LECTURE NOTES 4 FOR 247A

TERENCE TAO

1. The Hilbert transform

In this set of notes we begin the theory of *singular integral operators* - operators which are almost integral operators, except that their kernel $K(x,y)$ just barely fails to be integrable near the diagonal $x = y$. (This is in contrast to, say, fractional integral operators such as $Tf(y) := \int_{\mathbf{R}^d} \frac{1}{|x-y|^{d-s}} f(x)\, dx$, whose kernels go to infinity at the diagonal but remain locally integrable.) These operators arise naturally in complex analysis, Fourier analysis, and also in PDE (in particular, they include the *zeroth order pseudodifferential operators* as a special case, about which more will be said later). It is here for the first time that we must make essential use of *cancellation*: cruder tools such as Schur's test which do not take advantage of the sign of the kernel will not be effective here.

Before we turn to the general theory, let us study the prototypical singular integral operator[1], the *Hilbert transform*

$$Hf(x) := p.v. \frac{1}{\pi} \int_{\mathbf{R}} \frac{f(x-t)}{t}\, dt := \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_{|t|>\varepsilon} \frac{f(x-t)}{t}\, dt. \quad (1)$$

It is not immediately obvious that $Hf(x)$ is well-defined even for nice functions f. But if f is C^1 and compactly supported, then we can restrict the integral to a compact interval $t \in [-R, R]$ for some large R (depending on x and f), and use symmetry to write

$$\int_{|t|>\varepsilon} \frac{f(x-t)}{t}\, dt = \int_{\varepsilon<|t|<R} \frac{f(x-t)-f(x)}{t}\, dt.$$

The mean-value theorem then shows that $\frac{f(x-t)-f(x)}{t}$ is uniformly bounded on the interval $t \in [-R, R]$ for fixed f, x, and so the limit actually exists from the dominated convergence theorem. A variant of this argument shows that Hf is also well-defined for f in the Schwartz class, though it does not map the Schwartz class to itself. Indeed it is not hard to see that we have the asymptotic

$$\lim_{|x|\to\infty} xHf(x) = \frac{1}{\pi} \int_{\mathbf{R}} f$$

for such functions, and so if f has non-zero mean then Hf only decays like $1/|x|$ at infinity. In particular, we already see that H is not bounded on L^1. However, we do see that H at least maps the Schwartz class to $L^2(\mathbf{R})$.

[1] The other prototypical operator would be the identity operator $Tf = f$, but that is too trivial to be worth studying.

Let us first make comment on some algebraic properties of the Hilbert transform. It commutes with translations and dilations (but not modulations):

$$H\mathrm{Trans}_{x_0} = \mathrm{Trans}_{x_0} H; \quad \mathrm{Dil}^p_\lambda H = H\mathrm{Dil}^p_\lambda. \tag{2}$$

In fact, it is essentially the only such operator (see Q1). It is also formally skew-adjoint, in that for C^1, compactly supported f, g:

$$\int_{\mathbf{R}} Hf(x)\overline{g(x)}\ dx = -\int_{\mathbf{R}} f(x)\overline{Hg(x)}\ dx;$$

we leave the verification as an exercise.

The Hilbert transform is connected to complex analysis (and in particular to Cauchy integrals) by the following identities.

Proposition 1.1 (Plemelj formulae)**.** *Let $f \in C^1(\mathbf{R})$ obey a qualitative decay bound $f(x) = O_f(\langle x\rangle^{-1})$ (say - these conditions are needed just to make Hf well-defined). Then for any $x \in \mathbf{R}$*

$$\frac{1}{2\pi i}\lim_{\varepsilon\to 0}\int_{\mathbf{R}} \frac{f(y)}{y-(x\pm i\varepsilon)}\ dy = \frac{\pm f(x) + iHf(x)}{2}.$$

Proof By translation invariance we can take $x = 0$. By taking complex conjugates we may assume that the $\pm$ sign is $+$. Our task is then to show that

$$\lim_{\varepsilon\to 0}\frac{1}{2\pi i}\int_{\mathbf{R}}\frac{f(y)}{y-i\varepsilon}\ dy - \frac{1}{2}f(0) - \frac{i}{2\pi}\int_{|y|>\varepsilon}\frac{f(y)}{-y}\ dy = 0.$$

Multiplying by $2\pi i$ and making the change of variables $y = \varepsilon w$, we reduce to showing

$$\lim_{\varepsilon\to 0}\int_{\mathbf{R}} f(\varepsilon w)(\frac{1}{w-i} - 1_{|w|>1}\frac{1}{w})\ dw - \pi i f(0) = 0.$$

Direct computation shows that the expression in parentheses is absolutely integrable and

$$\int_{\mathbf{R}}(\frac{1}{w-i} - 1_{|w|>1}\frac{1}{w}) = \pi i.$$

So we reduce to showing that

$$\lim_{\varepsilon\to 0}\int_{\mathbf{R}}(f(\varepsilon w) - f(0))(\frac{1}{w-i} - 1_{|w|>1}\frac{1}{w})\ dw = 0.$$

The claim then follows from dominated convergence. ■

Now suppose that f not only obeys the hypotheses of the Plemelj formulae, but also extends holomorphically to the upper half-plane $\{z \in \mathbf{C} : \mathrm{Im}(z) \geq 0\}$ and obeys the decay bound $f(z) = O_f(\langle z\rangle^{-1})$ in this region. Then Cauchy's theorem gives

$$\frac{1}{2\pi i}\lim_{\varepsilon\to 0}\int_{\mathbf{R}}\frac{f(y)}{y-(x+i\varepsilon)}\ dy = f(x+i\varepsilon)$$

and

$$\frac{1}{2\pi i}\lim_{\varepsilon\to 0}\int_{\mathbf{R}}\frac{f(y)}{y-(x-i\varepsilon)}\ dy = 0$$

and thus by either of the Plemelj formulae we see that $Hf = -if$ in this case. In particular, comparing real and imaginary parts we conclude that $\mathrm{Im}(f) = H\mathrm{Re}(f)$

and $\mathrm{Re}(f) = -H\mathrm{Im}(f)$. Thus for reasonably decaying holomorphic functions on the upper half-plane, the real and imaginary parts of the boundary value are connected via the Hilbert transform. In particular this shows that such functions are uniquely determined by just the real part of the boundary value.

The above discussion also strongly suggests the identity $H^2 = -1$. This can be made more manifest by the following Fourier representation of the Hilbert transform.

Proposition 1.2. *If $f \in \mathcal{S}(\mathbf{R})$, then*

$$\widehat{Hf}(\xi) = -i\mathrm{sgn}(\xi)\hat{f}(\xi) \tag{3}$$

for (almost every) $\xi \in \mathbf{R}$. (Recall that Hf lies in L^2 and so its Fourier transform is defined via density by Plancherel's theorem.)

Proof Let us first give a non-rigorous proof of this identity. Morally speaking we have

$$Hf = f * \frac{1}{\pi x}$$

and so since Fourier transforms ought to interchange convolution and product

$$\widehat{Hf}(\xi) = \hat{f}(\xi)\widehat{\frac{1}{\pi x}}(\xi).$$

On the other hand we have

$$\widehat{1}(\xi) = \delta(\xi)$$

so on integrating this (using the relationship between differentiation and multiplication) we get

$$\widehat{\frac{1}{-2\pi i x}}(\xi) = \frac{1}{2}\mathrm{sgn}(\xi)$$

and the claim follows.

Now we turn to the rigorous proof. As the Hilbert transform is odd, a symmetry argument allows one to reduce to the case $\xi > 0$. It then suffices to show that $\frac{-f+iHf}{2}$ has vanishing Fourier transform in this half-line. Define the Cauchy integral operator

$$\mathcal{C}_\varepsilon f(x) = \frac{1}{2\pi i}\int_{\mathbf{R}} \frac{f(y)}{y-(x-i\varepsilon)}\ dy.$$

The Plemelj formulae show that these converge pointwise to $\frac{-f+iHf}{2}$ as $\varepsilon \to 0$; since f is Schwartz, it is also not hard to show via dominated convergence that they also converge in L^2. Thus by the L^2 boundedness of the Fourier transform it suffices to show that each of the $\mathcal{C}_\varepsilon f$ also have vanishing Fourier transform on the half-line[2]. Fix $\varepsilon > 0$ (we could rescale ε to be, say, 1, but we will not have need of

[2]This is part of a more general phenomenon, that functions which extend holomorphically to the upper half-space in a controlled manner tend to have vanishing Fourier transform on $\mathbf{R}^-$, while those which extend to the lower half-space have vanishing Fourier transform on $\mathbf{R}^+$. Thus there is a close relationship between holomorphic extension and distribution of the Fourier transform. More on this later.

this normalisation). We can truncate and define

$$\mathcal{C}_{\varepsilon,R} f(x) = \frac{1}{2\pi i} \int_{\mathbf{R}} \frac{f(y)}{y - (x - i\varepsilon)} 1_{|y-x|<R}\ dy;$$

more dominated convergence shows that $\mathcal{C}_{\varepsilon,R} f$ converges to $\mathcal{C}_\varepsilon f$ in L^2 as $R \to +\infty$ and so it will suffice to show that the Fourier coefficients of $\mathcal{C}_{\varepsilon,R} f$ converge pointwise to zero as $R \to +\infty$ on the half-line $\xi > 0$. From Fubini's theorem we easily compute

$$\widehat{\mathcal{C}_{\varepsilon,R} f}(\xi) = \frac{1}{2\pi i} \hat{f}(\xi) \int_{\mathbf{R}} \frac{e^{-2\pi i y \xi}}{-(y - i\varepsilon)} 1_{|y|<R}\ dy;$$

but then by shifting the contour to the lower semicircle of radius R and then letting $R \to \infty$ we obtain the claim. ■

From this proposition and Plancherel's theorem we conclude that H is an isometry:

$$\|Hf\|_{L^2(\mathbf{R})} = \|f\|_{L^2(\mathbf{R})} \text{ for all } f \in \mathcal{S}(\mathbf{R}).$$

Because of this, H has a unique dense extension to $L^2(\mathbf{R})$, and (by another application of Plancherel's theorem to justify taking the limit) the formula (3) is valid for all $f \in L^2(\mathbf{R})$. There is still the question of whether the original definition is valid for $f \in L^2(\mathbf{R})$, or in other words whether the sequence of functions

$$\frac{1}{\pi} \int_{|t|>\varepsilon} \frac{f(x-t)}{t}\ dt$$

(note that Cauchy-Schwarz ensures that the integrand is absolutely integrable) converges as $\varepsilon \to 0$ to Hf for $f \in L^2(\mathbf{R})$, and in what sense this convergence holds. It turns out that the convergence is true both pointwise almost everywhere and in the $L^2(\mathbf{R})$ metric sense, and in fact it follows from variants of the Calderón-Zygmund theory presented in this notes, but we shall not do so here (as it may be covered in the participating student seminar).

From (3) we also see that we do indeed have the identity $H^2 = -1$ on $L^2(\mathbf{R})$, and that H is indeed skew-adjoint on $L^2(\mathbf{R})$ ($H^* = -H$); in particular, H is unitary.

Let us briefly mention why the Hilbert transform is also connected to the theory of partial Fourier integrals.

Proposition 1.3. *Let $f \in L^2(\mathbf{R})$ and $N_+, N_- \in \mathbf{R}$. Then*

$$\int_{-N_-}^{N_+} \hat{f}(\xi) e^{2\pi i x \xi}\ d\xi = \frac{i}{2} (\mathrm{Mod}_{N_+} H \mathrm{Mod}_{-N_+} f - \mathrm{Mod}_{-N_-} H \mathrm{Mod}_{N_-} f).$$

Proof By limiting arguments (and because everything is continuous on $L^2(\mathbf{R})$) it suffices to verify this identity for f in the Schwartz class. But then the claim follows by taking the Fourier transform of both sides and using (3). ■

It is thus clear that in order to address the question of the extent to which the partial Fourier integrals $\int_{-N_-}^{N_+} \hat{f}(\xi) e^{2\pi i x \xi}\ d\xi$ actually converges back to f, we will need to understand the properties of the Hilbert transform, and in particular its boundedness properties.

The formula (3) suggests that H behaves "like" multiplication by $\pm i$ in some sense; informally, H behaves like $-i$ for positive frequency functions and $+i$ for negative frequency functions. Now multiplication by $\pm i$ is a fairly harmless operation - it is bounded on every normed vector space - so one might expect H to similarly be bounded on more normed spaces than just $L^2(\mathbf{R})$; in particular, it might be bounded on $L^p(\mathbf{R})$. This is indeed the case for $1 < p < \infty$, and we now turn to the theory which will generate such bounds.

Problem 1.4. Demonstrate by example that the Hilbert transform is not bounded on $L^1(\mathbf{R})$ or $L^\infty(\mathbf{R})$.

2. Calderón-Zygmund theory

In the Hilbert transform, we have an operator which is bounded on $L^2(\mathbf{R})$, and we wish to extend this boundedness to other $L^p(\mathbf{R})$ spaces as well. This is of course not automatic - the "enemy" is that the operator might map a broad shallow function into a tall narrow function of comparable L^2 norm (which will increase the L^p norms markedly for all $p > 2$), or conversely map a tall narrow function down to a broad shallow function of comparable L^2 norm (which will increase the L^p norms markedly for $p < 2$). It turns out that these two phenomena are more or less dual to each other, and so when working with an operator which is "comparable" to its adjoint (which is for instance the case with the skew-adjoint Hilbert transform, $H^* = -H$) eliminating one of these enemies will automatically remove the other.

For operators such as the Hilbert transform, there are basically two approaches[3] to exclude these scenarios. One is to exploit the phenomenon that if f is not too large locally, then Hf also doesn't fluctuate too much locally; this prevents the "broad shallow to tall narrow" enemy which blocks boundedness for large p. The dual approach is to show that if f is localised and fluctuating, then Hf is also localised; this prevents the "tall narrow to broad shallow" enemy which blocks boundedness for small p.

We shall first discuss the latter approach, which is the traditional approach to the subject, and return to the former approach later. We shall generalise from the Hilbert transform to a useful wider class of operators, which are integral operators "away from the diagonal". To motivate this concept, observe that if $f \in L^2(\mathbf{R})$ is compactly supported and y lies outside of the support of f, then

$$Hf(y) = \int_{\mathbf{R}} \frac{1}{\pi(y-x)} f(x)\ dx,$$

the point being that the integral is now absolutely convergent, Thus while H is not, strictly speaking, an integral operator of the type studied in preceding sections, it can be described as such as long as one avoids the diagonal $x = y$.

[3]There are also more algebraic methods, relying on explicit formulae for things like the L^4 norm of Hf, or of the size of level sets of $H1_E$ for various sets E, but we will not discuss those here as they do not extend well to the more general class of singular kernels considered here.

Definition 2.1 (Singular kernels)**.** A *singular kernel* is a function $K : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{C}$ (defined away from the diagonal $x = y$) which obeys the decay estimate

$$|K(x,y)| \lesssim |x-y|^{-d} \tag{4}$$

for $x \neq y$ and the Hölder-type regularity estimates

$$K(x,y') = K(x,y) + O(\frac{|y-y'|^\sigma}{|x-y|^{d+\sigma}}) \text{ whenever } |y-y'| \leq \frac{1}{2}|x-y| \tag{5}$$

and

$$K(x',y) = K(x,y) + O(\frac{|x-x'|^\sigma}{|x-y|^{d+\sigma}}) \text{ whenever } |x-x'| \leq \frac{1}{2}|x-y| \tag{6}$$

for some *Hölder exponent* $0 < \sigma \leq 1$.

Example 2.2. *In one dimension,* $\frac{1}{x-y}$ *and* $\frac{1}{|x-y|}$ *are both singular kernels. In higher dimensions, the kernel* $K(x,y) := \Omega(\frac{x-y}{|x-y|})/|x-y|^d$ *is a singular kernel whenever* $\Omega : S^{d-1} \to \mathbf{C}$ *is a Hölder-continuous function.*

Remarks 2.3. The regularity estimates may seem strange at first sight, although the fundamental theorem of calculus shows that these estimates would follow (with $\sigma = 1$) from the estimates

$$\nabla_x K(x,y), \nabla_y K(x,y) = O(|x-y|^{-d-1});$$

given that the gradient of $|x-y|^{-d}$ is $O(|x-y|^{d-1})$, these estimates are then consistent with (4) but of course do not follow from it. In most applications one can take $\sigma = 1$ (in which case the Hölder type conditions become Lipschitz type conditions). It is more desirable (5), (6) for large σ than for small σ (note in particular that the $\sigma = 0$ version of the estimate already follows from (4)). It is possible to weaken the Hölder type conditions further, but there are few applications that require such a weakening and so we shall stay with the Hölder condition. The factor $\frac{1}{2}$ is not significant; it is not hard to show that the regularity conditions are equivalent to the condition in which $\frac{1}{2}$ is replaced by any other constant strictly between zero and one.

Definition 2.4 (Calderón-Zygmund operators)**.** A *Calderón-Zygmund operator* (CZO) is a linear operator $T : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$ which is bounded on $L^2(\mathbf{R}^d)$,

$$\|Tf\|_{L^2(\mathbf{R}^d)} \lesssim \|f\|_{L^2(\mathbf{R}^d)},$$

and such that there exists a singular kernel K for which

$$Tf(y) = \int_{\mathbf{R}^d} K(x,y) f(x)\ dx$$

whenever $f \in L^2(\mathbf{R}^d)$ is compactly supported, and y lies outside of the support of f.

Remarks 2.5. It is not hard to see that T uniquely determines K. However the converse is not true; for instance, given any bounded function $b : \mathbf{R}^d \to \mathbf{C}$, the physical space multiplier $Tf := bf$ is a CZO with kernel 0. However, this is the only source of ambiguity (Q2). A CZO is an example of a *singular integral operator*; the kernel K just barely fails to be locally integrable (Schur's test, for instance, encounters a logarithmic divergence), and so there is a singularity in the integration which must be understood in order to analyse T correctly. In particular, not all singular kernels K lead to an CZO, due to possible failure of L^2 boundedness. We will return to the question of working out which kernels do give CZOs later.

Remark 2.6. As with the theory of the Hardy-Littlewood maximal function, one can define CZOs on more general homogeneous spaces than $\mathbf{R}^d$, but we will not do so here for sake of concreteness.

By convention, we allow all implied constants in what follows to depend on the Hölder exponent σ and on the implied constants in (4), (5), (6) and the operator norm bound on T. (One could try to formalise things by imposing a norm on CZOs with a given exponent σ, but this gets a bit confusing and is largely unnecessary since one mostly deals with just a single CZO at a time.)

The Hilbert transform is thus a CZO, with kernel $K(x,y) = \frac{1}{\pi(y-x)}$. We will discuss other examples of CZOs, such as the Riesz transforms, Hörmander-Mikhlin multipliers, and pseudodifferential operators of order zero, later on.

The Hilbert transform was skew-adjoint, translation-invariant, and scale-invariant. In general, CZOs are neither. However, the *class* of CZOs is self-adjoint (and skew-adjoint), translation-invariant, and scale-invariant; if T is a CZO then so is T^*, and the translates $\mathrm{Trans}_{x_0} T \mathrm{Trans}_{-x_0}$ and dilates $\mathrm{Dil}^p_{1/\lambda} T \mathrm{Dil}^p_\lambda$ of T are also CZOs uniformly in $\lambda > 0$ and $x_0 \in \mathbf{R}^d$. (What happens to the kernels K in each case?)

Now we formalise the key phenomenon alluded to earlier, that CZOs map fluctuating localised functions to nearly localised functions.

Lemma 2.7. *Let $B = B(x_0, r)$ be a ball in $\mathbf{R}^d$, let T be a CZO, and let $f \in L^1(B)$ have mean zero, thus $\int_B f = 0$. Then we have the pointwise bound*

$$|Tf(y)| \lesssim_d \frac{r^\sigma}{|y-x_0|^{d+\sigma}} \int_B |f| \text{ for all } y \notin 2B.$$

In particular, we have

$$\|Tf\|_{L^1(\mathbf{R}^d \backslash 2B)} \lesssim_d \|f\|_{L^1(B)}.$$

Proof We remark that we could use the translation and scaling symmetries of CZOs to normalise $(x_0, r) = (0, 1)$, but we shall not do so here. For $y \notin 2B$ we can use the mean zero hypothesis to obtain

$$Tf(y) = \int_B K(x,y) f(x)\ dx = \int_B [K(x,y) - K(x_0,y)] f(y)\ dy,$$

and the first claim then follows from (5) and the triangle inequality. The second claim of course follows from the first (recall we allow our implied constants to depend on σ and d). ■

Remark 2.8. Observe that the argument also gives the slightly more general estimate

$$Tf(y) = K(x_0, y) \int_B f + O(\frac{r^\sigma}{|y-x_0|^{d+\sigma}} \int_B |f|) \text{ for all } y \notin 2B \tag{7}$$

when the mean zero condition $\int_B f = 0$ is dropped.

This lemma asserts that T is sort of bounded on L^1 - but only when the input function is localised and mean zero, and as long as one doesn't look too closely at the region near the support of the function. This extremely conditional L^1 boundedness is a far cry from true L^1 boundedness; indeed, it is not hard to show that L^1 boundedness can fail. However, Lemma 2.7, combined with the Calderón-Zygmund decomposition from the previous section (which decomposes any function into a bounded part and a lot of localised, fluctuating parts) is strong enough to obtain the weak-type boundedness instead.

Corollary 2.9. *All CZOs are of weak type* $(1,1)$*, with an operator norm of* $O_d(1)$.

Proof Let $f \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$ (the L^2 condition ensuring that T is well-defined), let T be a CZO, and let $\lambda > 0$. Our task is to show that

$$|\{|Tf| \geq \lambda\}| \lesssim_d \frac{\|f\|_{L^1(\mathbf{R}^d)}}{\lambda}.$$

We spend scaling and homogeneity symmetries to normalise $\|f\|_{L^1(\mathbf{R}^d)}$ and λ to both be 1; this avoids a certain amount of book-keeping regarding these factors later. Applying the Calderón-Zygmund decomposition, we can split $f = g + \sum_{Q \in \mathcal{Q}} b_Q$, where the "good" function g obeys the bounds

$$\|g\|_{L^1(\mathbf{R}^d)}, \|g\|_{L^\infty(\mathbf{R}^d)} \lesssim_d 1,$$

each b_Q is supported on Q, has mean zero, and has the L^1 bound

$$\|b_Q\|_{L^1(\mathbf{R}^d)} \lesssim_d |Q|,$$

and

$$\sum_{Q \in \mathcal{Q}} |Q| \lesssim_d 1.$$

Now since $Tf = Tg + \sum_{Q \mathcal{Q}} Tb_Q$, we can split

$$|\{|Tf| \geq 1\}| \leq |\{|Tg| \geq 1/2\}| + |\{\sum_{Q \in \mathcal{Q}} |Tb_Q| \geq 1/2\}|.$$

For the first term, we exploit the L^2 boundedness of T via Chebyshev's inequality and log-convexity of Hölder norms:

$$|\{|Tg| \geq 1/2\}| \lesssim_d \|Tg\|_{L^2(\mathbf{R}^d)}^2 \lesssim_d \|g\|_{L^2(\mathbf{R}^d)}^2 \lesssim_d \|g\|_{L^1(\mathbf{R}^d)} \|g\|_{L^\infty(\mathbf{R}^d)} \lesssim_d 1.$$

For the second term, we cannot rely on L^2 bounds because the b_Q have no reason to be bounded in L^2. But the mean zero condition, combined with Lemma 2.7, shows that each Tb_Q has L^1 bounds outside of a dilate CQ of Q (for some fixed constant[4] C depending only on d);

$$\|Tb_Q\|_{L^1(\mathbf{R}^d \backslash CQ)} \lesssim_d \|b_Q\|_{L^1(Q)} \lesssim_d |Q|.$$

Restricting to the common domain $\mathbf{R}^d \backslash \bigcup_{Q \in \mathcal{Q}} CQ$ and summing in Q using the triangle inequality we obtain

$$\| \sum_{Q \in \mathcal{Q}} Tb_Q\|_{L^1(\mathbf{R}^d \backslash \bigcup_{Q \in \mathcal{Q}} CQ)} \lesssim_d \sum_{Q \in \mathcal{Q}} |Q| \lesssim_d 1$$

[4]Actually one could take C arbitrarily close to 1 by iterating (5), (6) to obtain a variant of Lemma 2.7, but we will not do so here.

and hence by Chebyshev

$$|\{x \in \mathbf{R}^d \backslash \bigcup_{Q \in \mathcal{Q}} CQ : |\sum_{Q \in \mathcal{Q}} Tb_Q| \geq \frac{1}{2}\}| \lesssim_d 1.$$

This still leaves the question of what happens in the exceptional set $\bigcup_{Q \in \mathcal{Q}} CQ$. But here is the great advantage of weak-type estimates: we can tolerate arbitrarily bad behaviour as long as it is localised to small sets. Indeed we have

$$|\bigcup_{Q \in \mathcal{Q}} CQ| \leq \sum_{Q \in \mathcal{Q}} |CQ| \lesssim_d \sum_{Q \in \mathcal{Q}} |Q| \lesssim_d 1.$$

Putting it all together, we obtain

$$|\{|Tf| \geq 1\}| \lesssim_d 1$$

as desired. ∎

Combining this with Marcinkiewicz interpolation and the L^2 boundedness hypothesis gives that all CZOs are of strong-type (p,p) for $1 < p \leq 2$. The self-adjointness of the CZOs *as a class* then gives the same statement for $2 \leq p < \infty$, thus we have proven

Corollary 2.10. *CZOs are of strong type (p,p) for all $1 < p < \infty$, with an operator norm of $O_{d,p}(1)$.*

Remark 2.11. We thus see that the localising properties of the singular kernel K allow one to automatically leverage L^2 boundedness into L^p boundedness for $1 < p < \infty$. In fact there is nothing particularly special about L^2 here (Q3). But in practice L^2 boundedness is easier to obtain than other types of boundedness.

In particular, we see that the Hilbert transform is bounded on $L^p(\mathbf{R})$ for $1 < p < \infty$ (or more precisely, it is bounded on a dense subclass, and thus has a unique bounded extension). This does not by itself directly settle the question of whether the formula (1) makes sense for L^p functions in a norm or pointwise convergence sense (for that, one needs bounds on the truncated Hilbert transforms and the maximal truncated maximal Hilbert transforms), though it does lend confidence that this is the case. However, we can now easily conclude an analogous claim for the inversion formula.

Theorem 2.12. *Let $1 < p < \infty$, and for every $N > 0$ let S_N be the Dirichlet summation operator, defined on Schwartz functions $f \in \mathcal{S}(\mathbf{R})$ by*

$$S_N f(x) := \int_{-N}^{N} \hat{f}(\xi) e^{2\pi i x \xi}\, d\xi.$$

Then S_N extends uniquely to a bounded operator on $L^p(\mathbf{R})$, and $S_N f$ converges in L^p norm to f as $N \to \infty$ for all $f \in L^p(\mathbf{R})$.

Proof From Proposition 1.3 we obtain the identity

$$S_N = \frac{i}{2}(\mathrm{Mod}_N H \mathrm{Mod}_{-N} f - \mathrm{Mod}_{-N} H \mathrm{Mod}_N f)$$

and so the S_N are uniformly bounded on $L^p(\mathbf{R})$ as desired. (Note however that the S_N are not CZOs, since modulation does not preserve the CZO class.) Since $S_N f$ converges in L^p norm to f for Schwartz functions (by dominated convergence), the claim then follows by the usual limiting arguments. ■

Remark 2.13. The corresponding question about pointwise almost-everywhere convergence is the Carleson-Hunt theorem, which is quite difficult. When $p = 1$ an example of Kolmogorov shows that both norm convergence and almost-everywhere convergence can fail. In higher dimensions, the norm convergence for $p = 2$ is true by Plancherel's theorem, but a surprising result of Fefferman shows that it fails for $p \neq 2$. The almost everywhere convergence problem remains open (and is considered very difficult) in higher dimensions.

3. Bounded mean oscillation

Let us now study what a CZO T does on the endpoint space $L^\infty(\mathbf{R}^d)$. At first glance things are rather bad; not only is T unbounded on $L^\infty(\mathbf{R}^d)$ in general, T is not even known yet to be defined on a dense subset of $L^\infty(\mathbf{R}^d)$. Nevertheless, a closer inspection of things will show that T actually does have a fairly reasonable definition on $L^\infty(\mathbf{R}^d)$, up to constants.

Let us try to define Tf for $f \in L^\infty(\mathbf{R}^d)$ directly. A naive implementation of the kernel formula

$$Tf(y) = \int_{\mathbf{R}^d} K(x,y) f(x)\ dx$$

runs into problems both for x near y, and x near infinity, since the decay (4) is not strong enough to give absolute integrability in either case. Localising f to be supported away from y helps with the first problem, but not the second. But suppose we instead looked at the difference $Tf(y) - Tf(y')$ between Tf at two values. Then we formally have

$$Tf(y) - Tf(y') = \int_{\mathbf{R}^d} (K(x,y) - K(x,y')) f(x)\ dx.$$

Now (5) implies that for fixed y, y', the expression $K(x,y) - K(x,y')$ decays like $|x|^{-d-\sigma}$ as $x \to \infty$, and so the integral will be absolutely integrable as long as f vanishes near y and y'. At the other extreme, if f is compactly supported, then it lies in $L^2(\mathbf{R}^d)$ and there is no difficulty defining $Tf(y) - Tf(y')$ (up to the usual measure zero ambiguities).

Motivated by this, let us make the following definitions. We say that two functions $f, g : \mathbf{R}^d \to \mathbf{C}$ are *equivalent modulo a constant* if there exists a complex number C such that $f(x) - g(x) = C$ for almost every x; this is of course an equivalence relation. Given $f \in L^\infty(\mathbf{R}^d)$, we define Tf to be the equivalence class associated to the function

$$Tf(y) := T(f 1_B)(y) + \int_{\mathbf{R}^d \setminus B} (K(x,y) - K(x,0)) f(x)\ dx$$

where $B = B_y$ is any ball which contains both y and 0; it is not hard to see that the right-hand side is well-defined (in particular, the integral is absolutely convergent) and is independent (modulo a constant) of the choice of B, and that T remains linear. One can also check that this definition is consistent (modulo a constant) with the existing definition of T on $L^2(\mathbf{R}^d)$ (or on $L^p(\mathbf{R}^d)$) on the common domain of definition (i.e. $L^2(\mathbf{R}^d) \cap L^\infty(\mathbf{R}^d)$ or $L^p(\mathbf{R}^d) \cap L^\infty(\mathbf{R}^d)$).

Because Tf is only defined modulo a constant, we cannot meaningfully control it in standard norms such as L^p. However, there is a norm which is suitable for this type of function.

Definition 3.1 (Bounded mean oscillation)**.** Let $f : \mathbf{R}^d \to \mathbf{C}$ be a function defined modulo a constant. The BMO (or Bounded Mean Oscillation) norm of f is defined as

$$\|f\|_{\mathrm{BMO}} := \sup_B \fint_B \left| f - \fint_B f \right|$$

where B ranges over all balls. Note that if one shifts f by a constant, the BMO norm is unchanged, so this norm is well-defined for functions defined modulo constants. We denote by $\mathrm{BMO}(\mathbf{R}^d)$ the space of all functions with finite BMO norm, modulo constants.

It is not hard to verify that this is indeed a norm (but note that one has to quotient out by constants before this is the case!), and that $\mathrm{BMO}(\mathbf{R}^d)$ is a Banach space. One can also easily check using the triangle inequality that

$$\fint_B \left| f - \fint_B f \right| \sim \inf_c \fint_B |f - c|$$

where c ranges over constants; thus the statement $\|f\|_{\mathrm{BMO}} \lesssim A$ is equivalent (up to changes in the implied constant) to the assertion that for any ball B there exists a constant c_B such that $\fint_B |f - c_B| \lesssim A$, thus f deviates from c_B by at most A on the average.

It is not hard to see that $\mathrm{BMO}(\mathbf{R}^d)$ contains $L^\infty(\mathbf{R}^d)$, and that $\|f\|_{\mathrm{BMO}(\mathbf{R}^d)} \lesssim \|f\|_{L^\infty(\mathbf{R}^d)}$. But it also contains unbounded functions:

Problem 3.2. Verify that the function $\log |x|$ is in $\mathrm{BMO}(\mathbf{R}^d)$.

The primary significance of BMO for Calderón-Zygmund theory is that it can serve as a substitute for L^∞, on which CZOs are often ill-behaved. Consider for instance the Hilbert transform applied (formally) to the signum function sgn. The resulting function is not bounded, but lies in BMO:

Problem 3.3. Verify that using the previous definitions, $H\mathrm{sgn} = \frac{2}{\pi} \log |x|$ modulo constants.

This phenomenon applies more generally to all CZOs:

Proposition 3.4 (CZOs map L^∞ to BMO)**.** *Let T be a CZO and $f \in L^\infty(\mathbf{R}^d)$. Then $Tf \in \mathrm{BMO}(\mathbf{R}^d)$ and*

$$\|Tf\|_{\mathrm{BMO}(\mathbf{R}^d)} \lesssim_d \|f\|_{L^\infty(\mathbf{R}^d)}.$$

Proof We can normalise $\|f\|_{L^\infty(\mathbf{R}^d)} = 1$. By definition of the BMO norm, it suffices to show that for every ball B there exists a constant c_B such that

$$⨍_B |Tf - c_B| \lesssim_d 1.$$

By translation and dilation symmetry we may take $B = B(0,1)$. We split $Tf = T(f1_{2B}) + T(f1_{\mathbf{R}^d\backslash 2B})$. Since T is bounded on L^2, we have

$$\|T(f1_{2B})\|_{L^2(\mathbf{R}^d)} \lesssim_d \|f1_{2B}\|_{L^2(\mathbf{R}^d)} \lesssim_d 1$$

and hence by Hölder (or Cauchy-Schwarz)

$$⨍_B |T(f1_{2B})| \lesssim_d 1.$$

This deals with the "local" part of Tf. For the "global" part, observe that for $x \in B$ we have

$$T(f1_{\mathbf{R}^d\backslash 2B})(x) = \int_{|y|\geq 2} (K(x,y) - K(0,y))f(y) + c$$

for some arbitrary constant c. But by (6) we have $K(x,y) - K(0,y) = O(|y|^{-d-\sigma})$; since f is bounded by $O(1)$, we thus see that $T(f1_{\mathbf{R}^d\backslash 2B})(x) = c + O(1)$, and so

$$⨍_B |T(f1_{\mathbf{R}^d\backslash 2B}) - c| \lesssim_d 1.$$

Adding the two facts, we obtain the claim. ∎

In order to exploit this proposition, we need to understand to what extent BMO bounds on a function control its size. The basic insight here is that BMO bounds are "almost as good as" L^∞ control for many purposes, despite the logarithmically unbounded members of BMO such as $\log|x|$.

Let us first work informally. Let f be a real-valued BMO function, normalised so that $\|f\|_{\mathrm{BMO}(\mathbf{R}^d)} = 1$, thus

$$\int_B |f - ⨍_B f| \leq |B|$$

for all balls B. By Markov's inequality, this means that

$$|\{x \in B : |f(x) - ⨍_B f| \geq \lambda\}| \leq |B|/\lambda \tag{8}$$

for any $\lambda > 0$. Note that this bound is only non-trivial for $\lambda > 1$, otherwise the trivial bound of $|B|$ is superior. It asserts, for instance, that f only exceeds its average by (say) 10 on at most 10% of any given ball.

It turns out that one can iterate the above fact to give far better estimates in the limit $\lambda \to \infty$ that (8) initially suggests. This is because of a basic principle in harmonic analysis: bad behaviour on a small exceptional set (such as 10% of any given ball) can often be iterated away if we know that the exceptional sets are small *at every scale*. Roughly speaking, this principle allows us to say that f only exceeds its average by 20 on at most 10% of 10% of a given ball (i.e. on at most 1% of the ball), f exceeds its average by 30 on at most 0.1%, and so forth, leading to a bound of the form (8) which tails off exponentially in λ rather than polynomially. More precisely, we have

Proposition 3.5 (John-Nirenberg inequality)**.** *Let $f \in \mathrm{BMO}(\mathbf{R}^d)$. Then for any ball B we have*

$$|\{x \in B : |f(x) - \fint_B f| \geq \lambda \|f\|_{\mathrm{BMO}(\mathbf{R}^d)}\}| \lesssim_d e^{-c_d \lambda}|B|$$

for all $\lambda > 0$ and some constant $c_d > 0$ depending only on d.

Proof We can normalise $\|f\|_{\mathrm{BMO}(\mathbf{R}^d)} = 1$. For each $\lambda > 0$ let $A(\lambda)$ denote the best constant for which we have

$$|\{x \in B : f(x) \geq \fint_B f + \lambda\}| \leq A(\lambda)|B|$$

whenever B is a ball and f is a real-valued function with BMO norm 1. From (8) and the trivial bound of $|B|$ we have the bounds

$$A(\lambda) \leq \min(1, 1/\lambda)$$

and our task is to somehow iterate this to obtain the improved estimate

$$A(\lambda) \lesssim_d e^{-c_d/\lambda}.$$

We shall do this by a variant of the Calderón-Zygmund decomposition. We'll need a moderately large quantity $\lambda_0 > 0$ to be chosen later, and assume that $\lambda > \lambda_0$. Let B be any ball, and let $F := \max((f - \fint_B f)1_{2B}, 0)$. Then we have

$$\|F\|_{L^1(\mathbf{R}^d)} \leq \int_{2B} |f - \fint_B f| \lesssim |2B| \|f\|_{\mathrm{BMO}(\mathbf{R}^d)} \lesssim_d |B|.$$

We consider the set $E := B \cap \{MF > \lambda_0\}$. Let $\varepsilon > 0$ be small, and let K be any compact set in E such that

$$|E \backslash K| \leq \varepsilon.$$

Applying the Vitali covering lemma as in the proof of the Hardy-Littlewood maximal inequality (one can also use the Whitney decomposition or Calderón-Zygmund decomposition here) we can cover

$$K \subset \bigcup_n B_n$$

where B_n are a finite collection of balls intersecting K such that $\fint_{B_n} F \sim_d \lambda_0$ and

$$\sum_n |B_n| \lesssim_d \|F\|_{L^1(\mathbf{R}^d)}/\lambda_0 \leq |B|/\lambda_0.$$

Note that as

$$\fint_{B_n} F \leq \frac{\|F\|_{L^1(\mathbf{R}^d)}}{|B_n|} \lesssim_d \frac{|B|}{|B_n|}$$

we conclude that

$$|B_n| \lesssim_d |B|/\lambda_0.$$

Since B_n intersects K, which is contained in B, we can conclude that $B_n \subset 2B$ if we choose λ_0 large enough.

Observe that F is pointwise bounded almost everywhere by MF (by the Lebesgue differentiation theorem), and so

$$|\{x \in B : |F(x)| \geq \lambda\}| = |\{x \in B : |F(x)| \geq \lambda, MF \geq \lambda_0\}| \leq \sum_n |\{x \in B : |F(x)| \geq \lambda\} \cap B_n| + \varepsilon.$$

On the other hand, on each B_n we see that $f = ⨍_B f + O(F)$, and hence

$$⨍_{B_n} f = ⨍_B f + O(⨍_{B_n} F) = ⨍_B f + O(\lambda_0)$$

and so by definition of A

$$|\{x \in B : |F(x)| \geq \lambda\} \cap B_n| \leq |\{x \in B_n : |f(x) - ⨍_B f| \geq \lambda - O(\lambda_0)\}| \leq A(\lambda - O(\lambda_0))|B_n|.$$

Summing this we conclude

$$|\{x \in B : |F(x)| \geq \lambda\}| \leq A(\lambda - O(\lambda_0)) \sum_n |B_n| + \varepsilon \lesssim_d A(\lambda - O(\lambda_0))|B|/\lambda_0 + \varepsilon.$$

Since ε is arbitrary we can remove it. Taking suprema over all B and f we conclude the recursive inequality

$$A(\lambda) \lesssim_d A(\lambda - O(\lambda_0))/\lambda_0.$$

Taking λ_0 sufficiently large depending on d, and iterating starting from the trivial bound $A(\lambda) \leq 1$, we obtain the claim. ∎

Using the distributional characterisation of the L^p norms, we conclude

Corollary 3.6 (Alternate characterisation of BMO norm)**.** *For any $1 \leq p < \infty$ and all locally integrable f we have*

$$\sup_B (⨍_B |f - ⨍_B f|^p)^{1/p} \sim_{p,d} \sup_B \inf_c (⨍_B |f - c|^p)^{1/p} \sim_{p,d} \|f\|_{\mathrm{BMO}(\mathbf{R}^d)}.$$

We leave the proof as an exercise to the reader.

One consequence of this corollary, when combined with the Calderón-Zygmund decomposition, is that the BMO norm serves as a substitute for the L^∞ norm as far as log-convexity of Hölder norms is concerned:

Lemma 3.7. *Let $0 < p < q < \infty$ and $f \in L^p(\mathbf{R}^d) \cap \mathrm{BMO}(\mathbf{R}^d)$. Then $f \in L^q(\mathbf{R}^d)$ and*

$$\|f\|_{L^q(\mathbf{R}^d)} \lesssim_{p,q,d} \|f\|_{L^p(\mathbf{R}^d)}^{p/q} \|f\|_{\mathrm{BMO}(\mathbf{R}^d)}^{1-p/q}.$$

Proof By rescaling one can normalise $\|f\|_{L^p(\mathbf{R}^d)} = \|f\|_{\mathrm{BMO}(\mathbf{R}^d)} = 1$. Let us understand the size of the distribution function $|\{|f| \geq \lambda\}|$. From the L^p bound we already have

$$|\{|f| \geq \lambda\}| \leq \lambda^{-p}.$$

We can get a better bound for large λ as follows. Applying the L^p version of the Calderón-Zygmund decomposition at level 1, we can find a collection of disjoint dyadic cubes Q with $\sum_Q |Q| \lesssim_d 1$, such that $|f| \leq 1$ almost everywhere outside of

$\bigcup_Q Q$, and such that $⨍_Q |f|^p \lesssim_d 1$ on each cube Q. If we let B be a ball containing Q of comparable diameter, we then easily verify that $⨍_B |f|^p \lesssim_d 1$ also, and thus $⨍_B f = O(1)$. From the John-Nirenberg inequality we thus see that

$$|\{x \in B : |f(x)| \geq \lambda\}| \lesssim_d e^{-c_d \lambda} |B|$$

for λ large enough; summing over B we conclude that

$$|\{|f(x)| \geq \lambda\}| \lesssim_d e^{-c_d \lambda}$$

for large λ. Combining this with the L^p bound for small λ and using the distributional formula for the L^q norm we obtain the claim. ■

Because of this, we can also use BMO as a substitute for L^∞ for real interpolation purposes. For instance, let T be a CZO and $2 < p < \infty$. It is now easy to check that T is restricted type (p,p). Indeed, for any sub-step function f of height H and width W, the boundedness on L^2 gives

$$\|Tf\|_{L^2(\mathbf{R}^d)} \lesssim HW^{1/2}$$

and Proposition 3.4 gives

$$\|Tf\|_{\mathrm{BMO}(\mathbf{R}^d)} \lesssim_d H$$

and so Lemma 3.7 gives

$$\|Tf\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} HW^{1/p}$$

for $2 < p < \infty$, which is the restricted type. Further interpolation then gives strong type. (Note that these arguments are essentially the dual of those which establish weak-type (p,p) for $1 \leq p < 2$; in particular, note how the Calderón-Zygmund decomposition continues to make an appearance.)

4. Multiplier theory

An important subclass of CZOs are given by certain types[5] of *Fourier multipliers*.

Definition 4.1. If $m : \mathbf{R}^d \to \mathbf{C}$ is a tempered distribution we define the *Fourier multiplier* $m(D) : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)^*$ on Schwartz functions $f \in \mathcal{S}(\mathbf{R}^d)$ by the formula

$$\widehat{m(D)f}(\xi) = m(\xi)\hat{f}(\xi).$$

We refer to the function m as the *symbol* of the multiplier $m(D)$.

Note that if f is Schwartz, then so is $\hat{f}$, and that is enough to make $m\hat{f}$ a tempered distribution, and so $m(D)f$ is well-defined as a tempered distribution. If m is bounded, we observe from Plancherel's identity that

$$\|m(D)f\|_{L^2(\mathbf{R}^d)} \leq \|m\|_{L^\infty(\mathbf{R}^d)} \|f\|_{L^2(\mathbf{R}^d)}$$

[5]We also have *spatial multipliers* $m(X) : f(x) \mapsto m(x)f(x)$, but their behaviour is much more elementary to analyse thanks to Hölder's inequality; for instance, a spatial multiplier $m(X)$ is bounded on L^p if and only if the symbol m is in L^∞, and indeed the operator norm of the former is the sup norm of the latter.

and hence the Fourier multiplier extends continuously to all of L^2 in this case. Indeed it is not hard to show that in fact

$$\|m(D)\|_{L^2(\mathbf{R}^d)\to L^2(\mathbf{R}^d)} = \|m\|_{L^\infty(\mathbf{R}^d)}.$$

Examples 4.2. The partial derivative operator $\frac{\partial}{\partial x_j}$ is a Fourier multiplier with symbol $2\pi i\xi_j$. More generally, any constant-coefficient differential operator $P(\nabla)$, where $P : \mathbf{R}^d \to \mathbf{C}$ is a polynomial, has symbol $P(2\pi i\xi)$; thus one can formally view D in $m(D)$ as representing the vector-valued operator $D = \frac{1}{2\pi i}\nabla$. The translation operator Trans_{x_0} is a Fourier multiplier with symbol $e^{-2\pi i x\cdot\xi_0}$, thus we formally have $\mathrm{Trans}_{x_0} = e^{-2\pi i D\cdot x_0} = e^{-x_0\cdot\nabla}$. A convolution operator $T : f \mapsto f * K$, with K a distribution, is a Fourier multiplier with symbol $m = \hat{K}$; conversely, every Fourier multiplier is a convolution operator with the distribution $K = \check{m}$. The Hilbert transform is a Fourier multiplier with symbol $-i\mathrm{sgn}(\xi)$, thus $H = -i\mathrm{sgn}(D)$.

Let $1 \le p \le \infty$. We say that m is an L^p multiplier if $m(D)$ extends to a bounded linear operator on $L^p(\mathbf{R}^d)$, and in that case denote $\|m\|_{M^p}$ to be the L^p operator norm of m on M^p. Thus for instance the M^2 norm is the same as the L^∞ norm. At first glance, it seems one can say quite a lot about these quantities; for instance, since the adjoint of $m(D)$ is easily seen to be $\overline{m}(D)$, which has the same L^p operator norm as $m(D)$, we see that

$$\|m\|_{M^p(\mathbf{R}^d)} = \|\overline{m}\|_{M^p(\mathbf{R}^d)} = \|m\|_{M^{p'}(\mathbf{R}^d)}$$

and thus by Riesz-Thorin interpolation

$$\|m\|_{L^\infty(\mathbf{R}^d)} = \|m\|_{M^2(\mathbf{R}^d)} \le \|m\|_{M^p(\mathbf{R}^d)}$$

and so all L^p multipliers are bounded (though the converse turns out to be false, as we shall see later). The space M^p of L^p multipliers turns out to be a Banach algebra, thus the M^p norm is a Banach space norm and furthermore

$$\|m_1m_2\|_{M^p(\mathbf{R}^d)} \le \|m_1\|_{M^p(\mathbf{R}^d)}\|m_2\|_{M^p(\mathbf{R}^d)}.$$

One can also show (from the symmetries of the Fourier transform) that the M^p norm is invariant under translations, modulations, and L^∞ dilations:

$$\|\mathrm{Trans}_{x_0}m\|_{M^p(\mathbf{R}^d)} = \|\mathrm{Mod}_{\xi_0}m\|_{M^p(\mathbf{R}^d)} = \|\mathrm{Dil}^\infty_\lambda m\|_{M^p(\mathbf{R}^d)} = \|m\|_{M^p(\mathbf{R}^d)}. \tag{9}$$

The dilation invariance can be generalised: given any invertible linear transformation $L : \mathbf{R}^d \to \mathbf{R}^d$ we can check that

$$\|m \circ L\|_{M^p(\mathbf{R}^d)} = \|m\|_{M^p(\mathbf{R}^d)}. \tag{10}$$

Because of the translation invariance, the same is true if L is an affine transformation rather than a linear one.

Thus for instance, by modulating the Hilbert transform we see that

$$\|1_{[\xi_0,+\infty)}\|_{M^p(\mathbf{R})} \lesssim_p 1$$

for $1 < p < \infty$, which by Riemann-Stieltjes integration and Minkowski's inequality shows that

$$\|m\|_{M^p(\mathbf{R})} \lesssim_p \|m\|_{BV(\mathbf{R})}$$

when m has bounded variation, where

$$\|m\|_{BV(\mathbf{R})} := \|m\|_{L^\infty(\mathbf{R})} + \sup_{\xi_1 < \ldots < \xi_N} \sum_{j=1}^{N-1} |m(\xi_{j+1}) - m(\xi_j)|.$$

For M^1 or M^∞, it is also not hard to show that

$$\|m\|_{M^1(\mathbf{R}^d)} = \|m\|_{M^\infty(\mathbf{R}^d)} = \|\check{m}\|_{L^1(\mathbf{R}^d)}$$

whenever the inverse Fourier transform of m is absolutely integrable; with a little more work one can also show that $m(D)$ is unbounded on L^1 and L^∞ when $\check{m}$ does not lie in $L^1(\mathbf{R}^d)$, although this is a little tricky. Thus for $1 \leq p \leq \infty$ we have (by Riesz-Thorin again)

$$\|m\|_{L^\infty(\mathbf{R}^d)} \leq \|m\|_{M^p(\mathbf{R}^d)} \leq \|\check{m}\|_{L^1(\mathbf{R}^d)}.$$

One consequence of this is that if m is a bump function adapted to $L(\Omega)$ for some fixed bounded domain Ω and some arbitrary affine linear transformation L, then m is an M^p multiplier with
$|m\|_{M^p(\mathbf{R}^d)} = O_{d,\Omega}(1)$ for all $1 \leq p \leq \infty$.

Suppose a multiplier m on $\mathbf{R}^d$ does not depend on the final coordinate ξ_d, thus

$$m(\xi_1, \ldots, \xi_d) = \tilde{m}(\xi_1, \ldots, \xi_{d-1})$$

for some $\tilde{m} : \mathbf{R}^{d-1} \to \mathbf{C}$. Then the d-dimensional Fourier multiplier $m(D)$ and the $d-1$-dimensional Fourier multiplier $\tilde{m}(D)$ are related by the identity

$$[m(D)f]_{x_d} = \tilde{m}(D)[f_{x_d}]$$

for all $x_d \in \mathbf{R}$ and Schwartz $f \in \mathcal{S}(\mathbf{R}^d)$, where $f_{x_d} : \mathbf{R}^{d-1} \to \mathbf{C}$ is the slice

$$f_{x_d}(x_1, \ldots, x_{d-1}) := f(x_1, \ldots, x_d).$$

From this and Fubini's theorem one easily verifies that if $\tilde{m}$ is in $M^p(\mathbf{R}^{d-1})$, then m is in $M^p(\mathbf{R}^d)$ with

$$\|m\|_{M^p(\mathbf{R}^d)} = \|\tilde{m}\|_{M^p(\mathbf{R}^d)}. \tag{11}$$

Of course, most multipliers m on $\mathbf{R}^d$ do depend on the final coordinate ξ_d. Nevertheless we can obtain a partial generalisation of (11) for these more general multipliers:

Theorem 4.3 (De Leeuw's theorem)**.** *If $m \in M^p(\mathbf{R}^d)$ is continuous, then $\|m_{\xi_d}\|_{M^p(\mathbf{R}^{d-1})} \leq \|m\|_{M^p(\mathbf{R}^d)}$ for any ξ_d.*

Proof By translation invariance we can take $\xi_d = 0$. For any $\lambda > 0$, let $L_\lambda : \mathbf{R}^d \to \mathbf{R}^d$ be the linear transformation

$$L(\xi_1, \ldots, \xi_d) := (\xi_1, \ldots, \xi_{d-1}, \xi_d/\lambda),$$

then by (10) we have

$$\|m \circ L_\lambda\|_{M^p(\mathbf{R}^d)} = \|m\|_{M^p(\mathbf{R}^d)}.$$

Now we send $\lambda \to \infty$. We observe from the hypothesis that m is continuous that $m \circ L_\lambda \to m \circ L_\infty$ pointwise, where L_∞ is the (non-invertible) projection

$$L_\infty(\xi_1, \dots, \xi_d) = (\xi_1, \dots, \xi_{d-1}, 0).$$

On the other hand, from (11) we have

$$\|m \circ L_\infty\|_{M^p(\mathbf{R}^d)} = \|m_0\|_{M^p(\mathbf{R}^{d-1})}.$$

So to conclude the theorem it suffices to show that

$$\|m \circ L_\infty\|_{M^p(\mathbf{R}^d)} \leq \limsup_{\lambda \to \infty} \|m \circ L_\lambda\|_{M^p(\mathbf{R}^d)}.$$

This can be achieved by a weak convergence[6] argument. If f, g are Schwartz functions, we see from Hölder's inequality that

$$|\langle m \circ L_\lambda(D) f, g\rangle| \leq \|m \circ L_\lambda\|_{M^p(\mathbf{R}^d)} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^{p'}(\mathbf{R}^d)}.$$

Taking limit superior as $\lambda \to \infty$ and using Plancherel and dominated convergence on the left-hand side (using the pointwise convergence, as well as the boundedness of m) we conclude

$$|\langle m \circ L_\infty(D) f, g\rangle| \leq \limsup_{\lambda \to \infty} \|m \circ L_\lambda\|_{M^p(\mathbf{R}^d)} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^{p'}(\mathbf{R}^d)}.$$

Using the converse to Hölder's inequality we obtain the claim. ∎

One consequence of de Leeuw's theorem is that we expect multiplier theory on higher dimensions to be harder than multiplier theory for lower dimensions.

One can view de Leeuw's theorem as a *transference theorem*, transferring a bound on one space ($\mathbf{R}^d$) to another ($\mathbf{R}^{d-1}$), basically by exploiting the fact that there was a projection from $\mathbf{R}^d$ to $\mathbf{R}^{d-1}$. For a result of similar flavour, see Q12.

Despite these encouraging initial results, it appears very difficult to decide in general whether any given multiplier m lies in M^p or not for any $1 < p < \infty$ other than $p = 2$. For instance, the question of determining the precise values of p and δ for which the multiplier $m^\delta(\xi) := \max(1 - |\xi|^2, 0)^\delta$ is an $L^p(\mathbf{R}^d)$ multiplier is not fully resolved in three and higher dimensions; this problem is known as the *Bochner-Riesz conjecture*, which we will not discuss in detail here. Broadly speaking, the difficulty with the Bochner-Riesz conjecture is that the symbol m^δ is singular on a large, curved set (in this case, the sphere), which necessarily leads to some delicate and still incompletely understood geometric issues (most notably the Kakeya conjecture, which we will also not discuss here).

However, one can do a lot better if the multiplier has no singularities, or is only singular at one point (such as the origin). For instance, if m is Schwartz, then $\check{m}$ is definitely in L^1, and so m lies in M^p for all $1 \leq p \leq \infty$. Note from (9) that any dilate, translate, and/or modulation of a Schwartz function is then also in M^p uniformly in the dilation, translation, and modulation parameters.

Now we address multipliers which only have singularities at the origin.

[6] As usual, weak convergence preserves upper bounds, but does not necessarily preserve lower bounds (cf. Fatou's lemma).

Theorem 4.4 (Hörmander-Mikhlin multiplier theorem)**.** *Let $m : \mathbf{R}^d \to \mathbf{C}$ obey the* homogeneous symbol estimates of order 0

$$|\nabla^k m(\xi)| \lesssim |\xi|^{-k} \tag{12}$$

for all $\xi \neq 0$ and $0 \leq k \leq d+2$. (In particular $\|m\|_{L^\infty(\mathbf{R}^d)} \lesssim 1$.) Then m is a CZO, and we have $\|m\|_{M^p(\mathbf{R}^d)} \lesssim_{p,d} 1$ for all $1 < p < \infty$.

Remark 4.5. It is possible to use less regularity than $d+2$ derivatives here, but in practice this type of control is sufficient for many applications. The conditions (12) are essentially asserting that m is smooth and does not oscillate on each annulus $2^j \leq |\xi| \leq 2^{j+1}$, although it essentially allows different annuli to behave independently. For instance, if ψ_j is a bump function adapted to the annulus $2^j \leq |\xi| \leq 2^{j+1}$, and $\varepsilon_j \in \{-1,+1\}$ are arbitrary signs, then $\sum_j \varepsilon_j \psi_j$ obeys the homogeneous symbol estimates of order 0 uniformly in the choice of ε_j.

Proof The multiplier $m(D)$ is already bounded on $L^2(\mathbf{R}^d)$. Formally, we know that $m(D)$ is a convolution operator: $m(D)f = f * \check{m}$, but we run into the formal difficulty that $\check{m}$ need not be locally integrable. To compute things rigorously it is convenient to use the Littlewood-Paley decomposition. Let $\phi : \mathbf{R}^d \to \mathbf{R}^+$ be a fixed bump function which equals one on the ball $B(0,1)$ and is supported on the ball $B(0,2)$, and write

$$\psi_j(\xi) := \phi(\xi/2^j) - \phi(\xi/2^{j-1})$$

for $j \in \mathbf{Z}$. Then we have the partition of unity

$$1 = \sum_{j \in \mathbf{Z}} \psi_j(\xi)$$

for almost every ξ, and hence we have the pointwise convergent sum

$$m = \sum_{j \in \mathbf{Z}} m_j$$

where $m_j := \psi_j m$. Using Plancherel and dominated convergence we conclude that

$$m(D)f = \sum_{j \in \mathbf{Z}} m_j(D)f$$

for all $f \in L^2(\mathbf{R}^d)$, where the sum is conditionally convergent in the L^2 sense. In particular for any $f, g \in L^2(\mathbf{R}^d)$ we have

$$\langle m(D)f, g\rangle = \sum_{j \in \mathbf{Z}} \langle m_j(D)f, g\rangle$$

where the sum is at least conditionally convergent (indeed by Parseval one can show that it is absolutely convergent). Now let us assume that f, g are also bounded and compactly supported with disjoint supports. We can write

$$\langle m_j(D)f, g\rangle = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} K_j(y-x) f(x) g(y)\, dxdy$$

where

$$K_j(x) = \check{m}_j(x) = \int_{\mathbf{R}^d} m_j(\xi) e^{2\pi i x \cdot \xi}\, d\xi.$$

One can justify these formulae via the Fourier inversion formula and Fubini's theorem using the fact that f, g, m_j are bounded with compact support; we omit the details. Now observe that m_j is supported on the annulus $|\xi| \sim 2^j$, and multiple applications of the Leibnitz rule give us the bounds

$$|\nabla^k m_j(\xi)| \lesssim_{k,d} 2^{-jk}$$

for $0 \leq k \leq d+2$. From the $k=0$ case of this bound and the triangle inequality we obtain the bound

$$|K_j(x)| \lesssim_d 2^{dj};$$

on the other hand, by integrating by parts k times (using the identity $e^{2\pi i x\cdot\xi} = (\frac{x}{2\pi i|x|^2} \cdot \nabla_\xi)^k e^{2\pi i x\cdot \xi})$ we obtain the bound

$$|K_j(x)| \lesssim_d 2^{dj} 2^{-jk} |x|^{-k}$$

for $k = 0, \ldots, d+2$, and hence

$$|K_j(x)| \lesssim_d \min(2^{dj}, 2^{dj} 2^{-j(d+2)} |x|^{-(d+2)}) = |x|^{-d} \min((2^j|x|)^d, (2^j|x|)^{-2}).$$

Also, we can compute gradients

$$\nabla K_j(x) = \check{m}_j(x) = \int_{\mathbf{R}^d} 2\pi i \xi m_j(\xi) e^{2\pi i x\cdot\xi}\ d\xi.$$

The additional $2\pi i\xi$ factor essentially just contributes an extra factor of 2^j to the analysis, and we end up with

$$|\nabla K_j(x)| \lesssim_d \min(2^{dj}, 2^{dj} 2^{-j(d+2)} |x|^{-(d+2)}) = |x|^{-d-1} \min((2^j|x|)^{d+1}, (2^j|x|)^{-1}).$$

These estimates are enough to show that away from a neighbourhood of $x = 0$, the series $\sum_{j\in\mathbf{Z}} K_j$ is convergent in C^1_{loc} to a kernel $K : \mathbf{R}^d\backslash\{0\} \to \mathbf{C}$ obeying the bounds

$$|K(x)| \lesssim_d |x|^{-d}; \quad |\nabla K(x)| \lesssim_d |x|^{-d-1} \tag{13}$$

and by Fubini's theorem we have

$$\langle m(D)f, g\rangle = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} K(y-x) f(x) g(y)\ dxdy \tag{14}$$

when f, g are bounded with disjoint support. By a monotone convergence argument (noting that $m(D)$ is bounded on L^2) the same is true when f, g are L^2 with compact support.

The bounds (13) (and the fundamental theorem of calculus) ensure that $K(y-x)$ is a singular kernel, and (14) shows that $m(D)$ is a CZO with kernel K. Thus, by Corollary 2.10, $m(D)$ is of strong-type (p,p) with an operator norm of $O_p(1)$ for all $1 < p < \infty$, and the claim follows. ∎

Note that the Hörmander-Mikhlin multiplier theorem includes the L^p boundedness of the Hilbert transform for $1 < p < \infty$ as a special case, since the multiplier $m(\xi) = i\mathrm{sgn}(\xi)$ certainly obeys the symbol estimates. Another useful instance of the theorem is to conclude that the homogeneous fractional derivative operators $|\nabla|^{it} = |2\pi D|^{it}$, with symbol $m(\xi) = |2\pi\xi|^{it}$ and t real, are also bounded on $L^p(\mathbf{R}^d)$, with an operator norm of $O_d(\langle t\rangle^{d+2})$ (in particular, it grows at most polynomially in t). This fact turns out to be very useful when one wishes to use Stein interpolation involving fractional differentiation operators $|\nabla|^z$. A similar argument also applies to the inhomogeneous fractional derivative[7] operators $\langle\nabla\rangle^{it} = \langle 2\pi D\rangle^{it}$ with symbol $\langle 2\pi\xi\rangle^{it}$.

Another example of multipliers covered by this theorem are the *Riesz transforms* $R_j := iD_j/|D| = \partial_{x_j}/|\nabla|$ for $j = 1, \ldots, d$, with symbol $i\xi_j/|\xi|$. These symbols are easily verified to obey the homogeneous symbol estimates and are thus in $M^p(\mathbf{R}^d)$ with norm $O_d(1)$ for all $1 < p < \infty$. The significance of the Riesz transforms is that they connect general second order derivatives to the Laplacian; indeed from Fourier analysis we quickly see that

$$\frac{\partial^2}{\partial_j \partial_k} f = -R_j R_k \Delta f$$

for all $1 \leq j, k \leq d$ and Schwartz f. In particular, since $R_j R_k$ is bounded on L^p, we obtain the *elliptic regularity estimate*

$$\|\nabla^2 f\|_{L^p(\mathbf{R}^d)} \lesssim_{d,p} \|\Delta f\|_{L^p(\mathbf{R}^d)}$$

for all Schwartz f and all $1 < p < \infty$. In a similar vein, we can show that

$$\|\nabla^k f\|_{L^p(\mathbf{R}^d)} \lesssim_{d,p,k} \||\nabla|^k f\|_{L^p(\mathbf{R}^d)}$$

for any $k \geq 0$, Schwartz f, and $1 < p < \infty$, where $|\nabla|^k = |2\pi D|^k$ is the Fourier multiplier with symbol $|2\pi\xi|^k$.

More generally, any multiplier which is homogeneous of degree zero (thus $m(\lambda\xi) = m(\xi)$ for all $\lambda > 0$), and smooth on the sphere S^{d-1}, obeys the homogeneous symbol estimates and is thus in M^p for all $1 < p < \infty$.

The Fourier multipliers covered by the Hörmander-Mikhlin multiplier theorem are essentially special cases of the *pseudodifferential operators of order* 0, which we shall discuss in later notes.

5. Vector-valued analogues

In the above discussion we considered only scalar-valued (complex-valued) functions, and scalar-valued operators with scalar kernels K. However, it is possible to generalise basically all of the above discussion to vector-valued cases, in which the functions take values in a Hilbert space, and the kernel takes values in the bounded linear operators from one Hilbert space to another. We rapidly present these generalisations here; as it turns out only some notational changes need to be made. We will then give an important application of this generalisation, namely the *Littlewood-Paley inequality.*

Let H be a Hilbert space. For $p < \infty$, we let $L^p(X \to H)$ be the space of measurable functions $f : X \to H$ such that

$$\int_X \|f(x)\|_H^p \, d\mu(x) < \infty,$$

[7] The factor of 2π is of course irrelevant here and the same results hold if this factor is dropped. Indeed, in PDE it is customary to drop the 2π from the definition of the Fourier transform so as not to see these factors then arise from differential operators.

where of course we identify functions which agree almost everywhere as before. We also define $L^\infty(X \to H)$ in the usual manner. One can (rather laboriously) rederive most of the scalar L^p theory to this vector-valued case (mostly by replacing absolute values with H norms, and any scalar form $f\overline{g}$ with the inner product $\langle f, g\rangle_H$). For instance, one can verify the duality relationship

$$\|f\|_{L^p(X\to H)} = \sup\{\int_X \langle f(x), g(x)\rangle_H \; d\mu(x) : \|g\|_{L^{p'}(X\to H)} \leq 1\}$$

whenever $1 \leq p \leq \infty$ and $f \in L^p(X \to H)$. Also, the real interpolation theory continues to be valid in this setting (the basic reason being that the layer-cake decompositions, which depend only on the magnitude of the underlying function, continue to be valid).

Observe that if $f : X \to H$ is absolutely integrable on some set E, thus $\int_E \|f(x)\|_H \; d\mu(x) < \infty$, then one can define the vector-valued integral $\int_E f(x) \; d\mu(x) \in H$ in several ways, for instance by using duality

$$\langle \int_E f(x) \; d\mu(x), g\rangle_H := \int_E \langle f(x), g\rangle_H \; d\mu(x).$$

In particular, for locally integrable vector-valued functions f, we can compute averages $⨍_B f$ for balls or cubes B.

Let $B(H \to H')$ denote the bounded linear operators of H to H', with the usual operator norm.

Definition 5.1 (Vector-valued singular kernels)**.** Let H, H' be Hilbert spaces. A *singular kernel from* H *to* H' is a function $K : \mathbf{R}^d \times \mathbf{R}^d \to B(H \to H')$ (defined away from the diagonal $x = y$) which obeys the decay estimate

$$\|K(x, y)\|_{B(H\to H')} \lesssim |x - y|^{-d} \tag{15}$$

for $x \neq y$ and the Hölder-type regularity estimates

$$\|K(x, y') - K(x, y)\|_{B(H\to H')} = O(\frac{|y - y'|^\sigma}{|x - y|^{d+\sigma}}) \text{ whenever } |y - y'| \leq \frac{1}{2}|x - y| \tag{16}$$

and

$$\|K(x', y) - K(x, y)\|_{B(H\to H')} = O(\frac{|x - x'|^\sigma}{|x - y|^{d+\sigma}}) \text{ whenever } |x - x'| \leq \frac{1}{2}|x - y| \tag{17}$$

for some *Hölder exponent* $0 < \sigma \leq 1$.

Definition 5.2 (Vector-valued Calderón-Zygmund operators)**.** A *Calderón-Zygmund operator* (CZO) from H to H' is a linear operator $T : L^2(\mathbf{R}^d \to H) \to L^2(\mathbf{R}^d \to H')$ which is bounded,

$$\|Tf\|_{L^2(\mathbf{R}^d\to H')} \lesssim \|f\|_{L^2(\mathbf{R}^d\to H)},$$

and such that there exists a singular kernel K from H to H' for which

$$Tf(y) = \int_{\mathbf{R}^d} K(x, y)f(x) \; dx$$

whenever $f \in L^2(\mathbf{R}^d \to H)$ is compactly supported, and y lies outside of the support of f. Note that the integral here is an absolutely convergent Hilbert space-valued integral.

One can then repeat the derivation of Corollary 2.10 more or less verbatim to show that vector-valued Calderón-Zygmund operators continue to be bounded from $L^p(\mathbf{R}^d \to H)$ to $L^p(\mathbf{R}^d \to H')$ with bound $O_{d,p}(1)$. Note that the constants never depend on the dimension of H and H', and indeed this dimension may be infinite.

An important application of vector-valued Calderón-Zygmund theory is the *Littlewood-Paley inequality.*

Proposition 5.3 (Upper Littlewood-Paley inequality)**.** *For each integer j, let $\psi_j : \mathbf{R}^d \to \mathbf{C}$ be a bump function adapted to an annulus $\{\xi : |\xi| \sim 2^j\}$. Then for any $1 < p < \infty$ and $f \in L^p(\mathbf{R}^d)$ we have*

$$\|(\sum_{j \in \mathbf{Z}} |\psi_j(D) f|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}$$

and for any $f_j \in L^p(\mathbf{R}^d)$ for $j \in \mathbf{Z}$ we have

$$\|\sum_{j \in \mathbf{Z}} \psi_j(D) f_j\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|(\sum_{j \in \mathbf{Z}} |f_j|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \tag{18}$$

where the sum in the left is conditionally convergent in L^p.

Proof By monotone convergence it suffices to prove the claim assuming that only finitely many of the ψ_j are non-zero. (This step is not strictly necessary, but it lets us avoid unnecessary technicalities in what follows.) It suffices to prove the first claim, as the second follows from a standard duality argument.

We consider the vector-valued operator $T : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d; l^2(\mathbf{Z}))$ defined by

$$Tf := (\psi_j(D) f)_{j \in \mathbf{Z}},$$

thus

$$\|Tf(x)\|_{l^2(\mathbf{Z})} = (\sum_{j \in \mathbf{Z}} |\psi_j(D) f(x)|^2)^{1/2}.$$

Since $\sum_j |\psi_j|^2 = O(1)$, we quickly see from Plancherel's theorem that T is strong-type $(2,2)$ with operator norm $O(1)$. Next, we observe that T has a vector-valued kernel

$$K(x,y) = (\check{\psi}_j(y - x))_{j \in \mathbf{Z}}$$

where we identify such vectors with operators from $\mathbf{C}$ to $l^2(\mathbf{Z})$ in the obvious manner. But as in the proof of the Hörmander-Mikhlin multiplier theorem we have

$$\check{\psi}_j(x) = O(\frac{1}{|x|^d} \min((2^j|x|)^d, (2^j|x|)^{-2}))$$

and

$$\nabla \check{\psi}_j(x) = O(\frac{1}{|x|^{d+1}} \min((2^j|x|)^{d+1}, (2^j|x|)^{-1}))$$

from which we easily verify (15), (16), (17). The claim then follows from the vector-valued analogue of Corollary 2.10. ■

Corollary 5.4 (Two-sided Littlewood-Paley inequality)**.** *For each integer j, let $\psi_j : \mathbf{R}^d \to \mathbf{C}$ be a bump function adapted to an annulus $\{\xi : |\xi| \sim 2^j\}$. Suppose also that $\sum_j |\psi_j|^2 \sim 1$. Then for any $1 < p < \infty$ and $f \in L^p(\mathbf{R}^d)$ we have*

$$\|(\sum_{j \in \mathbf{Z}} |\psi_j(D) f|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \sim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}.$$

Proof By density (and the preceding proposition) it suffices to prove this when f is Schwartz.

From the preceding proposition we already have the upper bound; the task is to establish the lower bound. Let $\tilde{\psi}_j$ be bump functions adapted to a slightly wider annulus than ψ_j, which equal 1 on the support of ψ_j. Writing

$$\varphi_j := \tilde{\psi}_j \frac{\psi_j}{\sum_j |\psi_j|^2}$$

we verify that φ_j is also adapted to the same annulus as ψ_j, and that we have the partition of unity $1 = \sum_j \varphi_j \psi_j$. This leads to the identity

$$f = \sum_j \varphi_j(D) \psi_j(D) f$$

when f is Schwartz, and hence by (18) we obtain

$$\|f\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|(\sum_j |\psi_j(D) f|^2)^{1/2}\|_{L^p(\mathbf{R}^d)}$$

as desired. ■

The quantity $(\sum_{j \in \mathbf{Z}} |\psi_j(D) f|^2)^{1/2}$ is known as a *Littlewood-Paley square function* of f. One can view $\psi_j(D) f$ as the component of f whose frequencies have magnitude $\sim 2^j$. The point of the inequality is that even though each component $\psi_j(D) f$ may individually oscillate, no cancellation occurs between the components in the square function. This makes the square function more amenable than the original function f when it comes to analyse operators (such as differential or pseudodifferential operators) which treat each frequency component in a different way.

5.5. **An alternate approach via Khinchine's inequality.** One can avoid dealing with vector-valued CZOs by instead using randomised signs as a substitute. The key tool is the following inequality, which roughly speaking asserts that a random walk with steps $x_1, \ldots, x_N$ is extremely likely to have magnitude about $(\sum_{j=1}^N |x_j|^2)^{1/2}$, and can be viewed as a variant of the law of large numbers:

Lemma 5.6 (Khinchine's inequality for scalars)**.** *Let $x_1, \ldots, x_N$ be complex numbers, and let $\epsilon_1, \ldots, \epsilon_N \in \{-1, 1\}$ be independent random signs, drawn from $\{-1, 1\}$ with the uniform distribution. Then for any $0 < p < \infty$*

$$(\mathbf{E}|\sum_{j=1}^N \epsilon_j x_j|^p)^{1/p} \sim_p (\sum_{j=1}^N |x_j|^2)^{1/2}.$$

Proof By the triangle inequality it suffices to prove this when the x_j are real. When $p = 2$ we observe that

$$\mathbf{E}|\sum_{j=1}^N \epsilon_j x_j|^2 = \mathbf{E} \sum_{1 \leq j,j' \leq N} \epsilon_j \epsilon_{j'} x_j x_{j'} = \sum_{j=1}^N x_j^2$$

since $\epsilon_j \epsilon_{j'}$ has expectation 0 when $j \neq j'$ and 1 when $j = j'$. This proves the claim for $p = 2$. By Hölder's inequality this gives the upper bound for $p \leq 2$ and the lower bound for $p \geq 2$. Hölder also shows us that the lower bound for $p \leq 2$ will follow from the upper bound for $p \geq 2$, so it remains to prove the upper bound for $p \geq 2$.

We will use the "exponential moment method". We normalise $\sum_{j=1}^N |x_j|^2 = 1$. Now consider the expression

$$\begin{aligned}\mathbf{E} \exp \sum_{j=1}^N \epsilon_j x_j &= \mathbf{E} \prod_{j=1}^N \exp \epsilon_j x_j \\ &= \prod_{j=1}^N \mathbf{E} \exp \epsilon_j x_j \\ &= \prod_{j=1}^N \cosh x_j.\end{aligned}$$

A comparison of Taylor series reveals the inequality

$$\cosh x_j \leq \exp(x_j^2/2)$$

and hence

$$\mathbf{E} \exp \sum_{j=1}^N \epsilon_j x_j \leq \exp \prod_{j=1}^N x_j^2/2 \lesssim 1.$$

By Markov's inequality we conclude

$$\mathbf{P}(\sum_{j=1}^N \epsilon_j x_j \geq \lambda) \lesssim e^{-\lambda}$$

and hence (by the symmetry of the ϵ_j)

$$\mathbf{P}(|\sum_{j=1}^N \epsilon_j x_j| \geq \lambda) \lesssim e^{-\lambda}.$$

The claim then follows from the distributional representation of the L^p norm. ∎

Remark 5.7. It is instructive to prove Khinchine's inequality directly for even exponents such as $p = 4$.

Corollary 5.8 (Khinchine's inequality for functions)**.** *Let $f_1, \ldots, f_N \in L^p(X)$ for some $0 < p < \infty$, and let $\epsilon_1, \ldots, \epsilon_N \in \{-1, 1\}$ be independent random signs, drawn from $\{-1, 1\}$ with the uniform distribution. Then for any $0 < p < \infty$*

$$(\mathbf{E}\|\sum_{j=1}^N \epsilon_j f_j\|_{L^p(X)}^p)^{1/p} \sim_p \|(\sum_{j=1}^N |f_j|^2)^{1/2}\|_{L^p(X)}.$$

Proof Apply the above inequality to the sequence $f_1(x), \ldots, f_N(x)$ for each $x \in X$ and then take L^p norms of both sides. ∎

This gives the following alternate proof of Proposition 5.3. Once again we may restrict to only finitely many of the ψ_j being non-zero. Observe that for arbitrary choices of signs $\epsilon_j \in \{-1, 1\}$, that the sum $\sum_j \epsilon_j \psi_j$ obeys the homogeneous symbol estimates of order 0. Thus by the Hörmander-Mikhlin theorem we have

$$\| \sum_j \epsilon_j \psi_j(D) f\|_{L^p(\mathbf{R}^d)}^p \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}^p$$

for $1 < p < \infty$. Taking expectations and using Khinchine's inequality we obtain the claim.

6. Exercises

- Q1. Let $T : \mathcal{S}(\mathbf{R}) \to C^0(\mathbf{R})$ be a continuous linear operator which maps Schwartz functions to bounded continuous, and which commutes with translations and dilations as in (2). Show that T is a linear multiple of the Hilbert transform and the identity. (Hint: first establish the existence of a distribution $K \in \mathcal{S}(\mathbf{R})^*$ such that $Tf(x) = \langle K, \mathrm{Trans}_{-x} f\rangle$ for all $f \in \mathcal{S}(\mathbf{R})$, and such that K is homogeneous of degree -1. Restrict the action of K to Schwartz functions on $\mathbf{R}^+$ or $\mathbf{R}^-$ and show that K is a constant multiple of $1/x$ on these functions.) (The claim is also true if we relax $C^0(\mathbf{R})$ to $\mathcal{S}(\mathbf{R})^*$, the space of distributions, but this is a little trickier to prove.)
- Q2. Let T and T' be two CZOs with the same kernel K. Show that there exists a bounded function $b \in L^\infty(\mathbf{R}^d)$ such that $Tf = T'f + bf$ for all $f \in L^2(\mathbf{R}^d)$. (Hint: by subtraction we can assume $T' = 0$ and $K = 0$. Now observe that $E \mapsto \langle T1_E, 1_E\rangle$ is an absolutely continuous measure, and use the Radon-Nikodym theorem.)
- Q3. Show that if we redefine the notion of a CZO by weakening the hypothesis of $L^2(\mathbf{R}^d)$ boundedness, to merely being of restricted weak-type (p_0, p_0) for some $1 < p_0 < \infty$, then Corollary 2.10 remains unchanged (but with an operator norm of $O_{p_0}(1)$ instead of $O(1)$. In particular this apparent expansion of the notion of a CZO does not in fact increase the class of CZOs (up to changes in the implied constants).
- Q4. (Lacunary exponential sums lie in BMO) Let $(\xi_n)_{n\in\mathbf{Z}}$ be a sequence of non-zero frequencies in $\mathbf{R}$ which are *lacunary* in the sense that $|\xi_n| \sim 2^n$ for all n. Let $(c_n)_{n\in\mathbf{Z}}$ be a sequence of complex numbers which obeys the l^2 bound $\sum_n |c_n|^2 \lesssim 1$. To avoid issues of convergence let us assume that only finitely many of the c_n are non-zero. Let $f : \mathbf{R} \to \mathbf{C}$ be the exponential sum
$$f(x) := \sum_n c_n e^{2\pi i \xi_n x}.$$
If I is an interval in $\mathbf{R}$, and $\phi_I : \mathbf{R} \to \mathbf{R}$ is a bump function adapted to I, establish the local L^2 estimate
$$\int_{\mathbf{R}} \phi_I(x)|f(x) - c_I|^2\, dx \lesssim |I|$$

for some constant c_I. (Hints: split the n summation into the high frequencies where $2^n \geq |I|^{-1}$, and the low frequencies where $2^n < |I|^{-1}$. For the high frequencies, forget about c_I and expand out the sum, using some good estimate on the Fourier transform of ϕ_I. For the low frequencies, choose c_I in order to obtain some cancellation and then use the triangle inequality.) Conclude that

$$\|f\|_{\mathrm{BMO}(\mathbf{R})} \lesssim 1.$$

- Q5: Show that Corollary 3.6 is in fact true for all $0 < p < \infty$.
- Q6 (Dyadic BMO) Given any $\Omega \subset \mathbf{R}^d$ and any locally integrable $f : \Omega \to \mathbf{C}$, define the dyadic BMO norm $\|f\|_{\mathrm{BMO}_\Delta(\Omega)}$ of f as

$$\|f\|_{\mathrm{BMO}_\Delta(\Omega)} = \sup_{Q \subset \Omega} ⨍_Q |f - ⨍_Q f|$$

where Q ranges over all dyadic cubes inside Ω. Show that we have the John-Nirenberg inequality

$$|\{f \in Q : |f - ⨍_Q f| \geq \lambda\}| \lesssim e^{-c\lambda/\|f\|_{\mathrm{BMO}_\Delta(\Omega)}} |Q|$$

for all $\lambda > 0$ and all $Q \subset \mathbf{R}^d$, and in particular

$$\|f\|_{\mathrm{BMO}_\Delta(\Omega)} \sim_p \sup_{Q \subset \Omega} (⨍_Q |f - ⨍_Q f|^p)^{1/p}$$

for all $1 \leq p < \infty$. (As in Q5, this extends to $0 < p < 1$ as well.)
- Q7 (BMO and dyadic BMO) For any locally integrable f in $\mathbf{R}^d$, show that

$$\|f\|_{\mathrm{BMO}_\Delta(\mathbf{R}^d)} \lesssim_d \|f\|_{\mathrm{BMO}(\mathbf{R}^d)}$$

but that the converse inequality is not true. On the other hand, if we let $\mathcal{D}_{d,\alpha}$ be the shifted dyadic meshes from the previous week's notes, and set $\mathrm{BMO}_{\Delta,\alpha}$ to be the associated shifted dyadic BMO norms, show that

$$\|f\|_{\mathrm{BMO}(\mathbf{R}^d)} \sim_d \sum_\alpha \|f\|_{\mathrm{BMO}_{\Delta,\alpha}(\mathbf{R}^d)}.$$

- Q8 (Sharp function, I) For any locally integrable function $f : \mathbf{R}^d \to \mathbf{C}$, define the function $f^\sharp : \mathbf{R}^d \to \mathbf{R}^+$ as

$$f^\sharp(x) := \sup_{B \ni x} ⨍_B |f - ⨍_B f|$$

where the supremum ranges over all balls containing x. In particular we see that

$$\|f^\sharp\|_{L^\infty(\mathbf{R}^d)} = \|f\|_{\mathrm{BMO}(\mathbf{R}^d)}.$$

Show that we have the pointwise bound

$$|f^\sharp| \lesssim_d Mf.$$

- Q9 (Sharp function, II) Let f be locally integrable, let $\varepsilon, \lambda > 0$, and let B be any ball on which $Mf(x) \lesssim \lambda$ for at least one $x \in B$, and such that $f^\sharp(y) \leq \varepsilon\lambda$ for at least one $y \in B$. Use the Hardy-Littlewood maximal inequality to show that for any sufficiently large $K \gg 1$, that

$$|\{x \in B : Mf \geq K\lambda\}| \lesssim_d \frac{\varepsilon}{K}|B|.$$

From this and a Vitali covering lemma or Calderón-Zygmund decomposition argument, conclude the *good λ inequality*

$$|\{Mf \geq K\lambda \text{ and } f^\sharp \leq \varepsilon\lambda\}| \lesssim_d \frac{\varepsilon}{K}|\{Mf \geq \lambda\}|$$

and then (by taking K large, and ε very small) show that

$$\|f\|_{L^p(\mathbf{R}^d)} \sim_{p,d} \|Mf\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f^\sharp\|_{L^p(\mathbf{R}^d)}$$

for all $1 < p < \infty$ and $f \in L^p(\mathbf{R}^d)$. Use this to give an alternate proof of Lemma 3.7.

- Q10 Show that if T is a bounded linear operator on $L^2(\mathbf{R}^d)$ which commutes with translations, then it is a Fourier multiplier operator with bounded symbol.
- Q11 Let $m : \mathbf{R}^d \to \mathbf{C}$ and $m' : \mathbf{R}^{d'} \to \mathbf{C}$ be L^p multipliers. Show that the tensor product $m \otimes m' : \mathbf{R}^{d+d'} \to \mathbf{C}$ is also an L^p multiplier with

$$\|m \otimes m'\|_{M^p(\mathbf{R}^{d+d'})} = \|m\|_{M^p(\mathbf{R}^d)}\|m'\|_{M^p(\mathbf{R}^{d'})}.$$

(Hint: it may help to first establish the cases when $m \equiv 1$ or $m' \equiv 1$.)

- Q12 (Transference theorem for $\mathbf{Z}^d$) If $m : \mathbf{Z}^d \to \mathbf{C}$ is bounded, define the multiplier $m(D)$ on $L^2(\mathbf{T}^d)$ by the formula

$$\widehat{m(D)f}(\xi) := m(\xi)\hat{f}(\xi) \text{ for } \xi \in \mathbf{Z}^d$$

where we use the Fourier transform on the torus $\mathbf{T}^d := \mathbf{R}^d/\mathbf{Z}^d$,

$$\hat{f}(\xi) := \int_{\mathbf{T}^d} e^{-2\pi i x\cdot\xi} f(x)\, dx.$$

We then define $\|m\|_{M^p(\mathbf{Z}^d)}$ to be the $L^p(\mathbf{T}^d)$ operator norm of $m(D)$.

Let $m : \mathbf{R}^d \to \mathbf{C}$ be continuous and an $M^p(\mathbf{R}^d)$ multiplier for some $1 \leq p \leq \infty$. Show that $\|m|_{\mathbf{Z}^d}\|_{M^p(\mathbf{Z}^d)} \leq \|m\|_{M^p(\mathbf{R}^d)}$. (Hint: take functions $f, g \in C^\infty(\mathbf{T}^d)$, extend them periodically to $\mathbf{R}^d$, multiply them by some cutoff to 1 (such as $e^{-\varepsilon|x|^2}$), apply $m(D)$ on $\mathbf{R}^d$, and use some appropriate limiting argument (and possibly a Poisson summation formula) to get down to $\mathbf{R}^d$.)

- Q13 (Continuous Littlewood-Paley inequality) For every $t > 0$, let $\psi_t : \mathbf{R}^d \to \mathbf{C}$ be a function obeying the estimates

$$|\nabla^j \psi_t(\xi)| \lesssim_d t^{-j} \min((|\xi|/2^j)^{-\varepsilon}, (|\xi|/2^j)^{\varepsilon})$$

for all $0 \leq j \leq d+2$, all $\xi \in \mathbf{R}^d$, and some $\varepsilon > 0$ (independent of t). Conclude that

$$\|(\int_0^\infty |\psi_t(D)f|^2 \frac{dt}{t})^{1/2}\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}$$

for $f \in L^p(\mathbf{R}^d)$. If we also have the pointwise bound $\int_0^\infty |\psi_t|^2 \frac{dt}{t} \sim 1$, improve this bound to

$$\|(\int_0^\infty |\psi_t(D)f|^2 \frac{dt}{t})^{1/2}\|_{L^p(\mathbf{R}^d)} \sim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}.$$

- Q14. (Chernoff's inequality) If $x_1, \dots, x_N$ are complex numbers of magnitude at most one, establish the inequality
$$\mathbf{P}(|\sum_{j=1}^N \epsilon_j x_j| \geq \lambda(\sum_{j=1}^N |x_j|^2)^{1/2}) \lesssim e^{-\lambda^2/4}$$
for all $0 < \lambda \leq 2(\sum_{j=1}^N |x_j|^2)^{1/2}$. (Hint: look at $\exp(t \sum_{j=1}^N \epsilon_j x_j)$ for a real parameter t.)
- Q15. (Marcinkiewicz-Zygmund vector extension) Let T be a linear operator on $L^p(\mathbf{R}^d)$ with operator norm 1. Show that for any functions $f_1, f_2, \dots \in L^p(\mathbf{R}^d)$ we have
$$\|(\sum_{n=1}^\infty |Tf_n|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \lesssim_p \|(\sum_{n=1}^\infty |f_n|^2)^{1/2}\|_{L^p(\mathbf{R}^d)}.$$
(Hint: reduce to finite summation and use Khinchine's inequality.)
- Q16. (BMO Littlewood-Paley inequality) For each integer j, let $f_j \in L^\infty(\mathbf{R}^d)$; assume that all but finitely many of the f_j are non-zero. Let ψ_j be as in Proposition 5.3. Show that
$$\|\sum_j \psi_j(D) f_j\|_{\mathrm{BMO}(\mathbf{R}^d)} \lesssim_d \|(\sum_j |f_j|^2)^{1/2}\|_{L^\infty(\mathbf{R}^d)}.$$
(It is possible to remove the hypothesis that finitely many of the f_j are non-zero, but then some care must be taken to define $\sum_j \psi_j(D) f_j$; in particular it will only be defined modulo a constant.)
- Q17. (Zygmund's inequality) For each positive integer n, let ξ_n be an integer with $|\xi_n| \sim 2^n$, and let c_n be a complex number with $\sum_n |c_n|^2$ finite. Show that
$$\|c_n e^{2\pi i \xi_n x}\|_{e^{L^2}(\mathbf{T})} \sim (\sum_n |c_n|^2)^{1/2}$$
and in particular that
$$\|c_n e^{2\pi i \xi_n x}\|_{L^p(\mathbf{T})} \sim_p (\sum_n |c_n|^2)^{1/2}$$
for all $2 \leq p < \infty$. Dually, for any $f \in L \log^{1/2} L(\mathbf{T})$, show that
$$(\sum_n |\hat f(\xi_n)|^2)^{1/2} \lesssim \|f\|_{L \log^{1/2} L(\mathbf{T})}$$
and in particular that
$$(\sum_n |\hat f(\xi_n)|^2)^{1/2} \lesssim_p \|f\|_{L^p(\mathbf{T})}$$
for all $1 < p \leq \infty$.
- Q18. (John-Nirenberg via Bellman functions) Let $V : [0,1] \to \mathbf{R}$ be the function $V(x) := 10 - (x-2)^2$. Use a Bellman function approach to establish the John-Nirenberg type inequality
$$|\{x \in I : f(x) - ⨍_I f \geq \lambda\}| \leq C_0 |I| V(\frac{\|f - ⨍_I f\|_{L^2(I)}^2}{|I|}) e^{-\varepsilon\lambda}$$
for some absolute constants $C_0, \varepsilon > 0$, all intervals I, and all functions $f : I \to \mathbf{R}$ with $\|f\|_{\mathrm{BMO}_\Delta(I)} \leq 1$.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: tao@math.ucla.edu