by

$$\mathrm{B}\,\phi(z) = \int \phi(\zeta)\frac{d\zeta \wedge d\overline{\zeta}}{(z-\zeta)^2}\,.$$

But all of these are convolution kernels. A set of examples which were of great significance to the development of the subject were the Calderón Commutators. Let $A : \mathbb{R} \to \mathbb{R}$ be a Lipschitz function, and define kernels by

$$K_j(x,y) = \frac{(A(x)-A(y))^j}{(x-y)^{j+1}}\,. \tag{2.8.5}$$

Let C_j be the associated operators. These are standard Calderón Zygmund Operators, with a slight extension of (2.8.4).

It is sometimes useful to replace (2.8.4) by a weaker condition. For instance, in most instances, this assumption is sufficient. For some $o < \delta < 1$, we have

$$|K(x,y) - K(x',y)| + |K(y,x) - K(y,x')| \lesssim \frac{|x-x'|^{\alpha}}{|x-y|^{1+\alpha}} \qquad |x-x'| \le \tfrac{1}{2}|x-y| \tag{2.8.6}$$

The size condition (2.8.3) only permits the integral representation (2.8.1) to be interpreted in a distributional, or principal value sense. If we truncate the integrals involved, these difficulties are no longer present. So, for $0 < \epsilon < \delta < \infty$, let us set

$$\mathrm{T}_{\epsilon,\delta} f(x) = \int_{\epsilon < |x-y| < \delta} K(x,y) f(y)\, dy \, .$$

We agree that T is bounded on L^2 iff

$$\|\mathrm{T}\|_{2\to 2} := \sup_{0<\epsilon<\delta} \|\mathrm{T}_{\epsilon,\delta}\|_{2\to 2} < \infty \, . \tag{2.8.7}$$

It is natural to wonder why we insist on L^2 boundedness. As it turns out, L^2 boundedness implies L^p boundedness, for all $1 < p < \infty$.

2.8.8 Proposition. *Suppose* T *is a standard Calderón Zygmund Operator which is bounded on* L^2*. Then, for* $1 < p < \infty$*,*

$$\sup_{0<\epsilon<\delta} \|\mathrm{T}_{\epsilon,\delta}\|_{p\to p} \lesssim 1 \, ,$$

where the implied constant depends upon p*,* $\|\mathrm{T}\|_{2\to 2}$ *and* $\|K\|_{CZ}$*, where* T *has kernel* K*.*

Proof. One proof is to apply the Calderón Zygmund Decomposition to get the weak-type bound at L^1, following the model of the proof of Theorem 1.7.7.

We offer this argument as an alternative. Using Proposition 2.8.9 below, we see that T maps L^∞ to *BMO*. Appeal to Theorem 2.7.13 to finish the proof.

□

This endpoint estimate is central to the T 1 Theorem.

2.8.9 Proposition. *Suppose* T *is a standard Calderón Zygmund Operator which is bounded on* L^2*. Then* T *extends to a bounded operator from* L^∞ *to BMO. That is, the truncated operators* $\mathrm{T}_{\epsilon,\delta}$ *admit a uniform bound as maps from* L^∞ *to BMO.*

This proof requires a little attention.

Proof. Fix $0 < \epsilon < \delta$, and $f \in L^\infty$ of norm one. Also fix an interval I on which we are to test the *BMO* norm of $\mathrm{T} f$. Suppose $|I| < 4\delta$. Then, we estimate

$$\begin{aligned} \|\mathrm{T}_{\epsilon,\delta} f\|_{L^2(I)} &= \|\mathrm{T}_{\epsilon,\delta} f\mathbf{1}_{3I}\|_{L^2(I)} \\ &\le \|T\|_{2\to 2}\|f\mathbf{1}_{3I}\|_2 \\ &\le \sqrt{3}\|T\|_{2\to 2}|I|^{1/2}\,. \end{aligned}$$

That is, we need not subtract off the constant in this case.

Suppose $4|I| \le \epsilon$. We then argue that $\mathrm{T}_{\epsilon<\delta} f$ is essentially constant on I. That is, for any two $x, x' \in I$, we have $|\mathrm{T}_{\epsilon<\delta} f(x) - \mathrm{T}_{\epsilon<\delta} f(x')| \lesssim 1$. Let us estimate

$$\begin{aligned} |\mathrm{T}_{\epsilon<\delta} f(x) - \mathrm{T}_{\epsilon<\delta} f(x')| &= \Big|\int_{\epsilon<|x-y|<\delta} f(y)K(x,y)\,dy - \int_{\epsilon<|x'-y|<\delta} f(y)K(x',y)\,dy\Big| \\ &\le I_1 + I_2 + I_3\,, \\ I_1 &= \int_{\epsilon<|x-y|,|x'-y|} |K(x,y) - K(x',y)|\,dy \\ I_2 &= \int_{|x'-y|<\epsilon<|x-y|} |K(x,y)|\,dy \\ I_3 &= \int_{|x-y|<\epsilon<|x'-y|} |K(x',y)|\,dy \end{aligned}$$

By design, $4|x - x'| < \epsilon$, so that (2.8.4) applies to control I_1, namely

$$I_1 \le \int_{\epsilon<|x-y|,|x'-y|} \frac{|I|}{|x-y|^2}\,dy \lesssim \frac{|I|}{\epsilon}\,.$$

The remaining two estimates are similar, and we discuss I_2. The integration is over an interval of length $\lesssim |I|$, and by (2.8.3), the integrand is at most $|x - y|^{-1}$, so that $I_2 \lesssim \frac{|I|}{\epsilon}$.

Thus, if the interval that we test on is relatively large, or relatively small, with respect to the truncation levels, we have the desired estimate. Let us take an interval $\frac{1}{4}\epsilon < \ell = |I| \le 4\delta$. We can reduce this case to the previous two, by writing

$$\mathrm{T}_{\epsilon,\delta} = \mathrm{T}_{\epsilon,\frac{1}{4}\ell} + \mathrm{T}_{\frac{1}{4}\ell,4\ell} + \mathrm{T}_{4\ell,\delta}\ .$$

Of the three terms on the right, I is large with respect to the first term, and small with respect to the last term. As concerns the middle term, note that

$$|T_{\epsilon,\frac{1}{4}\ell} f(x)| \lesssim 1\,, x \in \mathbb{R}\,,$$

so it clearly has bounded mean oscillation on I. □

We can now turn to the T 1 Theorem.

2.8.10 Definition. A standard Calderón Zygmund Operator T is said to be weakly-bounded iff for all Schwartz functions φ, ϕ supported on the unit interval, we have

$$\sup_I \sup_{0<\epsilon<\delta} |\langle \mathrm{T}_{\epsilon,\delta} \operatorname{Dil}_I^{(2)} \varphi, \operatorname{Dil}_I^{(2)} \phi\rangle| \lesssim 1 \tag{2.8.11}$$

where the implied constant is uniform as φ and ϕ vary over compact sets of Schwartz functions.

2.8.12 T 1 **Theorem of David and Journé.** *A standard Calderón Zygmund Operator* T *is bounded on* L^2 *iff these three conditions are met*

- T *is weakly bounded;*
- $\mathrm{T}\,1 \in BMO$*;*
- $\mathrm{T}^*\,1 \in BMO$*.*

This is a striking result. Let us see how it applies to the Calderón Commutators. We will show that C_k are bounded by induction on k, assuming only that A is Lipschitz. All the operators C_k are standard Calderón Zygmund Operators, if we replace (2.8.4) by (2.8.6), for $\alpha = 1/2$ say. They also have odd kernels, whence $\mathrm{C}_k^* = -\mathrm{C}_k$, and they are weakly bounded—which point we leave as an exercise. Now, C_0 is the Hilbert transform, so that it is clearly an L^2-bounded operator. Note that an integration by parts gives us

$$\begin{aligned} \mathrm{C}_{k+1}\,1 &= \int \frac{(A(x)-A(y))^k}{(x-y)^{k+1}}\,dy \\ &= -\tfrac{1}{k}\int (A(x)-A(y))^k\,\frac{dy}{(x-y)^k} \\ &= \mathrm{C}_k\,A'(x)\,. \end{aligned}$$

Thus, the boundedness of C_k implies that of C_{k+1}.

We pass to a dyadic model of the T 1 Theorem, which will permit a transparent proof.

2.8.13 Definition. An operator T will be a *perfect Calderón Zygmund Operator* if it has a kernel $K(x,y)$ that satisfies (2.8.3), and the condition (2.8.4) is replaced by the stronger condition

$$\mathrm{K}(x,y) \text{ is constant on each cube of the form } I \times (I + |I|). \tag{2.8.14}$$

See Figure 2.3

2.8.15 Theorem. *[T1 Theorem of Perfect Calderón Zygmund Operators] A perfect Calderón Zygmund Operator* T *is bounded on* L^2 *iff these three conditions are met*

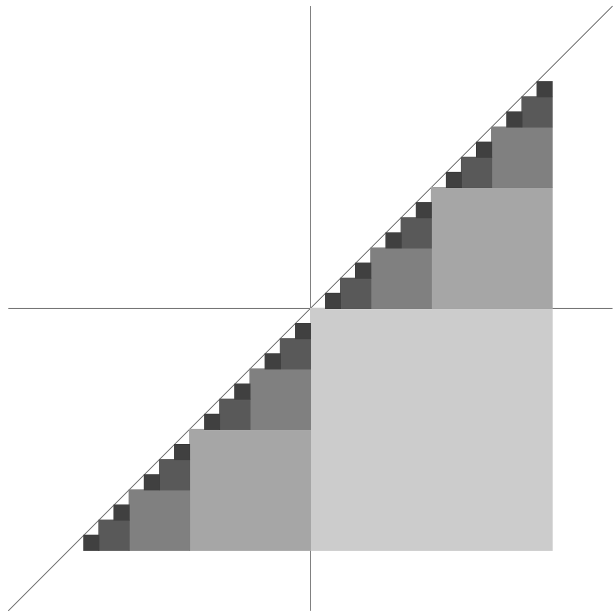


Figure 2.3: A Perfect Calderón Zygmund Kernel. The kernel is constant on each of the cubes indicated.

- *(Weakly bounded)* $\sup_I |\langle \mathrm{T}\, h_I^{\sigma}, h_I^{\sigma'} \rangle| \lesssim 1$,
- $\mathrm{T}\, 1 \in BMO$,
- $\mathrm{T}^*\, 1 \in BMO$.

Here, BMO is dyadic-BMO.

Proof. The three conditions are necessary, with an adaptation of the proof of Proposition 2.8.9.

In the converse direction, we will write T as

$$\mathrm{T}\, f = \mathrm{A}\, f + \mathrm{P}^{0,1,0}(\mathrm{T}\, 1, f) + \mathrm{P}^{0,0,1}(\mathrm{T}^*\, 1, f)\,, \tag{2.8.16}$$

where A is a Haar multiplier, and the next two terms are paraproducts. (Indeed, this is the essential idea of the proof of the T 1 Theorem.)

The kernel $K(x, y)$ is a function on the plane, and therefore, admits an expansion in the two-dimensional Haar basis. Recalling our notation from § 1.3, we write

$$K(x, y) = \sum_{\sigma \in \{(0,0),(0,1),(1,0)\}} \sum_{Q} \langle K, h_Q^{\sigma} \rangle h_Q^{\sigma}(x, y)$$

where Q ranges over all dyadic cubes in the plane. But, $K(x, y)$ is perfect, see (2.8.14) hence, $\langle K, h_Q^{\sigma} \rangle \neq 0$ implies that $Q = I \times I$. Thus, we have an expansion

$$\mathrm{T} = \sum_{\sigma \in \{(0,0),(0,1),(1,0)\}} \mathrm{A}^{\sigma}\,,$$

$$\begin{aligned} A^{(\sigma_1,\sigma_2)} f &= \sum_I \sum_I \langle K, h_{I\times I}^{(\sigma_1,\sigma_2)} \rangle \langle f, h_I^{\sigma_2} \rangle h_I^{\sigma_1}(x) \\ &= \sum_I \sum_I \langle T\, h_I^{\sigma_2}, h_I^{\sigma_1} \rangle \langle f, h_I^{\sigma_2} \rangle h_I^{\sigma_1}(x) \end{aligned}$$

For the term $A^{(0,0)}$, note that weak-boundedness implies that it is a Haar multiplier with a bounded sequence of multipliers, namely

$$A^{\sigma} f = \sum_I \epsilon_I \langle f, h_I \rangle h_I, \, , \qquad \sup_I |\epsilon_I| \lesssim 1 \, .$$

Observe that $A^{(0,1)} = P^{0,1,0}(T\,1, \cdot)$. Indeed, the fact that the kernel is perfect gives us the identity

$$\langle T\, h_I^1, h_I^0 \rangle = |I|^{-1/2} \langle T\, 1, h_I^0 \rangle$$

and so the claim follows. Similarly, it follows that $A^{(1,0)} = P^{0,0,1}(T^* \, 1, \cdot)$

□

Chapter 3

Weighted Inequalities

We take up the subject of weighted inequalities. Namely, in the simplest instance, we want to consider two Borel measures μ, ω on $\mathbb{R}^d$, and ask the question of whether or not the maximal function maps $L^2(\mu)$ into $L^2(\omega)$.

The rationale for this arose at first in the setting of, say, elliptic equations, where one or both of the measures would be a Harmonic measure. Increasingly, the interest of these questions lies in ongoing investigations of Geometric Measure Theory, spectral theory, and other subjects. As we will explain, some of these questions are of great interest, and very difficult.

The situation can be quite general. One can consider arbitrary Borel measures. Scale invariance properties, in full generality, are lost, so that one can consider inequalities for e. g. the Maximal Function mapping $L^2(\mu)$ into $L^4(\omega)$. For the purposes of this monograph, we will make some simplifications. In the first place, all measures will be continuous with respect to Lebesgue measure, and by abuse of notation, we will let μ denote both the measure and the density. We will also concentrate on the scale invariant case, namely L^p being mapped into L^p. These restrictions still encompass a number of difficulties.

3.1 Elementary Remarks

Throughout the subject is a notion of 'duality.'

3.1.1 Definition. Let (μ, ω) be a pair of positive measures. Let $1 < p < \infty$, and let $p' = p/(p-1)$. The *p-dual measure* (or just dual measure) is $\sigma = \mu^{-p'+1}$.

The point of this definition is that it is the measure that we can place on both sides of the inequality, as we explain in this proposition.

3.1.2 Proposition. *Consider a pair of strictly positive measures (μ, ω). We have the inequality*

$$\|\mathrm{M}(f)\|_{L^p(\omega)} \lesssim \|f\|_{L^p(\mu)} \tag{3.1.3}$$

if and only if we have the inequality below for the p-dual measure $\sigma = \mu^{1-p'} = \mu^{1/(1-p)}$*:*

$$\|\mathrm{M}(f\sigma)\|_{L^p(\omega)} \lesssim \|f\|_{L^p(\sigma)} \tag{3.1.4}$$

Note that the second inequality (3.1.4) makes sense regardless of assumption on the measure σ. Thus, even in the one-weight case of $\mu = \omega$, it can be useful to pass to a two-weight inequality.

Proof. Assuming (3.1.3), note that $p - \frac{1}{p'-1} = 1$, and so

$$\begin{aligned}\|\mathrm{M}(f\sigma)\|_{L^p(\omega)} &\lesssim \|f\sigma\|_{L^p(\mu)}\\ &= \Big[\int |f|^p\sigma^p\mu dx\Big]^{1/p}\\ &= \Big[\int |f|^p\sigma^{p-1/(p'-1)}dx\Big]^{1/p}\\ &= \|f\|_{L^p(\sigma)}\,.\end{aligned}$$

Assuming (3.1.4), note that $(p-1)(p'-1) = 1$, so that

$$\begin{aligned}\|\mathrm{M}(f)\|_{L^p(\omega)} &\lesssim \|f\sigma^{-1}\|_{L^p(\sigma)}\\ &= \Big[\int |f|\sigma^{-p+1}\,dx\Big]^{1/p}\\ &= \|f\|_{L^p(\mu)}\,,\end{aligned}$$

□

Some weighted inequalities are always available to us, and so we expect that they will be important in the development of the subject. The first, and most naive is that for any Borel measure μ we can define a Maximal Function adapted to μ by the definition

$$\mathrm{M}_\mu f(x) = \sup_{I\in\mathcal{D}} \frac{\mathbf{1}_I(x)}{\mu(I)} \int_I |f(y)|\,d\mu(y). \tag{3.1.5}$$

3.1.6 Proposition. *For any positive measure* μ *we have*

$$\|\mathrm{M}_\mu f\|_{L^{1,\infty}(\mu)} \le \|f\|_{L^1(\mu)} \tag{3.1.7}$$

$$\|\mathrm{M}_\mu f\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)}\,, 1 < p < \infty\,. \tag{3.1.8}$$

Proof. standard

□

Similarly, we have

3.1.9 Proposition. *For any positive measure* μ *we have*

$$\|\mathrm{M}\, f\|_{L^{1,\infty}(\mu)} \le \|f\|_{L^1(\mathrm{M}\,\mu)} \tag{3.1.10}$$

Proof. Fix $f \in \|f\|_{L^1(\mathrm{M}\,\mu)}$ of positive function of norm one, and take $\lambda > 0$. Let $\boldsymbol{Q}$ be the collection dyadic cubes with

$$|Q|^{-1} \int_Q f(y)\, dy \geq \lambda\,.$$

Let $\boldsymbol{Q}^*$ be the maximal cubes. Then,

$$\begin{aligned}\lambda \sum_{Q \in \boldsymbol{Q}^*} \omega(Q) &\leq \sum_{Q \in \boldsymbol{Q}^*} \frac{\omega(Q)}{|Q|} \int_Q f(y)\, dy \\ &\leq \|f\|_{L^1(\mathrm{M}\,\mu)}\end{aligned}$$

□

We have already seen that there is a close association between Maximal Functions and Carleson Measures. There is a corresponding version of such results for arbitrary measures. We make a series of definitions which are a companion to those of § 2.3. Given a positive measure μ and $\alpha = \{\alpha_Q \,:\, Q \in \mathcal{D}^d\}$, define

$$\|\alpha\|_{CM,\mu} = \sup_{R \in \mathcal{D}^d} \mu(R)^{-1} \sum_{Q\,:\,Q \subset R} \alpha_Q\,. \tag{3.1.11}$$

3.1.12 Weighted Dyadic Carleson Embedding Theorem. *For the notation above, and* $1 < p < \infty$ *we have the inequality below for positive functions* f

$$\sum_{Q \in \mathcal{D}^d} \alpha_Q \frac{\langle f\mu, \mathbf{1}_Q\rangle^p}{\mu(Q)^p} \lesssim \|\alpha\|_{CM,\mu} \|f\|_{L^p(\mu)}^p \tag{3.1.13}$$

Proof. Fix $f \in L^p$ For integers j let $\boldsymbol{Q}_j$ denote those dyadic cubes Q with

$$\langle f\mu, \mathbf{1}\rangle \simeq 2^j \mu(Q)\,.$$

and let $\boldsymbol{Q}_j^*$ be the maximal elements of $\boldsymbol{Q}_j$. It follows from (3.1.8) that we have

$$\sum_{j=-\infty}^{\infty} 2^{jp} \sum_{Q^* \in \boldsymbol{Q}_j^*} \mu(Q^*) \lesssim \|f\|_{L^p(\mu)}^p\,. \tag{3.1.14}$$

But then note that for each integer j,

$$\begin{aligned}\sum_{Q \in \boldsymbol{Q}_j} \alpha_Q \frac{\langle f\mu, \mathbf{1}_Q\rangle^p}{\mu(Q)^p} &\lesssim 2^{jp} \sum_{Q^* \in \boldsymbol{Q}_j} \sum_{Q\,:\,Q \subset^*} \alpha_Q \\ &\lesssim 2^{jp} \|\alpha\|_{CM,\mu} \sum_{Q^* \in \boldsymbol{Q}_j^*} \mu(Q^*)\end{aligned}$$

and so by (3.1.14), our proof is finished. □

3.1.15 Lemma. *Let $f, g \geq 0$ be measurable functions. Then, if $0 < p < 1$,*

$$\int fg \geq \|f\|_p \|g\|_{p'}, \tag{3.1.16}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$ (hence $p' < 0$), $\|f\|_p = \left(\int |f|^p\right)^{\frac{1}{p}}$, and

$$\|g\|_{p'} = \left(\int |g|^{p'}\right)^{\frac{1}{p'}} = \frac{1}{\left(\int \frac{1}{|g|^{-p'}}\right)^{\frac{1}{-p'}}}.$$

As a consequence,

$$\|f\|_p = \inf_{g\,:\,\|g\|_{p'}=1} \int fg. \tag{3.1.17}$$

Proof. The inequality (3.1.16) follows easily from the usual Hölder's inequality (i.e. with $p > 1$.) The case of equality in (3.1.17) is attained by taking $g = \frac{f^{p-1}}{\|f\|_p^{p-1}}$. □

3.2 The A_p Theory for the Maximal Function

The characterization of the inequality

$$\|\mathrm{M} f\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)}, \qquad \mu > 0$$

for a strictly positive measure μ is given by the Definition of the A_p characteristic of the weight μ, due to Muckenhoupt and Wheeden.

3.2.1 Definition. For any measure μ, set its (Muckenhoupt-Wheeden) A_p-characteristic to be

$$\mathrm{A}_p(\mu) = \sup_Q \frac{\mu(Q)}{|Q|} \cdot \left[\frac{\mu^{-1/(p-1)}(Q)}{|Q|}\right]^{p-1}. \tag{3.2.2}$$

The supremum is over all cubes Q. For $p = 1, \infty$, we interpret this as

$$\mathrm{A}_1(\mu) = \sup_x \frac{\mathrm{M}\,\mu(x)}{\mu(x)}$$
$$\mathrm{A}_\infty(\mu) = \sup_Q \frac{\mu(Q)}{|Q|} \cdot \sup_{x \in Q} \mu(x)^{-1}.$$

If $\mathrm{A}_p(\mu) < \infty$, we will write $\mu \in A_p$, and use the notation $A_1 \subset A_2$, for instance.

The basic fact here is the next Theorem, which proof we turn to in the next section.

3.2.3 The Muckenhoupt-Wheeden Theorem. *For a strictly positive measure μ,*

$$\|\mathrm{M}\|_{L^p(\mu)\to L^p(\mu)} \simeq \mathrm{A}_p(\mu)^{1/(p-1)} . \tag{3.2.4}$$

The definition of A_p is a subtle one, worthy of several comments. The conditions weaken as $p \to \infty$, thus $A_p \subset A_q$ for $1 \le p < q < \infty$. Indeed, the inequality

$$\mathrm{A}_p(\mu) \le \mathrm{A}_q(\mu) ,$$

as follows from the (3.2.2) and Hölders Inequality.

This also gives us the useful property that an A_p-measure must be nicely distributed. Take $\mathrm{A}_p(\mu) = 1$, and cube Q. For a small ϵ to be chosen, suppose that the set

$$E = \{x \in Q \,:\, \mu(x) \le \epsilon\mu(Q)/|Q|\}$$

has measure greater than $\frac{1}{2}|Q|$. Then, we would have

$$\begin{aligned} \frac{\mu(Q)}{|Q|} \cdot \left[\frac{\mu^{-1/(p-1)}(Q)}{|Q|}\right]^{p-1} &\ge \frac{\mu(Q)}{|Q|} \cdot \left[|Q|^{-1} \int_E \mu^{-1/(p-1)}\right]^{p-1} \\ &\ge \epsilon^{-1} 2^{-p+1} . \end{aligned}$$

And so for $0 < \epsilon = 2^{-p+1}$, we see a contradiction.

There are many definitions of A_∞, one of which is $A_\infty = \bigcup_{p\ge 1} A_p$.

Duality is expressed in this way. For $\mu \in A_p$, let $\sigma = \mu^{-p+1}$ be the dual weight. Then $\sigma \in A_{p'}$, as this interchanges the two terms in (3.2.2). In fact, note that $p' - 1 = 1/(p-1)$, and that the $A_{p'}$ condition for σ becomes

$$\frac{\sigma(Q)}{|Q|}\left[\frac{\sigma^{1-p}(Q)}{|Q|}\right]^{1/(p-1)} = \frac{\mu^{-p+1}(Q)}{|Q|}\left[\frac{\mu(Q)}{|Q|}\right]^{1/(p-1)} \le \mathrm{A}_p(\mu)^{1/(p-1)} \tag{3.2.5}$$

3.3 Proof of Theorem 3.2.3

Set $\sigma = \mu^{-1/(p-1)}$ to be the dual measure. The definition of A_p becomes

$$A_p(\mu) = \sup_Q A_p(Q,\mu) ,$$
$$A_p(Q,\mu) := \frac{\mu(Q)}{|Q|} \cdot \frac{\sigma(Q)^{p-1}}{|Q|^{p-1}} .$$

Our purpose in the initial step is to bound the Maximal Function M by a composition of M_μ and M_σ, where we use the notation of (3.1.5).

We have

$$\begin{aligned}\frac{1}{|Q|}\int_Q |f|\,dx &\le A_p(Q,\mu)^{1/(p-1)}\bigg\{\frac{|Q|}{\mu(Q)}\bigg[\frac{1}{\sigma(Q)}\int_Q |f|\,dx\bigg]^{p-1}\bigg\}^{1/(p-1)}\\ &\le A_p(\mu)^{1/(p-1)}\bigg\{\frac{1}{\mu(Q)}\int_Q \mathrm{M}_\sigma(f\sigma^{-1})^{p-1}dx\bigg\}^{1/(p-1)}.\end{aligned}$$

This leads us to the inequality

$$\mathrm{M}\,f \le A_p(\mu)^{1/(p-1)}\,\mathrm{M}_\mu(\mathrm{M}_\sigma(f\sigma^{-1})^{p-1}\mu^{-1})^{1/(p-1)}\,.$$

We use the fact that M_w maps $L^q(w)$ to $L^q(w)$, for all $1 < q < \infty$, to estimate

$$\begin{aligned}\|\mathrm{M}\,f\|_{L^p(\mu)} &\le A_p(\mu)^{1/(p-1)}\Big\|\mathrm{M}_\mu(\mathrm{M}_\sigma(f\sigma^{-1})^{p-1}\mu^{-1})^{1/(p-1)}\Big\|_{L^p(\mu)}\\ &= A_p(\mu)^{1/(p-1)}\Big\|\mathrm{M}_\mu(\mathrm{M}_\sigma(f\sigma^{-1})^{p-1}\mu^{-1})\Big\|_{L^{p'}(\mu)}^{1/(p-1)}\\ &\lesssim A_p(\mu)^{1/(p-1)}\Big\|\mathrm{M}_\sigma(f\sigma^{-1})^{p-1}\mu^{-1}\Big\|_{L^{p'}(\mu)}^{1/(p-1)}\\ &= A_p(\mu)^{1/(p-1)}\Big\|\mathrm{M}_\sigma(f\sigma^{-1})\Big\|_{L^p(\sigma)}\\ &\lesssim A_p(\mu)^{1/(p-1)}\|f\|_{L^p(\mu)}\end{aligned}$$

And the proof is complete.[1]

3.4 Sawyer's Two Weight Maximal Function Theorems

There is a complete characterization of the two-weight inequalities for the maximal function, found by E. Sawyer. Let us begin with the weak-type inequalities.

3.4.1 Sawyer's Two Weight Weak-Type Inequalities for the Maximal Function. *Let μ, σ be two positive Borel measures. Fix $1 < p\infty$. We have the inequalities*

$$\|\mathrm{M}(f\sigma)\|_{L^{p,\infty}(\mu)} \le C_0\|f\|_{L^p(\sigma)}\,, \qquad f \in L^p(\sigma)\,, \tag{3.4.2}$$

if and only if the two weight A_p condition holds

$$\sup_Q\Big[\frac{\mu(Q)}{|Q|}\Big]^{1/p}\Big[\frac{\sigma(Q)}{|Q|}\Big]^{1/p'} < \infty\,. \tag{3.4.3}$$

[1]This proof, the simplest we are aware of, is taken from

Lerner, Andrei K. 2008. *An elementary approach to several results on the Hardy-Littlewood maximal operator*, Proc. Amer. Math. Soc. **136**, 2829–2833. 2399047

Thus, the weak-type characterization has a characterization that is very similar to the one-weight results. The strong-type inequality requires more.

3.4.4 Sawyer's Two Weight Strong-Type Inequalities for the Maximal Function. *Let μ, σ be two positive Borel measures. Fix $1 < p \le q < \infty$. We have the inequalities*

$$\|\mathrm{M}(f\sigma)\|_{L^p(\mu)} \le C_0 \|f\|_{L^p(\sigma)}\,, \qquad f \in L^p(\sigma)\,, \tag{3.4.5}$$

if and only if there is a constant C_1 so that

$$\|\mathrm{M}(\mathbf{1}_Q \sigma)\|_{L^p(\mu)} \le C_0 \sigma(Q)^{1/p}.\,, \qquad Q \subset \mathbb{R}^d \text{ is a cube}\,. \tag{3.4.6}$$

Note that the second condition (3.4.6) is clearly necessary, and so this Theorem is a very natural analog of the T 1 Theorem. We have stated the Theorem in its dual measure formulation, and the argument of Proposition 3.1.2 applies to derive the two-weight version. Clearly this Theorem contains the Muckenhoupt-Wheeden result Theorem 3.2.3, but the deduction is not straight forward, as the conditions (3.4.6) do not immediately contain the A_p condition, when $p = q$, and $\sigma = \mu^{1-p'}$.

We shall see that the condition (3.4.6) will be interpreted as a Carleson measure estimate, in the spirit of Theorem 3.1.12.

Proof of Theorem 3.4.1. The necessity of (3.4.3) follows from applying the weak-type inequality to the indicator of a cube. So we turn to the sufficiency of (3.4.3).

Fix $1 < p < \infty$, positive f and $0 < \lambda < \infty$. Consider a cube Q with

$$\frac{1}{|Q|} \int_Q f \; \sigma dy \simeq \lambda\,. \tag{3.4.7}$$

Assuming that the supremum in (3.4.3) is one, we can estimate

$$\begin{aligned} \lambda^p \mu(Q) &\le |Q|^{-p} \Big[\int_Q f \; \sigma dy \Big]^p \mu(Q) \\ &\le \sigma(Q)^{-p/p'} \Big[\int_Q f \; \sigma dy \Big]^p \\ &\le \int_Q |f|^p \; \sigma dy\,. \end{aligned}$$

This estimate can be summed over a disjoint collection of cubes Q satisfying (3.4.7), which then proves the Theorem. □

Proof of Theorem 3.4.4. Take positive $f \in L^p(\sigma)$, and consider a linearization as in (1.5.9). For appropriate linearization, we can estimate

$$\|\mathrm{M} f\|_{L^p(\omega)}^p \lesssim \|\mathrm{T} f\|_{L^p(\omega)}^p \tag{3.4.8}$$

$$\leq \sum_Q \omega(E(Q)) \Big[\frac{1}{|Q|} \int_Q f \, \sigma dy \Big]^p . \tag{3.4.9}$$

Write the summands as below.

$$\omega(E(Q)) \Big[\frac{1}{|Q|} \int_Q f \, \sigma dy \Big]^p = \gamma_Q \Big[\frac{1}{\sigma(Q)} \int f \, \sigma dy \Big]^p$$

$$\gamma_Q := \omega(E(Q)) \Big[\frac{\sigma(Q)}{|Q|} \Big]^p .$$

We interpret our assumption (3.4.6) as asserting a Carleson Measure condition on the $\{\gamma_Q\}$. Namely, the γ_Q correspond themselves to a linearization of the Maximal Function, whence

$$\sum_{Q' \subset Q} \gamma_{Q'} \leq \int_Q (\mathrm{M}\, \mathbf{1}_p \sigma)^p \, dy \lesssim \sigma(Q) , \qquad Q \in \mathcal{D}^d .$$

Therefore we can apply Theorem 3.1.12, to estimate

$$(3.4.8) \leq \sum_Q \gamma_Q \Big[\frac{1}{\sigma(Q)} \int f \sigma dy \Big]^p \lesssim \int |f|^p \sigma \, dy .$$

The proof is complete.[2]

□

3.4.1 Exercises

3.4.10 Exercise. Let $\mu = w$, and $\sigma = \omega^{-1}$. For $1 < p < \infty$, show that the Sawyer two-weight condition (3.4.6) implies the A^p condition (3.2.2).

3.4.11 Exercise. Let $\mu = w$, and $\sigma = \omega^{-1}$. For $1 < p < \infty$, show that the A^p condition (3.2.2) implies the Sawyer two-weight condition (3.4.6). [The A^∞ condition will be helpful.]

3.5 A Theorem of Nazarov, Treil and Volberg

It is a profound, open question to characterize those pairs of weight (α, β) for which the two-weight inequality

$$\|\mathrm{H}\|_{L^2(\beta)} \lesssim \|f\|_{L^2(\alpha)} , \tag{3.5.1}$$

[2] This is the proof one can find in Sawyer's original article.

Sawyer, Eric T. 1982. *A characterization of a two-weight norm inequality for maximal operators*, Studia Math. **75**, 1–11. 676801 (84i:42032)

where H is, say, the Hilbert transform, or some other singular integral operator, such as Riesz transform, or Beurling operator. This stands in sharp contrast to the case of the Maximal Function, for which E. Sawyer has the beautiful result Theorem 3.4.4. Here, we should stress that we are interested in characterizations of real-variable type. There is a beautiful result Cotlar and Sadosky which characterizes (3.5.1), an extension of the Helson-Szegö Theorem, which provides a complex-variable characterization.

Among the class of singular integral operators, the known results are far from completely satisfactory. There is however one outstanding result due to Nazarov-Treil-Volberg, that considers a restricted class of operators, defined below. But then, provides a complete characterization, in the desired terms. We state and prove the result in one-dimensional case here.

We define the operators here. They are a class of operators which include the Haar shifts, as appeared in Proposition 1.4.4, and defined in § 1.4.1.

For the purposes of this section, we need a notion of distance between between two dyadic intervals, this will in fact be the standard tree-distance. A precise description is as follows. View the elements of $\mathcal{D}$ as the vertices of a graph. Connect I and J by an undirected edge iff $I \subset J$ and $2|I| = |J|$. Call the resulting graph $\mathcal{T}$. This graph is an unrooted binary tree. Then for any two dyadic intervals, $\delta(I, J) = d$ iff d is the length of the minimal path in $\mathcal{T}$ that connects I and J. Thus, $\delta([0, 1], [3, 4]) = 2$ while $\delta([-1, 0], [0, 1]) = \infty$.

3.5.2 Definition. We say that $\mathfrak{T}$ is *well-localized* if there is an integer r so that formally, we can write

$$\mathrm{T} f = \sum_{\substack{I,J\in\mathcal{D} \\ \delta(I,J)\le r}} \epsilon_{I,J} \langle f, h_I \rangle h_J \,, \tag{3.5.3}$$

3.5.4 Theorem. *[Nazarov-Treil-Volberg[3]]*

Let σ, w be two non-negative Borel measures, and let T *be a well-localized operator as in* (3.5.3)*. We have the inequality*

$$\|\mathrm{T}(f\sigma)\|_{L^2(w)} \lesssim \|f\|_{L^2(\sigma)} \tag{3.5.5}$$

iff the following three conditions hold uniformly in $Q \in \mathcal{D}$.

$$\left|\left\langle \mathrm{T}(\sigma \mathbf{1}_Q), \mathbf{1}_R \right\rangle_w\right| \lesssim \sqrt{w(R)\sigma(Q)}\,, \qquad \delta(R, Q) \le r\,, \tag{3.5.6}$$

$$\int_Q \left|\mathrm{T}(\sigma \mathbf{1}_Q)\right|^2 w(dx) \lesssim \sigma(Q)\,, \tag{3.5.7}$$

3

Nazarov, F., S. Treil, and A. Volberg. *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Research Letters, to appear, arXiv:math/0702758.

$$\int_Q \left| \mathrm{T}^*(w\mathbf{1}_Q) \right|^2 \sigma(dx) \lesssim w(Q)\,. \tag{3.5.8}$$

In (3.5.6)*, we are taking the inner product with respect to w-measure.*

We stress three points. First, this result is a characterization for *individual* operators T, not a class of operators. In particular, the operator T could be unbounded on Lebesgue measure. Second, despite the power of this result, as it applies to Haar shifts used to recover the Hilbert transform in Proposition 1.4.4, it does not solve the two-weight problem for the Hilbert transform, as the transform is obtained as an *average* of Haar shifts. Third, the Theorem is in close analogy to the T 1 Theorem, Theorem 2.8.12, in that the three conditions are analogous to the same three conditions in that Theorem. The condition (3.5.6) is analogous to the weak-boundedness condition, (3.5.7) is analogous to $\mathrm{T}\,1 \in BMO$, and (3.5.8) is analogous to $\mathrm{T}^*\,1 \in BMO$. The comparison goes beyond analogous, we will see that there is a strong formal connection between the proofs as well.

3.5.1 Weighted Haar Functions, Two Paraproducts

For any measure α on $\mathbb{R}$, we can define a class of Haar functions adapted to α. Let us make the definition

$$h_I^\alpha = \frac{-\alpha(I_{\text{right}})\mathbf{1}_{I_{\text{left}}} + \alpha(I_{\text{left}})\mathbf{1}_{I_{\text{right}}}}{\sqrt{\alpha(I_{\text{right}})^2\alpha(I_{\text{left}}) + \alpha(I_{\text{right}})\alpha(I_{\text{left}})^2}}\,, \qquad I \in \mathcal{D}\,. \tag{3.5.9}$$

The basic properties of these functions are given in

3.5.10 Proposition. *The functions* $\{h_I^\alpha \,:\, I \in \mathcal{D}\}$ *satisfy these properties.*

$$\{h_I^\alpha \,:\, I \in \mathcal{D}\} \text{ is an orthonormal basis for } L^2(\alpha). \tag{3.5.11}$$

$$\frac{1}{\alpha(I)} \int_I f \;\alpha\, dx = \sum_{J \supsetneq J} \langle f, h_J^\alpha \rangle_\alpha \, h_J^\alpha\,. \tag{3.5.12}$$

In the last line, the inner product is with respect to α-measure. This property is equivalent to

$$\Big\{ \sum_{|I| \geq 2^n} \langle f, h_I^\alpha \rangle_\alpha h_I^\alpha \,:\, n \in \mathbb{Z} \Big\} \text{ is a martingale with respect to } \alpha\text{-measure.}$$

The proof is elementary, following the analysis in § 1.2, so we leave the details as an exercise.

This last Proposition mentions martingales, see (2.5.3), and our discussion of the Chang-Wilson-Wolff Inequality, § 2.5. We will introduce some notation that is influenced by this connection. Set

$$\mathbb{E}^\alpha(f \,:\, I) = \frac{\mathbf{1}_I}{\alpha(I)} \int_I f \;\alpha\, dx \tag{3.5.13}$$

be a conditional expectation operator. Observe that

$$\langle f, h_I^\alpha \rangle_\alpha h_I^\alpha = \mathbb{E}(f : I_{\text{left}}) + \mathbb{E}(f : I_{\text{right}}) - \mathbb{E}(f : I)\,. \tag{3.5.14}$$

We make an important definition for our proof. Recall that T is well-localized, with r as in Definition 3.5.2. We now define paraproducts particular to our pair of measures w and σ.

Define two paraproduct operators by

$$\mathrm{P}^\sigma_{\mathrm{T}}(f) := \sum_{I \in \mathcal{D}} \mathbb{E}^\sigma(f : I) \sum_{\substack{K \in \mathcal{D} \\ |K| = 2^{-r}|I|}} \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w h_K^w\,, \tag{3.5.15}$$

$$\mathrm{P}^w_{\mathrm{T}^*}(f) := \sum_{I \in \mathcal{D}} \mathbb{E}^w(f : I) \sum_{\substack{K \in \mathcal{D} \\ |K| = 2^{-r}|I|}} \langle \mathrm{T}^*(w \mathbf{1}_I), h_K^\sigma \rangle_\sigma h_K^\sigma\,. \tag{3.5.16}$$

In these definitions, we have only required $K \in \mathcal{D}$ and $|K| = 2^{-r}|I|$. But, in order for the inner product $\langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w$ to be non-zero, we must also have $K \subset I^{(r)}$, where $I^{(r)}$ is the r^{th} grandparent of I. (The unique dyadic interval of length $2^r|I|$ that contans I.) We single this out in the

3.5.17 Proposition. *If $|K| \le 2^{-r}|I|$, and $K \not\subset I^{(r)}$, we have*

$$\langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w = 0\,,$$
$$\langle \mathrm{T}^*(w \mathbf{1}_I), h_K^\sigma \rangle_\sigma = 0\,.$$

That is, we can modify (3.5.15) to

$$\mathrm{P}^\sigma_{\mathrm{T}}(f) = \sum_{I \in \mathcal{D}} \mathbb{E}^\sigma(f : I) \sum_{\substack{K \in \mathcal{D} \\ |K| = 2^{-r}|I| \\ K \subset I^{(r)}}} \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w h_K^w\,,$$

A similar comment holds for (3.5.16).

Proof. We consider the first assertation. It follows that

$$\begin{aligned} \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w &= \sum_{\substack{Q,R \\ \delta(Q,R) \le r}} \epsilon(Q,R) \langle \sigma \mathbf{1}_I, h_R \rangle \langle h_Q, h_K^w \rangle_w \\ &= \sum_{\substack{Q,R \\ \delta(Q,R) \le r \\ Q \subsetneq K\,,\ R \cap I \ne \emptyset}} \epsilon(Q,R) \langle \sigma \mathbf{1}_I, h_R \rangle \langle h_Q, h_K^w \rangle_w \end{aligned} \tag{3.5.18}$$

In order for any summand on the right to be non-zero, we would have $Q \subsetneq K$, for one of the inner products to be non-zero, and and $R \cap I \ne \emptyset$ for the other. But the distance of Q to any dyadic interval that intersects I exceeds r, so we have a contradiction. The second asseration has a similar proof. □

There is a further remark about these definitions that we single out as a proposition.

3.5.19 Proposition. *Suppose $J \supset I$ and $K \subset I$ with $|K| = 2^{-r}|I|$. Then we have*

$$\langle \mathrm{T}(\sigma \mathbf{1}_J), h_K^w \rangle_w = \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w \,, \tag{3.5.20}$$

$$\langle \mathrm{T}^*(w \mathbf{1}_J), h_K^\sigma \rangle_\sigma = \langle \mathrm{T}^*(w \mathbf{1}_I), h_K^\sigma \rangle_\sigma \,. \tag{3.5.21}$$

That is, in (3.5.15) and (3.5.16), we could make the formal substitution of $\mathbf{1}$ for $\mathbf{1}_I$ in the inner products, in analogy to the formulation of the T 1 Theorem.

Proof. It suffices to prove one of the equalities, and so we only explicitly discuss (3.5.20). For any dyadic interval L which is a non-zero translation of I, we have

$$\langle \mathrm{T}(\sigma \mathbf{1}_L), h_K^w \rangle_w = 0\,.$$

Indeed, applying (3.5.18), we have

$$\langle \mathrm{T}(\sigma \mathbf{1}_L), h_K^w \rangle_w = \sum_{\substack{Q,R \\ \delta(Q,R) \le r \\ Q \subsetneq K\,,\ R \cap I \neq \emptyset}} \epsilon(Q,R) \langle \sigma \mathbf{1}_L, h_R \rangle \langle h_Q, h_K^w \rangle_w$$

But, any path connecting Q and R must necessarily pass through both K and I, and this is a contradiction to $\delta(Q,R) \le r$. Thus, the sum above is vacuous.

Now, observe that J can be written as a disjoint union of intervals L as above. And so, by linearity, the proposition is proved.

□

3.5.22 Lemma. *We have the two inequalities*

$$\|\mathrm{P}_{\mathrm{T}}^\sigma(\cdot)\|_{L^2(\sigma) \to L^2(w)} \lesssim \sup_{I \in \mathcal{D}} \frac{\|\mathrm{T}\, \sigma \mathbf{1}_I\|_{L^2(w)}}{\sqrt{\sigma(I)}}\,,$$

$$\|\mathrm{P}_{\mathrm{T}^*}^w(\cdot)\|_{L^2(w) \to L^2(\sigma)} \lesssim \sup_{I \in \mathcal{D}} \frac{\|\mathrm{T}^*\, w \mathbf{1}_I\|_{L^2(\sigma)}}{\sqrt{w(I)}}\,.$$

Proof. This is a consequence of our weighted Carleson Embedding Theorem, Theorem 3.1.12. By (3.5.11), we can estimate

$$\begin{aligned} \|\mathrm{P}_{\mathrm{T}}^\sigma(f)\|_{L^2(w)}^2 &= \sum_{I \in \mathcal{D}} [\mathbb{E}^\sigma(f \ :\ I)]^2 \Bigg\| \sum_{\substack{K \in \mathcal{D} \\ |K| = 2^{-r}|I| \\ K \subset I^{(r)}}} \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w \rangle_w h_K^w \Bigg\|_{L^2(w)}^2 \\ &= \sum_{I \in \mathcal{D}} \beta_I [\mathbb{E}^\sigma(f \ :\ I)]^2\,, \end{aligned}$$

$$\text{where} \quad \beta_I := \sum_{\substack{K\in\mathcal{D}\\ |K|=2^{-r}|I|\\ K\subset I^{(r)}}} \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w\rangle_w^2 .$$

Notice that we are taking local averages, and summing up with respect to some coefficients. That is this expression is just like as in a Carleson Embedding Theorem. The particular Embedding Theorem is Theorem 3.1.12, with $p = 2$. Hence, we will conclude our estimate for this paraproduct if we have the inequalities

$$\sum_{\substack{I\in\mathcal{D}\\ I\subset J}} \beta_I \lesssim \sigma(J), \qquad J \in \mathcal{D}. \tag{3.5.23}$$

Fix an interval J as in (3.5.23). Using (3.5.20), and orthogonality properties of the Haar functions, we can estimate

$$\begin{aligned}
\sum_{\substack{I\in\mathcal{D}\\ I\subset J}} \beta_I &= \sum_{\substack{I\in\mathcal{D}\\ I\subset J}} \sum_{\substack{K\in\mathcal{D}\\ |K|=2^{-r}|I|\\ K\subset I^{(r)}}} \langle \mathrm{T}(\sigma \mathbf{1}_I), h_K^w\rangle_w^2 \\
&= \sum_{\substack{I\in\mathcal{D}\\ I\subset J}} \sum_{\substack{K\in\mathcal{D}\\ |K|=2^{-r}|I|\\ K\subset I^{(r)}}} \langle \mathrm{T}(\sigma \mathbf{1}_J), h_K^w\rangle_w^2 \\
&\lesssim \sum_{\substack{K\in\mathcal{D}\\ K\subset J}} \langle \mathrm{T}(\sigma \mathbf{1}_J), h_K^w\rangle_w^2 \\
&\le \|\mathrm{T}(\sigma \mathbf{1}_J)\|_{L^2(w)}^2 \\
&\le \sigma(J) .
\end{aligned}$$

The concludes the proof of the first claim. The second claim follows by duality. □

3.5.2 The Paraproducts and the Weighted Inequality

We want to prove the inequality

$$\|\mathrm{T}(\sigma f)\|_{L^2(w)} \lesssim \|f\|_{L^2(\sigma)} .$$

By duality, this amounts to proving the inequality

$$|\langle \mathrm{T}(\sigma f), g\rangle_w| \lesssim \|f\|_{L^2(\sigma)} \|g\|_{L^2(w)} .$$

It suffices to prove this estimate for f, g being finite expansions of Haar functions, with respect to the appropriate weight. This we assume below. Here is the main decomposition, which is a rather precise analog of the decomposition in (2.8.16).

3.5.24 Lemma. *We have this equality.*

$$\langle \mathrm{T}(\sigma f), g\rangle_w = \sum_{I\in\mathcal{D}} \sum_{\substack{J\in\mathcal{D}\\ 2^{-r}|I|\le|J|\le 2^r|I|}} \langle f, h_I^\sigma\rangle_\sigma \langle \mathrm{T}(\sigma h_I^\sigma), h_J^w\rangle_w \langle g, h_J^w\rangle_w \tag{3.5.25}$$

$$+ \langle \mathrm{P}_{\mathrm{T}}^\sigma(f), g\rangle_w + \langle f, \mathrm{P}_{\mathrm{T}^*}^w(g)\rangle_\sigma \,. \tag{3.5.26}$$

Proof. The function f is a finite sum of functions h_I^σ and g is a finite sum of functions h_J^w. So, we have

$$\langle \mathrm{T}(\sigma f), g\rangle_w = \sum_{I\in\mathcal{D}} \sum_{J\in\mathcal{D}} \langle f, h_I^\sigma\rangle_\sigma \langle \mathrm{T}(\sigma h_I^\sigma), h_J^w\rangle_w \langle g, h_J^w\rangle_w \,.$$

Hence, the equality amounts to the following two dual assertions. First

$$\langle \mathrm{T}(\sigma h_I^\sigma), h_J^w\rangle_w = \langle \mathrm{P}_{\mathrm{T}}^\sigma(h_I^\sigma), h_J^w\rangle_w \,, \qquad |I| > 2^r|J| \,. \tag{3.5.27}$$

$$\langle \mathrm{T}(\sigma h_I^\sigma), h_J^w\rangle_w = \langle h_I^\sigma, \mathrm{P}_{\mathrm{T}^*}^w(h_J^w)\rangle_\sigma \,, \qquad |I| < 2^{-r}|J| \,. \tag{3.5.28}$$

Let us prove (3.5.27). Let I' be the descendant of I with $J \subset I'$ and $|J| = 2^{-r}|I|$. Then, h_I^σ is constant on I' and we can write, by (3.5.20),

$$\begin{aligned}\langle \mathrm{T}(\sigma h_I^\sigma), h_J^w\rangle_w &= \mathbb{E}(h_I^\sigma \;:\; I')\langle \mathrm{T}(\sigma \mathbf{1}_{I'}), h_J^w\rangle_w \\ &= \langle \mathrm{P}_{\mathrm{T}}^\sigma(h_I^\sigma), h_J^w\rangle_w \,.\end{aligned}$$

The last line follows by definition, see (3.5.15). Equality (3.5.28) follows by the dual argument.

□

The conclusion of the proof of Theorem 3.5.4 is achieved by examining the three terms on the right in (3.5.25) and (3.5.26). We can use Cauchy-Schwartz to estimate the term on the right in (3.5.25).

$$\begin{aligned}\text{right-hand side of (3.5.25)} &\le r^2 S \cdot \Big[\sum_{I\in\mathcal{D}} \langle f, h_I^\sigma\rangle_\sigma^2 \cdot \sum_{J\in\mathcal{D}} \langle g, h_J^w\rangle_w^2\Big]^{1/2} \\ &\lesssim S \cdot \|f\|_{L^2(\sigma)} \cdot \|g\|_{L^2(w)} \\ S &:= \sup_{\substack{I,J\in\mathcal{D}\\ 2^{-r}|I|\le|J|\le 2^r|I|}} |\langle \mathrm{T}(\sigma h_I^\sigma), h_J^w\rangle_w| \,.\end{aligned}$$

But the first assumption (3.5.6) controls the quantity S above. Indeed, the Haar functions are L^2-normalized sums of indicator of dyadic intervals.

Finally, the two terms in (3.5.26) are controlled by (3.5.7) and (3.5.8), together with Lemma 3.5.22. The proof is complete.

Appendix A

Exponential Orlicz Classes

Let $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ be a symmetric convex function with $\psi(x) = 0$ iff $x = 0$. Define the Orlicz norm

$$\|f\|_{\psi} := \inf\{C > 0 : \mathbb{E}\psi(f/C) \leq 1\}. \tag{A.0.1}$$

We take the infimum of the empty set to be $+\infty$, and denote by L^{ψ} to be the collection of functions for which $\|f\|_{\psi} < \infty$.

It is straight forward to see that $\|\cdot\|_{\psi}$ is in fact a norm, with the triangle inequality following from Jensen's inequality. If $\psi(x) = x^p$, then $\|\cdot\|_{\psi}$ is the usual L^p norm.

We are especially interested in the class of ψ given by

$$\psi_{\alpha}(x) = \mathrm{e}^{|x|^{\alpha}}, \qquad |x| \gtrsim 1.$$

Here, we insist upon equality for $|x|$ sufficiently large, depending upon x. We will write $L^{\psi_{\alpha}} = \exp(L^{\alpha})$. These are the exponential Orlicz classes.

Especially important is the the case of $\alpha = 2$, which is the class $\exp(L^2)$, of exponentially square integrable functions, of which the Gaussian random variables are a canonical example. A function $f \in \exp(L^2)$ is said to be *sub-gaussian.*

Using Stirling's formula, and the Taylor expansion for e^x, one can check that

A.0.2 Proposition. *We have the equivalence of norms*

$$\begin{aligned} \|f\|_{\exp(L^{\alpha})} &\simeq \sup_{p\geq 1} p^{-1/\alpha}\|f\|_p \\ &\simeq \sup_{\lambda>0} \lambda^{-\alpha} \log \mathbb{P}(|f| > \lambda). \end{aligned}$$

One also has a familiar Lemma for the maximum of random variables.

A.0.3 Lemma. *Let $X_1, \ldots, X_N$ be random variables in L^{ψ} of norm at most one. Then, we have*

$$\mathbb{E}\sup_{n\leq N}|X_N| \lesssim \psi^{-1}(N).$$

So for $X_1, \dots, X_N \in \exp(L^2)$ of norm one, we have

$$\mathbb{E} \sup_{n\le N} |X_N| \lesssim \sqrt{\log N + 1}\,. \tag{A.0.4}$$

Indeed, we will leave to the reader to verify that under the assumptions above

$$\| \sup_{n\le N} |X_N| \, \|_{\exp(L^2)} \lesssim \sqrt{\log N + 1}\,. \tag{A.0.5}$$

Proof. By Jensen's inequality

$$\begin{aligned} \psi(\mathbb{E} \sup_{n\le N} |X_N|) &\le \mathbb{E} \sup_{n\le N} \psi(|X_N|) \\ &\le \sum_{n=1}^{N} \mathbb{E}\psi(|X_N|) \\ &\lesssim N\,. \end{aligned}$$

The proof is complete. ◻

Another class of relevant spaces are given by the convex functions

$$\varphi_\beta(x) := |x|(\log 2 + |x|)\,.$$

We denote $L^{\varphi_\beta} = L(\log L)^\beta$. The connection with the exponential Orlicz classes is by way of duality.

$$[\exp(L^\alpha)]^* = L(\log L)^{1/\alpha}\,. \tag{A.0.6}$$

These spaces are closely associated with the *extrapolation* principle.

A.0.7 Proposition. *Let* T *be a linear operator with*

$$\|\mathrm{T}\|_{L^p([0,1]^d)\to L^p([0,1]^d)} \lesssim (p-1)^\alpha\,, \qquad 1 < p \le 2\,,\ 0 < \alpha < 1\,. \tag{A.0.8}$$

We then have the inequality

$$\|\mathrm{T} f\|_{L^1} \lesssim \|f\|_{L(logL)^\alpha}\,. \tag{A.0.9}$$

More generally,

$$\|\mathrm{T} f\|_{L^1(\log L)^\beta} \lesssim \|f\|_{L(logL)^{\alpha+\beta}}\,, \qquad 0 < \beta < \infty\,. \tag{A.0.10}$$

Proof. Let us consider (A.0.9). This inequality is dual to

$$\|\mathrm{T}^* f\|_{\exp(L)^{1/\alpha}} \lesssim \|f\|_\infty\,.$$

But, taking $f \in L^\infty$, with $\|f\|_\infty = 1$, and using (A.0.8), we have for $2 < p < \infty$,

$$\|\mathrm{T}^* f\|_p \lesssim p^\alpha$$

and so the dual estimate follows Proposition A.0.2. The inequality (A.0.10) is entirely similar. ◻

Appendix B

Khintchine Inequalities

The utility of the exponential Orlicz classes is that they allow a concise expression of a range of inequalities. This is especially relevant to the classical Khintchine Inequalities. In other instances we shall see, that Orlicz spaces express sharp inequalities forms of different inequalities.

Let $\{r_k \,:\, k \geq 1\}$ be independent, identically distributed random variables, with $\mathbb{P}(r_1 = 1) = \mathbb{P}(r_1 = -1) = \frac{1}{2}$. Such random variables are referred to as Rademacher random variables. They admit different realizations, of which the most direct is

$$r_k = \operatorname{sgn}(\sin(2^k \pi x))\,, \qquad 0 \leq x \leq 1\,.$$

Such random variables are in particular orthogonal, so that we have

$$\Big\| \sum_k a_k r_k \Big\|_2 = \Big[\sum_k a_k^2 \Big]^{1/2}\,.$$

This holds for all finite sequences of constants $\{a_k\}$.

The Khintchine Inequality says that these sums, in all L^p, are controlled by the L^2 norms. In its sharp form, this inequality states

B.0.1 Khintchine Inequalities. *For all finite sequences of constants* $\{a_k\}$

$$\Big\| \sum_k a_k r_k \Big\|_{\exp(L^2)} \lesssim \Big[\sum_k a_k^2 \Big]^{1/2}\,. \tag{B.0.2}$$

Proof. The classical proof of this is quite elementary, passing through the Moment Generating Function. We can restrict attention to the case where

$$\Big[\sum_k a_k^2 \Big]^{1/2} = 1\,.$$

Consider the moment generating function, given by

$$\varphi(\lambda) = \mathbb{E}\, \mathrm{e}^{\lambda \sum_k a_k r_k}\,, \qquad \lambda > 0$$

$$
\begin{aligned}
&= \prod_k \mathbb{E}\, e^{\lambda a_k r_k} \\
&= \prod_k \tfrac{1}{2}(e^{-\lambda a_k} + e^{\lambda a_k}) \\
&\le \prod_k e^{\lambda^2 a_k^2} \\
&\le e^{\lambda^2}
\end{aligned}
$$

Here, we have relied statistical independence of the random variables. In particular, if X, Y are independent random variables, then

$$\mathbb{E} X \cdot Y = \mathbb{E} X \cdot \mathbb{E} Y\,.$$

We have also used the the elementary inequality

$$\tfrac{1}{2}(e^{-\mu} + e^{\mu}) = \sum_{j=1}^{\infty} \frac{\mu^{2j}}{(2j)!} \le e^{\mu^2}\,, \qquad \mu \in \mathbb{R}\,. \tag{B.0.3}$$

Now estimate

$$\mathbb{P}\Big(\sum_k a_k r_k > t\Big) \le \varphi(\lambda)\, e^{-\lambda t} \le e^{\lambda^2 - \lambda t}\,, \qquad \lambda > 0\,.$$

The minimum over $\lambda > 0$ of the right hand side occurs at $\lambda = t/2$, giving us the estimate

$$\mathbb{P}\Big(\sum_k a_k r_k > t\Big) \le e^{-t^2/4}\,.$$

In view of the symmetry of the Rademacher random variables and Proposition A.0.2, this proves the Theorem.

□