Notes on Fourier analysis and singular integrals

Sam Punshon-Smith

April 11, 2023

The purpose of these notes is to introduce some basics of Fourier analysis and its application to the analysis of singular integrals and regularity analysis. These notes are meant as supplementary material in the PDE II class at Tulane. The notes are rough, please excuse any typos or mistakes. In writing these notes, I have interpolated from many sources as well material in my head. Useful references for this material and much more advanced material are Grafakos [1], Hörmander [2] and Muscalau-Schlag [3]

1 The Fourier transform

The Fourier transform is a powerful tool that decomposes a function into its fundamental frequencies. It was originally discovered and used to provide explicit solutions to the heat equation and other linear PDEs. It has since become a powerful tool for analyzing regularity and fine properties of functions. In these notes, we will primarily focus on the space $\mathbb{R}^n$. However, there are also many analogous applications when considering the Fourier transform on the periodic domain $\mathbb{T}^n$.

Definition 1.1. Let $f \in L^1(\mathbb{R}^n)$, the Fourier transform of f is defined by

$$\mathcal{F}(f)(\xi) = \hat{f}(\xi) = f^\wedge(\xi) := \int_{\mathbb{R}^n} e^{-\mathrm{i}\xi\cdot x} f(x)\mathrm{d}x.$$

Remark 1.2. This is not the only definition of the Fourier transform. The one I am using is commonly (but not always) used by analysts (for instance Hörmander). However other common definitions are

$$\int_{\mathbb{R}^n} e^{-2\pi\mathrm{i}\xi\cdot x} f(x)\mathrm{d}x \quad \text{and} \quad \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n} e^{-\mathrm{i}\xi\cdot x} f(x)\mathrm{d}x,$$

which have the added benefit that they are unitary when extended to L^2.

It is useful to note that a simple application of Hölder's inequality gives the following bound

$$\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$$

showing that the Fourier transform is a bounded operator from L^1 to L^∞. However, the range of this operator in L^∞ is very poorly understood and in general cannot be inverted on a general L^∞ function.

1.1 Schwartz space

It is very convenient to introduce a "core" of functions where the Fourier transform can be easily inverted on and which is dense in most space we might wise to define the Fourier transform on (i.e. L^p for $p \in [1, \infty)$).

Definition 1.3 (Schwartz space)**.** A function $f : \mathbb{R}^n \to \mathbb{C}$, $f \in C^\infty$ is called a Schwartz function if for each multi-index $\alpha \in \mathbb{N}^n$, $\beta \in \mathbb{N}^n$

$$x^\alpha D^\beta f := x_1^{\alpha_1} \dots x_n^{\alpha_n} \partial_{x_1}^{\beta_n} \dots \partial_{x_n}^{\beta_n} f \in L^\infty(\mathbb{R}^n).$$

The collection of all such function we call the *Schwartz space* (or space of rapidly decreasing functions) and denote by $\mathcal{S}(\mathbb{R}^n)$. We say that a sequence of functions $\{f_n\} \subset \mathcal{S}(\mathbb{R}^n)$ converges to f in $\mathcal{S}(\mathbb{R}^n)$ if

$$\|x^\alpha D^\beta (f_n - f)\|_{L^\infty} \to 0 \quad \forall \alpha, \beta \in \mathbb{N}^n.$$

Note that any $f \in \mathcal{S}(\mathbb{R}^n)$ has the property that f and all of its derivatives decay faster than any power of x and therefore we see that $f \in W^{k,\infty} \cap W^{k,1}$ for any $k \geq 0$. Moreover its easily follows that $C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$ and therefore $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^P(\mathbb{R}^n)$ for $p \in [1, \infty)$

Remark 1.4. The space $\mathcal{S}(\mathbb{R}^n)$ is clearly a linear space of functions, however it is not a Banach space, in that it does not have a norm that can be used to define its topology. Instead $\mathcal{S}(\mathbb{R}^n)$ defines a Frèchét space, namely a locally convex topological space with a family of semi norms $\{\rho_\alpha\}$ that separate points (i.e. $\rho_\alpha(f) = 0, \quad \forall \alpha$ implies that $f = 0$). In the case of $\mathcal{S}(\mathbb{R}^n)$, the seminorms are given by

$$\rho_{\alpha,\beta}(f) := \|x^\alpha D^\beta f\|_{L^\infty}, \quad \alpha, \beta \in \mathbb{N}^n.$$

The following useful properties are easy to prove for the Fourier transform on $\mathcal{S}(\mathbb{R}^n)$ using basic change of variables or integration by parts. Their proof is left as an exercise.

Proposition 1.5. *Let $f \in \mathcal{S}(\mathbb{R}^n)$ then the following properties hold*

1. *Let $g(x) = f(x - y)$, then $\hat{g}(\xi) = e^{-\mathrm{i}y\cdot\xi} \hat{f}(\xi)$*
2. *Let $g(x) = e^{ix\cdot\xi} f(x)$, then $\hat{g}(\xi) = \hat{f}(\xi - \eta)$*
3. *Let $g(x) = f(Tx)$ for $T \in GL_n(\mathbb{R})$, then $\hat{g}(\xi) = |\det T|^{-1} \hat{f}(T^{-\top}\xi)$.*
4. *Let $g(x) = D^\alpha f(x)$, then $\hat{g}(\xi) = (i\xi)^\alpha \hat{f}(\xi)$*
5. *Let $g(x) = x^\alpha f(x)$, then $\hat{g}(\xi) = (iD)^\alpha \hat{f}(\xi)$*
6. *Let $g(x) = (k \star f)(x)$ for $k \in \mathcal{S}(\mathbb{R}^n)$, then $\hat{g} = \hat{k}(\xi)\hat{f}(\xi)$*

Properties 1 and 2 can be seen as saying that the Fourier transform exchanges shifts for multiplication by oscillating factors and vice-versa. Properties 3 and 4 are arguably the properties that make the Fourier transform so useful for the study of PDE namely that the Fourier transform turns differential operators into multiplication by monomials and multiplication by monomials into multiplication. Finally, property 6 shows that the Fourier transform converts convolutions of two functions into multiplication.

Example 1.6 (The Gaussian)**.** A canonical example of a function in $S(\mathbb{R}^n)$ is a Gaussian $e^{-\langle x, Ax \rangle}$, where A is some symmetric positive definite $n \times n$ matrix. It is well known that the Fourier transform of a Gaussian is again a Gaussian. Specifically the following formula holds

$$\int_{\mathbb{R}^n} e^{-\langle x, Ax \rangle} e^{\mathrm{i}\xi\cdot x} \mathrm{d}x = \pi^{n/2} (\det A)^{-1/2} e^{-\langle \xi, A^{-1}\xi \rangle/4}. \tag{1.1}$$

Exercise 1.1. Prove the Gaussian formula (1.1).

Next we see that the Fourier transform maps the Schwartz space continuously into itself.

Lemma 1.7. *Let $f \in \mathcal{S}(\mathbb{R}^n)$, then $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$. Moreover, if $f_n \to f$ in $\mathcal{S}(\mathbb{R}^n)$, then $\hat{f}_n \to \hat{f}$ in $\mathcal{S}(\mathbb{R}^n)$.*

Proof. Note that by properties 3 and 4 in Proposition 1.5 above, we have (up to factors of i)

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

Therefore by the L^1 to L^∞ bound

$$\|\xi^\alpha D^\beta \hat{f}\|_{L^\infty} = \|\widehat{D^\alpha (x^\beta f)}\|_{L^\infty} \leq \|D^\alpha (x^\beta f)\|_{L^1} < \infty.$$

The L^1 norm above is finite because we can bound f and its derivatives away from zero by any power of x to ensure integrability. □

As a corollary, we now have a relatively simple proof of the Riemann Lebesgue lemma

Corollary 1.8. *Let $f \in L^1(\mathbb{R}^n)$, then $\hat{f}$ is uniformly continuous and $f(\xi) \to 0$ as $\xi \to \infty$.*

Proof. Let $C_0(\mathbb{R}^n)$ be the space of continuous functions that vanish at infinity (hence uniformly continuous). Note that this is a Banach space with respect to the L^∞ norm. Now by density, for each $f \in L^1(\mathbb{R}^n)$, there exists $\{f_n\} \subset S(\mathbb{R}^n)$ such that $f_n \to f$ in $L^1(\mathbb{R}^n)$. Therefore

$$\|\hat{f}_n - \hat{f}\|_{L^\infty} \leq \|f_n - f\|_{L^1}.$$

Since $\hat{f}_n \in S(\mathbb{R}^n) \subset C_0(\mathbb{R}^n)$, we conclude that $f \in C_0(\mathbb{R}^n)$. □

It is clear that Lemma 1.7 implies that that $\mathcal{F}\mathcal{S}(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$. In fact we will see that this mapping is in fact an ismorphism of $\mathcal{S}(\mathbb{R}^n)$. To see this, we need the Fourier inversion theorem.

Theorem 1.9 (Fourier Inversion)**.** *Let $f \in \mathcal{S}(\mathbb{R}^n)$, then $\hat{f}^\wedge(x) = (2\pi)^n f(-x)$. In particular the Fourier transform is invertible on $\mathcal{S}(\mathbb{R}^n)$ with inverse given by*

$$\mathcal{F}^{-1} f(x) = \frac{1}{(2\pi)^n} \hat{f}(-x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{\mathrm{i}\xi \cdot x} f(\xi) \mathrm{d}\xi.$$

Proof. We want to show that $(2\pi)^{-n} \hat{f}^\wedge(-x) = f(x)$. Using the definition of the Fourier transform, we see that

$$(2\pi)^{-n} \hat{f}^\wedge(-x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (2\pi)^{-n} e^{\mathrm{i}(x-y)\cdot\xi} f(x) \mathrm{d}x \mathrm{d}\xi.$$

Hence we would be done if we could show that $\int_{\mathbb{R}^n} e^{\mathrm{i}(x-y)\cdot\xi} \mathrm{d}\xi = (2\pi)^n \delta(x-y)$. To do this, we regularize using a Gaussian, multiplying $e^{\mathrm{i}(x-y)\cdot\xi}$ by $e^{-\epsilon^2|\xi|^2/2}$ and defining

$$I_\epsilon(x) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (2\pi)^{-n} e^{\mathrm{i}(x-y)\cdot\xi} e^{-\epsilon^2|\xi|^2/2} f(x) \mathrm{d}x \mathrm{d}\xi.$$

Note that by dominated convergence, $I_\epsilon(x) \to (2\pi)^{-n} \hat{f}^\wedge(-x)$ as $\epsilon \to 0$ for each $x \in \mathbb{R}^n$. Using Fubini and the formula (1.1), we find

$$\begin{aligned} I_\epsilon(x) &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{\mathrm{i}(x-y)\cdot\xi} (2\pi)^{-n} e^{-\epsilon^2|\xi|^2/2} \mathrm{d}\xi \right) f(x) \mathrm{d}x \\ &= \int_{\mathbb{R}^n} (2\pi)^{-n/2} \epsilon^{-n} e^{-|x-y|^2/(2\epsilon^2)} f(x) \mathrm{d}x \\ &= (\phi_\epsilon \star f)(x) \end{aligned}$$

where $\phi_\epsilon(x) = \epsilon^{-n}\phi(x/\epsilon)$ and $\phi(x) = (2\pi)^{-n/2}e^{-|x|^2/2}$. Using the properties of mollifiers, we conclude that for each $x \in \mathbb{R}^n$

$$I_\epsilon(x) = \phi_\epsilon \star f(x) \to f(x) \quad \text{as } \epsilon \to 0.$$

Therefore $(2\pi)^{-n}\hat{f}^\wedge(-x) = f(x)$ as desired.

□

1.2 Extension to L^p

An important property of the Fourier transform is how it behaves with respect to the L^2 inner product.

Lemma 1.10 (Parseval/ Plancharel)**.** *Let $f, g \in \mathcal{S}(\mathbb{R}^n)$, then*

$$\int_{\mathbb{R}^n} \hat{f} g \mathrm{d}x = \int_{\mathbb{R}^n} f \hat{g} \mathrm{d}\xi.$$

In particular we have

$$\int_{\mathbb{R}^n} \hat{f}\bar{\hat{g}} \mathrm{d}\xi = (2\pi)^n \int_{\mathbb{R}^n} f\bar{g}\mathrm{d}x. \tag{1.2}$$

Proof. Using Fubini one readily computes

$$\begin{aligned}\int_{\mathbb{R}^n} \hat{f}(\xi)g(\xi)\mathrm{d}\xi &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{-\mathrm{i}\xi\cdot x} f(x)\mathrm{d}x\right) g(\xi)\mathrm{d}\xi \\ &= \int_{\mathbb{R}^n} f(x) \left(\int_{\mathbb{R}^n} e^{-\mathrm{i}\xi\cdot x} g(\xi)\mathrm{d}\xi\right) \mathrm{d}x \\ &= \int_{\mathbb{R}^n} f(x)\hat{g}(x)\mathrm{d}x.\end{aligned}$$

To see the second identity, we use properties of Fourier transforms to see $\bar{\hat{g}}(\xi) = \hat{\bar{g}}(-\xi)$ and therefore $(\bar{\hat{g}})^\wedge(x) = (2\pi)^n\bar{g}(x)$. This implies (1.2). □

In particular these estimates give a natural extension of the Fourier transform to L^2 that has (upon suitable normalization) unitary properties.

Corollary 1.11. *For each $f \in \mathcal{S}(\mathbb{R}^n)$ one has*

$$\|\hat{f}\|_{L^2}^2 = (2\pi)^n\|f\|_{L^2}^2. \tag{1.3}$$

In particular, the normalized Fourier transform $\widetilde{\mathcal{F}} = (2\pi)^{-n/2}\mathcal{F}$ extends to a unitary operator on L^2.

Proof. The bound $\|\hat{f}\|_{L^2}^2 = (2\pi)^n\|f\|_{L^2}^2$ follows immediately from (1.2). To extend the operator to L^2 we use that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$. Fix $f \in L^2$ and let $\{f_n\} \subseteq \mathcal{S}(\mathbb{R}^n)$ be such that $f_n \to f$ in L^2. Then we have

$$\|\hat{f}_n - \hat{f}_m\|_{L^2} = (2\pi)^{n/2}\|f_n - f_m\|_{L^2}.$$

Since the right-hand side is Cauchy, the left-hand side is Cauchy also, implying that there exists $\hat{f} \in L^2$ such that $\hat{f}_n \to \hat{f}$ in L^2. Define $\mathcal{F}f := \hat{f}$. It is not hard to see that this definition does not depend on the choice of approximating sequence. Moreover passing to the limit in (1.2) also implies that

$$\langle\widetilde{\mathcal{F}}f, \widetilde{\mathcal{F}}g\rangle_{L^2} = \langle f, g\rangle_{L^2},$$

hence $\widetilde{\mathcal{F}} = \mathcal{F}/(2\pi)^n$ is an isometry on L^2. To see that it is in fact unitary, one needs to show that $\widetilde{\mathcal{F}}$ is in fact surjective. To show this, let $R = \mathrm{Ran}(\widetilde{\mathcal{F}})$ be the range of $\widetilde{\mathcal{F}}$. Note that R is dense in L^2 since it contains $\mathcal{S}(\mathbb{R}^n)$ by Theorem 1.9. In fact being an isometry R must also be closed. Indeed, let $g \in R$ and let $\{f_n\} \subseteq L^2$ be such that $g_n = \widetilde{\mathcal{F}} f_n \to g$ then

$$\|g_n - g_m\|_{L^2} = \|f_n - f_m\|_{L^2}.$$

Therefore since $\{g_n\}$ is Cauchy, so is $\{f_n\}$. Let f be the L^2 limit of $\{f_n\}$, then by continuity of $\widetilde{\mathcal{F}}$ we have $g = \widetilde{\mathcal{F}} f$, which implies that $g \in R$. It follows by density that $R = L^2$ and therefore $\widetilde{\mathcal{F}}$ is unitary. □

More generally the Fourier transform can be extended continuously to L^p for $p \in [1, 2]$. To do this we will interpolate between the easy bound $\|\hat{f}\|_{L^1} \leq \|f\|_{L^\infty}$ and the L^2 estimate (1.3). To do this we will require the Riesz Thorin interpolation theorem

Theorem 1.12 (Riesz/Thorin Interpolation)**.** *Suppose that T is a bounded linear operator from $L^{p_0} + L^{p_1} \to L^{q_0} + L^{q_1}$ and that T boundedly maps $L^{p_0} \to L^{q_0}$ and $L^{p_1} \to L^{q_1}$*

$$\|Tf\|_{L^{q_0}} \leq C_0 \|f\|_{L^{p_0}}, \quad \|Tf\|_{L^{q_1}} \leq C_1 \|f\|_{L^{p_1}}.$$

For $p_0, p_1, q_0, q_1 \in [1, \infty]$. Then T boundedly maps $L^{p_\theta} \to L^{q_\theta}$, where

$$\frac{1}{p_\theta} = (1-\theta)\frac{1}{p_0} + \theta\frac{1}{p_1}, \quad \frac{1}{q_\theta} = (1-\theta)\frac{1}{q_0} + \theta\frac{1}{q_1}, \quad \text{for all } \theta \in [0,1]$$

and

$$\|Tf\|_{L^{q_\theta}} \leq C_0^{1-\theta} C_1^{\theta} \|f\|_{L^{p_\theta}}.$$

The proof of this theorem can be proved using complex interpolation, namely by holomorphically extending the parameters p, q to the complex plane (or a strip). A wonderful exposition and proof of this can be found in the blog post by Terrance Tao here.

As an immediate application of this is the following

Theorem 1.13. *Let $f \in \mathcal{S}(\mathbb{R}^n)$, then for each $p \in [1, 2]$ we have*

$$\|\hat{f}\|_{L^{p'}} \lesssim \|f\|_{L^p}, \tag{1.4}$$

where p' is the Hölder conjugate of p, $\frac{1}{p'} + \frac{1}{p} = 1$. Consequently $\mathcal{F}$ can be extended to a continuous bounded linear operator from L^p to $L^{p'}$.

Proof. By interpolating between $\|\hat{f}\|_{L^2} \lesssim \|f\|_{L^2}$ and $\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$ using Riesz/Thorin, we obtain the desired bound from L^{p_θ} to L^{q_θ}, where

$$\frac{1}{p_\theta} = (1-\theta) + \frac{\theta}{2}, \quad \frac{1}{q_\theta} = \frac{\theta}{2}.$$

It follows that $p_\theta \in [1, 2]$ and $q_\theta = p'$. The extension to L^p follows by the usual Cauchy sequence argument. □

As it turns out, the Hausdorf/Young inequality is essentially sharp in terms of the characterization boundedness between L^p spaces, showing that any bound between L^p spaces must be of the form (1.4) and that the Fourier transform cannot continuously extended to L^p for $p > 2$.

Theorem 1.14. *Suppose that for some $p, q \in [1,\infty]$ the following holds*

$$\|\hat{f}\|_{L^q} \lesssim \|f\|_{L^p},$$

for all $f \in \mathcal{S}(\mathbb{R}^n)$. Then $p \in [1,2]$ and $q = p'$.

Proof. We begin by showing $q = p'$. This can be easily seen by scaling. Let $f \in \mathcal{S}(\mathbb{R}^n)$, $f \neq 0$ and for each $\lambda > 0$ define $f_\lambda(x) = f(x/\lambda)$. Then it follows (by property 3 of Proposition 1.5) that $\hat{f}_\lambda = \lambda^n \hat{f}(\lambda\xi)$. Using change of variables we see that

$$\|f_\lambda\|_{L^p} = \lambda^{n/p}\|f\|_{L^p}$$

and

$$\|\hat{f}_\lambda\|_{L^q} = \lambda^n \lambda^{-n/q}\|f\|_{L^q}.$$

Substituting these into the desired inequality evaluated on f_λ and collecting powers of λ gives

$$\lambda^{n\left(1-\frac{1}{q}-\frac{1}{p}\right)}\|\hat{f}\|_{L^q} \lesssim \|f\|_{L^p}.$$

The only way this can hold for all $\lambda > 0$ is if $q = p'$.

Next to show that $p \in [1,2]$, we note that by the properties of Hölder exponents that it suffices to show that $p < p'$. To do this, consider $\varphi \in C_c^\infty(\mathbb{R}^n)$ with $\operatorname{supp}\varphi \in B(0,1/2)$ and define

$$\varphi_k(x) := e^{-i\lambda x\cdot(ke_i)}\varphi(x - ke_1),$$

where e_1 is an arbitrary unit vector in $\mathbb{R}^n$ and $k \in \{1,\ldots,n,\ldots\}$. Note that by construction $\{\varphi_k\}$ all have disjoint support. Note by properties 1 and 2 of Proposition 1.5

$$\hat{\varphi}_k(\xi) = e^{-i\xi\cdot(ke_1)}\hat{\varphi}(x - \lambda k e_1).$$

Choose $N > 1$ arbitrary and define

$$f = \sum_{k=1}^N \varphi_k,$$

then using that φ_k are disjoint, we have

$$\|f\|_{L^p} = \left(\sum_{k=1}^N \int |\varphi_k|^p \mathrm{d}x\right)^{1/p} \sim N^{1/p}.$$

The same property is not true for $\hat{f}$ since $\hat{\varphi}$ no longer has compact support. Instead, we split $\hat{f} = \hat{f}_1 + \hat{f}_2$ where

$$\hat{f}_1 = \sum_{k=1}^N \hat{\varphi}_k \mathbb{1}_{B(\lambda k e_1,\lambda/2)}, \quad f_2 = \sum_{k=1}^N \hat{\varphi}_k \mathbb{1}_{B^c(\lambda k e_1,\lambda/2)}$$

Since $\{\hat{\varphi}_k \mathbb{1}_{B(\lambda k e_1,\lambda/2)}\}$ now have disjoint support, we have $\|\hat{f}_1\|_{L^q} \sim N^{1/q}$. On the other hand

$$\|\hat{f}_2\|_{L^q} \leq \sum_{k=1}^N \left(\int_{B^c(\lambda k e_1,\lambda/2)} |\hat{\varphi}_k|^q \mathrm{d}\xi\right)^{1/q} = N\|\mathbb{1}_{|\xi|>\lambda/2}\hat{\varphi}\|_{L^q}$$

Since $\hat{\varphi}$ is in $S(\mathbb{R}^n)$, we can choose λ big enough so that

$$N\|\mathbb{1}_{|\xi|>\lambda/2}\hat{\varphi}\|_{L^q} \leq \frac{1}{2}N^{1/q}.$$

Therefore for such λ

$$\|\hat{f}\|_{L^q} \geq \|\hat{f}_1\|_{L^q} - \|\hat{f}_2\|_{L^q} \gtrsim N^{1/q}.$$

It follows that

$$\|\hat{f}\|_{L^q} \lesssim \|f\|_{L^p} \quad \Rightarrow \quad N^{1/q} \lesssim N^{1/p}$$

Since this holds for all $N > 1$, we must have $1/q \leq 1/p$, or $p \leq p'$. □

1.3 Fourier multipliers

The Fourier transform can be a very convenient way to define certain classes of operators that behave, in a sense, like functions of differential operators. Recall that the Fourier transform has the property that

$$\widehat{\nabla f}(\xi) = i\xi \hat{f}(\xi).$$

Namely the it turns differentiation in to multiplication by $i\xi$. Naturally one could consider multiplication of some more general function of $i\xi$.

Definition 1.15. Let $m \in L^1_{\text{loc}}(\mathbb{R}^n)$, called a *Fourier multiplier*. Define the operator $m(\nabla)$ for each $f \in S(\mathbb{R}^n)$ by

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

In general Fourier multipliers are not bounded operators if m is not bounded. However, when m is bounded, we have the following

Lemma 1.16. *Let $m \in L^\infty$, then $m(\nabla)$ extends to a bounded linear operator on L^2 with*

$$\|m(\nabla)f\|_{L^2} \leq \|m\|_{L^\infty}\|f\|_{L^2}.$$

Proof. By Parseval, we have

$$\begin{aligned}
\|m(\nabla)f\|_{L^2} &= (2\pi)^{n/2}\|m(i\xi)\hat{f}\|_{L^2} \\
&\leq (2\pi)^{n/2}\|m\|_{L^\infty}\|\hat{f}\|_{L^2} \\
&= \|m\|_{L^\infty}\|f\|_{L^2}.
\end{aligned}$$

□

Remark 1.17. The L^p, $p \neq 2$ version of this type of estimate is much harder and will be the focus of later sections.

1.4 Characterization of Sobolev spaces

Notationally, we will denote the operator $|\nabla|$ by the operator with Fourier multiplier $|\xi|$, or more generally $|\nabla|^s$ with $s \in \mathbb{R}$ (positive or negative) by the operator with Fourier multiplier $|\xi|^s$. Sometimes the notation $\Lambda = |\nabla|$ is used instead and the associated operator is called the Zygmund operator.

The Fourier transform can be used to define a convenient norm for L^2 based Sobolev spaces H^s. Indeed it is easy to see from the properties of Fourier transform that for each $k \in \mathbb{N}$ and $f \in \mathcal{S}(\mathbb{R}^n)$

$$\|f\|_{H^k} \approx \|\langle\nabla\rangle^k f\|_{L^2},$$

where $\langle\nabla\rangle$ is the operator with Fourier multiplier $\langle\xi\rangle = \sqrt{1 + |\xi|^2}$ (some times called the Japanese bracket). In this way, we see that the H^k norm on $\mathbb{R}^n$ is no other than a weighted L^2 norm on the

Fourier side. This allows to easily extend this norm to arbitrary $s \in \mathbb{R}$ in L^2. In what follows, we will denote

$$\|f\|_{H^s} = \|\langle \nabla \rangle^s f\|_{L^2}$$

for the H^s Sobolev norm and

$$\|f\|_{\dot{H}^s} = \||\nabla|^s f\|_{L^2}$$

for the corresponding "homogeneous" norm only depending on the highest derivative (note the dot over the H). The associated inner product is

$$\langle f, g \rangle_{H^s} = \int_{\mathbb{R}^n} (1 + |\xi|^2)^s \hat{f}(\xi)\overline{\hat{g}(\xi)} \mathrm{d}\xi.$$

The associated space H^s when $s \geq 0$ is then defined to be the the closure of $\mathcal{S}(\mathbb{R}^n)$ with respect to the H^s norm, or using the extension of the Fourier transform to L^2 (again for $s \geq 0$) we have

$$H^s := \left\{ f \in L^2(\mathbb{R}^n) \, : \, \int_{\mathbb{R}^n} (1 + |\xi|^2)^s |\hat{f}(\xi)|^2 \mathrm{d}\xi < \infty \right\}$$

When $s < 0$, however, H^s instead consists in a sense of a much large class of "distributions", which we will discuss in the next section.

Remark 1.18. The choice of the norm for the H^s spaces when $s > 0$ can equivalently be taken to be

$$\|f\|_{H^s} \approx \|(1 + |\xi|^s)\hat{f}\|_{L^2}$$

as is done (for instance) in Evan's. However such a choice does not work as well for $s < 0$ due to the fact that $(1 + |\xi|^s)$ and $(1 + |\xi|^2)^{s/2}$ have very different behavior near $\xi = 0$. For the case $s < 0$ the proper choice is really $(1 + |\xi|^2)^{s/2}$.

1.5 Tempered distributions and negative Sobolev spaces

In order to deal with the Fourier transform on functions that are not L^2 (or any L^p for that matter) it is useful to introduce a space of "distributions" that play well with the Fourier transform

Definition 1.19. We define the space of *tempered distributions* $\mathcal{S}'(\mathbb{R}^n)$ to be the topological dual space to $\mathcal{S}(\mathbb{R}^n)$. Specifically, this is the space of continuous linear functionals u on $S(\mathbb{R}^n)$, whose action (or pairing with) on $\phi \in \mathcal{S})(\mathbb{R}^n)$ is denoted by

$$u(\phi) = \langle u, \phi \rangle.$$

Here, continuity of the functional $u \in \mathcal{S}'(\mathbb{R}^n)$ is in the sense that there exists an $M \geq 0$ such that

$$|\langle u, \phi \rangle| \leq \sum_{|\alpha| \leq M} \sum_{\beta \leq M} \sup_{x \in \mathbb{R}^n} |x^\alpha D^\beta \phi|.$$

When u is just a locally integrable function, with growth bounded by some power of $|x|$ at infinity, then it is clear that u is also a tempered distribution with the pairing simply given by

$$\langle u, \phi \rangle = \int_{\mathbb{R}^n} u\phi \mathrm{d}x.$$

Remark 1.20. As mentioned, the space of tempered distributions contains the space of all locally integrable functions. But also contains much more. Indeed it also contains objects like the delta function δ_x, for which $\langle \delta_x, \phi \rangle = \phi(x)$ as well as "negative order" objects like derivatives of delta functions $D^\alpha \delta_x$, where $\langle D^\alpha \delta_x, \phi \rangle = (-1)^{|\alpha|} D^\alpha \phi(x)$. The terminology "tempered" comes from the fact that the rapidly decreasing property of $\mathcal{S}'(\mathbb{R}^n)$ imposes a growth restriction on a tempered distribution. Namely $u \in \mathcal{S}'(\mathbb{R}^n)$ cannot grow faster than any power of $|x|$. Generally, a function is in $S'(\mathbb{R}^n)$ if it is locally integrable or some finite number of derivatives of a locally integrable function.

Remark 1.21. Note that any tempered distribution u can be multiplied by a function $\phi \in \mathcal{S}(\mathbb{R}^n)$, yielding $\phi u \in \mathcal{S}'(\mathbb{R}^n)$ via the relation

$$\langle \phi u, \varphi \rangle = \langle u, \phi \varphi \rangle$$

for every $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Such a relation uniquely determines the tempered distribution ϕu.

The Fourier transform can naturally be extended from $\mathcal{S}(\mathbb{R}^n)$ to $S'(\mathbb{R}^n)$ by duality.

Definition 1.22. Let $u \in \mathcal{S}'(\mathbb{R}^n)$, then the Fourier transform $\hat{u} \in \mathcal{S}'(\mathbb{R}^n)$ is defined via the generalized Parseval relation

$$\langle \hat{u}, \phi \rangle = \langle u, \hat{\phi} \rangle. \tag{1.5}$$

Remark 1.23. Note that the Fourier transform is uniquely defined as a tempered distribution by the relation (1.5) since the Fourier transform is an isomorphism of $\mathcal{S}(\mathbb{R}^n)$. Moreover it is easy to see that continuity of the Fourier transform on $\mathcal{S}(\mathbb{R}^n)$ implies continuity of its extension to the much weaker space $\mathcal{S}'(\mathbb{R}^n)$

With this definition in place, we can define the Sobolev space H^s for $s < 0$ as a space of distributions

$$H^s := \left\{ f \in \mathcal{S}'(\mathbb{R}^n) \,:\, \int_{\mathbb{R}^n} (1+|\xi|^2)^s |\hat{f}(\xi)|^2 \mathrm{d}\xi < \infty \right\}.$$

Namely, H^s is the space of tempered distributions u who's Fourier transform $\hat{u}$ is a locally integrable functions such that $(1+|\xi|^2)^{s/2}\hat{u}(\xi) \in L^2$. In general we will rarely be considering distributions that belong only to $S'(\mathbb{R}^n)$ but not to some sufficiently negative Sobolev space.

Example 1.24. The negative Sobolev spaces actually contain a large class of important distributions. In particular one can show that the delta function δ_x belongs to H^s for all $s < -n/2$. While $D^\alpha \delta_x$ belongs to H^s for all $s < -n/2 - |\alpha|$. This is left as an exercise.

It is also possible to define the convolution of a tempered distribution and a Schwartz function (in fact one can often convolve two distributions).

Definition 1.25. Let $\phi, \varphi \in \mathcal{S}(\mathbb{R}^n)$ and $u \in \mathcal{S}'(\mathbb{R}^n)$, then the convolution $u \star \phi \in \mathcal{S}'(\mathbb{R}^n)$ is defined by

$$\langle u \star \phi, \varphi \rangle = \langle u, \phi \star \varphi \rangle.$$

This allows us to extend the useful relation between convolutions and multiplication to distributions.

Lemma 1.26. *Let $u \in \mathcal{S}'(\mathbb{R}^n)$ and $\phi \in \mathcal{S}(\mathbb{R}^n)$, the following holds*

$$\widehat{u \star \phi} = \hat{u}\hat{\phi},$$

where the right-hand side is interpreted as the multiplication between $\hat{\phi} \in \mathcal{S}(\mathbb{R}^n)$ and $\hat{u} \in \mathcal{S}'(\mathbb{R}^n)$.

2 Singular integrals and convolutions

Averaging is very important concept in analysis. Indeed, some form of averaging is usually used to approximate a function by a simpler one. A classical example of this is the mollification operation $f_\epsilon = \eta_\epsilon \star f$ obtained by convolving f with $\eta_\epsilon(x) = \epsilon^{-n}\eta(x/\epsilon) > 0$, where $\eta > 0$ is a smooth symmetric compactly supported function with $\int \eta = 1$.

In this section we will study properties of convolution operations where the function η used in the convolution operator is not so nice. Indeed by the properties of Fourier transform discussed above every integrable Fourier multiplier m gives rise to a convolution operator on $\mathcal{S}(\mathbb{R}^n)$

$$m(\nabla)f = \check{m(i\cdot)} \star f = \mathcal{F}^{-1}(m(i\cdot)\hat{f}),$$

where $\mathcal{F}^{-1}f(x) = \check{f}(x) = (2\pi)^{-n}\hat{f}(-x)$ denotes the inverse Fourier transform and we have interpreted $\check{m}$ in the sense of tempered distributions. There are certain relatively simple cases when convolution operators are bounded in L^p spaces. Recall Young's inequality

Theorem 2.1 (Young's Inequality). *Assume $f \in L^p$ and $g \in L^q$, then $f \star g$ satisfies*

$$\|f \star g\|_{L^r} \le \|f\|_{L^q}\|g\|_{L^p}, \quad \textit{provided} \quad 1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}.$$

However, as we will see, there are many interesting cases where the function one is convolving with is not globally integrable (or in some cases even locally) and therefore Young's inequality doesn't apply. Such integrals are usually referred to as *singular*.

2.1 Riesz Potential

As a starting point we will consider the following function

$$K_\alpha(x) = c_n|x|^{\alpha-n}, \quad \alpha \in (0, n),$$

and the constant $c_{\alpha,n}$ is given by

$$c_{\alpha,n} = \pi^{n/2}2^\alpha \frac{\Gamma(\alpha/2)}{\Gamma((n-\alpha)/2)}.$$

We are interested in properties of the convolution operator defined for each $f \in \mathcal{S}(\mathbb{R}^n)$ by

$$I_\alpha[f](x) := K_\alpha \star f(x) = \int_{\mathbb{R}^n} K_\alpha(y)f(x-y)\mathrm{d}y.$$

The above integral is indeed well defined since $\alpha \in (0, n)$ and therefore K_α is locally integrable, meaning that its convolution with any rapidly decreasing function $f \in \mathcal{S}(\mathbb{R}^n)$ is finite. However, K_α is not globally integrable and therefore when taking its Fourier transform it must be treated as a tempered distribution and, in general may not produce a function which is even locally integrable.

Note that when $\alpha = 2$, $K_\alpha(x)$ is exactly the fundamental solution of the Laplacian, implying that in this case I_2 is just the inverse Laplacian

$$I_2[f] = (-\Delta)^{-1}f.$$

In fact we will see that $I_\alpha = (-\Delta)^{-\alpha/2}f$, where the right-hand sides is interpreted as the operator with Fourier multiplier $|\xi|^{-\alpha/2}$.

Proposition 2.2. *The Fourier transform of K_α is given by*

$$\widehat{K_\alpha}(\xi) = |\xi|^{-\alpha},$$

where the Fourier transform is interpreted in the sense of tempered distributions.

Proof. We must work with distributions and therefore work with a pairing with an element $\phi \in \mathcal{S}(\mathbb{R}^n)$. Before we start, we will use the following useful representation of the function $|x|^{n-\alpha}$. Indeed, recall the definition of the Gamma function

$$\Gamma\left(\frac{n-\alpha}{2}\right) = \int_{\mathbb{R}_+} e^{-s} s^{(n-\alpha)/2} \frac{\mathrm{d}s}{s}.$$

Upon changing variables to $s = |x|^2 t$ for a fixed $x \in \mathbb{R}^n$, we see that

$$\Gamma\left(\frac{n-\alpha}{2}\right) |x|^{\alpha-n} = \int_{\mathbb{R}_+} e^{-t|x|^2} t^{(n-\alpha)/2} \frac{\mathrm{d}t}{t}.$$

It follows that as distribution we have

$$\begin{aligned}
\left\langle \Gamma\left(\tfrac{n-\alpha}{2}\right) \widehat{|x|^{\alpha-n}}, \phi \right\rangle &= \left\langle \Gamma\left(\tfrac{n-\alpha}{2}\right) |x|^{\alpha-n}, \hat{\phi} \right\rangle \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}_+} e^{-t|x|^2} t^{(n-\alpha)/2} \frac{\mathrm{d}t}{t} \right) \left(\int_{\mathbb{R}^n} e^{-\mathrm{i}x\cdot\xi} \phi(\xi) \mathrm{d}\xi \right) \mathrm{d}x \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}_+} \left(\int_{\mathbb{R}^n} e^{-t|x|^2} e^{-\mathrm{i}x\cdot\xi} \mathrm{d}\xi \right) t^{(n-\alpha)/2} \frac{\mathrm{d}t}{t} \phi(\xi) \mathrm{d}\xi \\
&= \langle G_\alpha, \phi \rangle,
\end{aligned}$$

where in the second to last line we used Fubini, and we defined

$$G_\alpha(\xi) := \int_{\mathbb{R}_+} \left(\int_{\mathbb{R}^n} e^{-t|x|^2} e^{-\mathrm{i}x\cdot\xi} \mathrm{d}\xi \right) t^{(n-\alpha)/2} \frac{\mathrm{d}t}{t}.$$

Using the Fourier transform of a Gaussian (1.1), we find $\widehat{e^{-t|x|^2}}(\xi) = \pi^{n/2} t^{-n/2} e^{-|\xi|^2/4t}$. Therefore

$$\begin{aligned}
G_\alpha(\xi) &= \pi^{n/2} \int_{\mathbb{R}_+} t^{-\alpha/2} e^{-|\xi|^2/4t} \frac{\mathrm{d}t}{t} \\
&= \pi^{n/2} 2^\alpha |\xi|^{-\alpha} \int_{\mathbb{R}_+} u^{\alpha/2} e^{-u} \frac{\mathrm{d}u}{u} \\
&= \pi^{n/2} 2^\alpha |\xi|^{-\alpha} \Gamma\left(\tfrac{\alpha}{2}\right),
\end{aligned}$$

where in the second line above we used the change of variables $u = |\xi|^2/4t$. It follows that in the sense of distributions we have

$$\Gamma\left(\tfrac{n-\alpha}{2}\right) \widehat{|x|^{\alpha-n}} = \pi^{n/2} 2^\alpha |\xi|^{-\alpha} \Gamma\left(\tfrac{\alpha}{2}\right),$$

which is equivalent to $\widehat{K_\alpha}(\xi) = |\xi|^{-\alpha}$. □

This result implies that the operator I_α is indeed given a Fourier multiplier operator

$$I_\alpha[f] = |\nabla|^{-\alpha} f = (-\Delta)^{-\alpha/2} f.$$

2.2 Hardy-Littlewood-Sobolev and the maximal function

Now we consider a fundamental inequality concerning Riesz potentials (or more general convolutions with functions of the type $|x|^{-\alpha}$, analogous to Young's inequality.

Theorem 2.3 (Hardy-Littlewood-Sobolev). *Let $\alpha \in (0, n)$ and $f \in \mathcal{S}(\mathbb{R}^n)$, then $|x|^{-\alpha} \star f$ satisfies*

$$\| |x|^{-\alpha} \star f \|_{L^r} \lesssim_{n,p} \|f\|_{L^p} \quad \textit{provided} \quad 1 + \frac{1}{r} = \frac{1}{p} + \frac{\alpha}{n}, \quad p, r \in (1, \infty).$$

The proof of this theorem will involve studying a very important function known as the *Hardy-Littlewood maximal function*

$$Mf(x) = \sup_{r>0} \fint_{B_r(x)} |f(y)| \mathrm{d}y, \tag{2.1}$$

where $B_r(x)$ is the ball of radius r centered at x and $\fint_B |f(y)|\mathrm{d}y = \frac{1}{|B|} \int_B |f(y)|\mathrm{d}y$ is the average of $|f|$ over the ball B. Hence, the maximal function describes the largest possible value that averages of f can take over ball centered around a certain point. Apriori it is not clear that Mf is even finite at every point. Indeed, if f is only L^p, then if x is not a Lebesgue point, the value of $Mf(x)$ may be infinite at that point. Nonetheless, the following fundamental theorem due to Hardy and Littlewood state properties of Mf and gives sharp conditions on its boundedness as an operator on L^p.

Theorem 2.4 (Hardy-Littlewood Maximal Inequality). *The maximal function Mf satisfies*

(a) For $f \in L^p$, $p \in (1, \infty]$

$$\|Mf\|_{L^p} \lesssim_{p,n} \|f\|_{L^p} \tag{2.2}$$

(b) For $f \in L^1$

$$|\{Mf > \lambda\}| \leq \frac{C_n}{\lambda} \|f\|_{L^1}. \tag{2.3}$$

Remark 2.5. Theorem 2.4 implies that Mf is almost surely finite for $f \in L^p$ for $p \in [1, \infty]$, however $M : L^p \to L^p$ is only a bounded operator if $p > 1$. The bound (2.3) is weaker than boundedness on L^1. It, for instance, allows for Mf to be $1/|x|^n$ which is not integrable. Such an estimate is called a *weak type* estimate as it maps L^1 to a space $L^{1,\infty}$ known as weak L^1. Indeed estimate (2.2) is false on L^1.

To see this, choose $\phi \in C_c^\infty$ with supp$\phi \subseteq B_1(0)$. Note that if $|x| > 2$, then $B_{1/2}(0) \subseteq B_{2|x|}(x)$ and therefore for $|x| > 2$

$$M\phi(x) \geq \fint_{B_{2|x|}(x)} \phi(y)\mathrm{d}y = \frac{1}{2^n |x|^n} \int_{B_{1/2}(0)} \phi(y)\mathrm{d}y \gtrsim \frac{1}{|x|^n}$$

which is not in L^1.

In order the prove this theorem, we will use certain covering Lemma due to Vitali.

Lemma 2.6 (Vitali). *Let $B_1, \ldots B_N$ be a finite collection of balls in $\mathbb{R}^n$. Then there is a sub-collection $B_{n_1}, \ldots B_{n_k}$ of disjoint balls such that*

$$\bigcup_i B_i \subseteq \bigcup_j 3B_{n_j},$$

where $3B_{n_k}$ denotes the ball B_{n_k} scaled up by a factor of 3.

Proof. If the collection of balls is empty, then the theorem is obviously true. Now we proceed with the following algorithm. Order the balls by radius arbitrarily breaking ties and let B_{n_1} be the ball with the largest radius and remove it from the list. Next discard all the balls which have a non-empty intersection with B_{n_1}. If there are any balls left, let B_{n_2} be the largest ball of the remaining balls and remove it from the list. If there are no balls left, then stop. Repeat this procedure discarding all remaining balls which have an intersection with B_{n_2} and choosing the largest remaining ball.

This procedure removes a least one ball from the ordered list at each step and therefore must terminate after finitely many iterations. The resulting disjoint collection of balls $B_{n_1}, B_{n_2}, \dots, B_{n_k}$ have the property that any ball B_i not in the disjoint collection must intersect a larger ball B_{n_j} in the disjoint collection. By the triangle inequality (draw a picture) we see that $B_i \subseteq 3B_{n_j}$. Therefore the proof is complete. □

Now we are equipped to prove the Maximal inequality.

Proof of maximal inequality Theorem 2.4. Step 1: We begin by proving the weak type L^1 estimate (2.3). To do this, fix $\lambda > 0$ and consider the set $\{Mf > \lambda\}$. Let $K \subseteq \{Mf > \lambda\}$ be an arbitrary compact set. For each $x \in K$ we have $Mf(x) > \lambda$ and therefore by the definition (2.1) of the maximal function there exists an r_x such that

$$\int_{B_{r_x}(x)} |f(y)| \mathrm{d}y > \lambda |B_{r_x}(x)|. \tag{2.4}$$

Moreover $\{B_{r_k}(x)\}$ is clearly an open covering of K and therefore there exists a finite sub-cover $B_1, \dots, B_N$ of K. By Vitali's covering Lemma, there is a disjoint sub collection $B_{n_1}, \dots, B_{n_k}$ such that

$$|K| \leq \left| \bigcup_i B_i \right| \leq 3^n \sum_j |B_{n_j}|.$$

It follows by (2.4) that

$$|K| \leq \frac{3^n}{\lambda} \sum_j \int_{B_{n_j}} |f(y)| \mathrm{d}y \leq \frac{3^n}{\lambda} \|f\|_{L^1}.$$

Since this holds for any compact $K \subseteq \{Mf > \lambda\}$, we conclude that it also holds for $\{Mf > \lambda\}$ itself by the inner regularity of Lebesgue measure.

Step 2: Next we prove (2.2) from (2.3). To do this, we first fix $\lambda > 0$ and denote

$$g = \mathbb{1}_{\{|f| > \lambda/2\}} |f|.$$

Note that

$$Mf(x) = \sup_{r>0} \fint_{B_r(x)} (\mathbb{1}_{\{|f|>\lambda/2\}} + \mathbb{1}_{\{|f|\leq\lambda/2\}}) |f(y)| \mathrm{d}y \leq Mg(x) + \lambda/2$$

Therefore we have $\{Mf > \lambda\} \subseteq \{Mg > \lambda/2\}$. Applying the maximal inequality (2.3) to Mg gives

$$|\{Mf > \lambda\}| \leq |\{Mg > \lambda/2\}| \leq \frac{2C}{\lambda} \int_{\mathbb{R}^n} \mathbb{1}_{\{|f|>\lambda/2\}}(y) |f(y)| \mathrm{d}y.$$

The layer cake formula then implies that for $p > 1$

$$\|Mf\|_{L^p}^p = p \int_0^\infty \lambda^{p-1} |\{Mf > \lambda\}| \mathrm{d}\lambda,$$

whereby using the above maximal function bound and applying Fubini gives

$$\begin{aligned}\|Mf\|_{L^p}^p &\lesssim \int_0^\infty \int_{\mathbb{R}^n} \lambda^{p-2} \mathbb{1}_{|f|>\lambda/2} |f(y)| \mathrm{d}y \\ &= \int_{\mathbb{R}^n} \left(\int_0^{2|f(y)|} \lambda^{p-2} \mathrm{d}\lambda \right) |f(y)| \mathrm{d}y \\ &= \frac{2^{p-1}}{p-1} \int_{\mathbb{R}^n} |f(y)|^p \mathrm{d}y.\end{aligned}$$

This proves the result for $p \in (1, \infty)$. The case $p = \infty$ is obvious. □

Now we are ready to prove the Hardy-Littlewood-Sobolev (HLS) inequality Theorem 2.3.

Proof of Hardy-Littlewood-Sobolev Theorem 2.3. To begin we see that for each x

$$|(f \star |x|^{-\alpha})(x)| \leq \int_{\mathbb{R}^n} |y|^{-\alpha} |f(x-y)| \mathrm{d}y = \int_{\mathbb{R}^n} |y|^{-\alpha} |f(x+y)| \mathrm{d}y.$$

Let $2^{\mathbb{Z}} = \{2^j : j \in \mathbb{Z}\}$ be the set of dyadic (powers of 2) integers. We divy up the integrand on the right-hand side over annuli of the form $\{2^j \leq |y| \leq 2^{j+1}\}$ and estimate

$$\begin{aligned}|(f \star |x|^{-\alpha})(x)| &\leq \sum_{s \in 2^{\mathbb{Z}}} \int_{s \leq |y| \leq 2s} |y|^{-\alpha} |f(x+y)| \mathrm{d}y \\ &\leq \sum_{s \in 2^{\mathbb{Z}}} s^{-\alpha} \int_{s \leq |y| \leq 2s} |f(x+y)| \mathrm{d}y.\end{aligned}$$

Note that by Hölder's inequality

$$s^{-\alpha} \int_{s \leq |y| \leq 2s} |f(x+y)| \mathrm{d}y \lesssim s^{n/p' - \alpha} \|f\|_{L^p} = s^{-n/r} \|f\|_{L^p},$$

however, $\sum_{s \in 2^{\mathbb{Z}}} s^{-n/r}$ is not generally summable near $s = 0$ since you get a growing geometric series. To remedy this, we chop up the sum into regions where $s \leq R$ and $s \geq R$ for some cut-off R (to be chosen later). When $s \geq R$ we simply use the above bound. When $s \leq R$ we will replace the integral with it's average to get and extra factor of s^n and bound by the maximal function

$$s^{-\alpha} \int_{s \leq |y| \leq 2s} |f(x+y)| \mathrm{d}y \leq s^{n-\alpha} ⨍_{|y| \leq 2s} |f(x+y)| \mathrm{d}y \leq s^{n-\alpha} Mf(x).$$

Since $\alpha \in (0, n)$, the series $\sum_{s \in 2^{\mathbb{Z}}, s \leq R} s^{n-\alpha}$ is summable as a geometric series. Putting this together gives

$$\begin{aligned}\sum_{s \in 2^{\mathbb{Z}}} s^{-\alpha} \int_{s \leq |y| \leq 2s} |f(x+y)| \mathrm{d}y &= \sum_{\substack{s \in 2^{\mathbb{Z}} \\ s \leq R}} s^{-\alpha} \int_{s \leq |y| \leq 2s} |f(x+y)| \mathrm{d}y + \sum_{\substack{s \in 2^{\mathbb{Z}} \\ s \geq R}} s^{-\alpha} \int_{s \leq |y| \leq 2s} |f(x+y)| \mathrm{d}y \\ &\leq \left(\sum_{\substack{s \in 2^{\mathbb{Z}} \\ s \leq R}} s^{n-\alpha} \right) Mf(x) + \left(\sum_{\substack{s \in 2^{\mathbb{Z}} \\ s \geq R}} s^{-n/r} \right) \|f\|_{L^p} \\ &\lesssim R^{n-\alpha} Mf(x) + R^{-n/r} \|f\|_{L^p},\end{aligned} \tag{2.5}$$

where in the last inequality we used the property that any convergent geometric series can be controlled by the largest term in the series. □

Now we optimize in the value of R. We do this by setting $R^{n-\alpha}Mf(x) = R^{-n/r}\|f\|_{L^p}$ and solving for R, giving

$$R = \left(\frac{\|f\|_{L^p}}{Mf(x)}\right)^{p/n},$$

assigning the value $+\infty$ when $Mf(x) = 0$. Substituting this into (2.5) gives

$$|f \star |x|^{-\alpha}(x)| \leq |Mf(x)|^{p/r}\|f\|_{L^p}^{1-p/r}.$$

It now follows from the maximal inequality (2.2) that since $p > 1$

$$\begin{aligned}\|f \star |x|^{-\alpha}\|_{L^r} &\leq \|Mf\|_{L^p}^{p/r}\|f\|_{L^p}^{1-p/r}\\ &\lesssim \|f\|_{L^p}.\end{aligned}$$

2.3 Sobolev embeddings and interpolation

The Hardy-Littlewood-Sobolev inequality can be used to prove most of the L^p based Sobolev embeddings on $\mathbb{R}^n$ and naturally extends them to the setting of fractional Sobolev spaces.

Theorem 2.7 (Sobolev Embedding)**.** *Let $s > 0$ and $p \in (1, \infty)$, then for every $f \in \mathcal{S}(\mathbb{R}^n)$*

$$\|f\|_{L^r} \lesssim \||\nabla|^s f\|_{L^p}, \quad \textit{if} \quad \frac{1}{r} = \frac{1}{p} - \frac{s}{n}, \quad r \in (1, \infty)$$

Proof. To prove this we write for $f \in \mathcal{S}(\mathbb{R}^n)$

$$\|f\|_{L^r} = \sup_{\substack{g\in L^{r'}\\ \|g\|_{L^{r'}}=1}} \langle g, f\rangle = \sup_{\substack{g\in \mathcal{S}(\mathbb{R}^n)\\ \|g\|_{L^{r'}}=1}} \langle g, f\rangle = \sup_{\substack{g\in \mathcal{S}(\mathbb{R}^n)\\ \|g\|_{L^{r'}}=1}} \langle \hat{g}, \hat{f}\rangle = \sup_{\substack{g\in \mathcal{S}(\mathbb{R}^n)\\ \|g\|_{L^{r'}}=1}} \langle |\xi|^{-s}\hat{g}, |\xi|^s\hat{f}\rangle$$

where in the second equality we used that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^{r'}$. Ideally we would like to apply Plancharel again to obtain $\langle |\nabla|^{-s}g, |\nabla|^s f\rangle$, however neither $|\xi|^{-s}\hat{g}$ nor $|\xi|^s\hat{f}$ are in $S'(\mathbb{R}^n)$ due to the singularity at 0 (we need one to be in $\mathcal{S}(\mathbb{R}^n)$ to treat the other as in $\mathcal{S}'(\mathbb{R}^n)$. To remedy this we consider the modified set

$$\mathcal{S}_0(\mathbb{R}^n) := \left\{f \in \mathcal{S}(\mathbb{R}^n) \,:\, \hat{f} \text{ vanishes in a neighborhood of } 0\right\}.$$

We claim that this set is dense in $L^{r'}$. Indeed to see this we simply need to show $\mathcal{S}_0(\mathbb{R}^n)$ is dense in $\mathcal{S}(\mathbb{R}^n)$ in the $L^{r'}$ metric. For each $g \in \mathcal{S}(\mathbb{R}^n)$ denote

$$\hat{g}_\epsilon(\xi) = \hat{g}(\xi)(1 - \varphi(\xi/\epsilon)),$$

where $\varphi \in C_c^\infty$, $\varphi \geq 0$, $\varphi = 1$ on $B_1(0)$ and vanishes outside of $B_2(0)$. It follows that

$$\hat{g}(\xi) - \hat{g}_\epsilon(\xi) = \hat{g}(\xi)\varphi(\xi/\epsilon)$$

Since $\mathcal{F}^{-1}\varphi(x/\epsilon) = \epsilon^n\check{\varphi}(\epsilon x)$, we see that so that $g - g_\epsilon = \epsilon^n g \star (\check{\varphi}(\cdot\epsilon))$. Therefore by Young's inequality

$$\|g - g_\epsilon\|_{L^{r'}} \leq \epsilon^n\|g\|_{L^1}\|\check{\varphi}(\cdot\epsilon)\|_{L^{r'}} = \epsilon^{n(1-1/r')}\|g\|_{L^1}\|\check{\varphi}\|_{L^{r'}}.$$

Therefore $\|g - g_\epsilon\|_{L^{r'}} \to 0$ as $\epsilon \to 0$ since $r' > 1$.

Using just proven density of $\mathcal{S}_0(\mathbb{R}^n)$, we replace $\mathcal{S}(\mathbb{R}^n)$ with $\mathcal{S}_0(\mathbb{R}^n)$ so that $|\xi|^{-s}\hat{g}$ belongs to $\mathcal{S}(\mathbb{R}^n)$. It follows that

$$\|f\|_{L^r} = \sup_{\substack{g\in\mathcal{S}_0(\mathbb{R}^n)\\ \|g\|_{L^{r'}}=1}} \langle |\xi|^{-s}\hat{g}, |\xi|^s\hat{f}\rangle = \sup_{\substack{g\in\mathcal{S}_0(\mathbb{R}^n)\\ \|g\|_{L^{r'}}=1}} \langle |\nabla|^{-s}g, |\nabla|^s f\rangle \le \sup_{\substack{g\in\mathcal{S}_0(\mathbb{R}^n)\\ \|g\|_{L^{r'}}=1}} \||\nabla|^{-s}g\|_{L^{p'}} \||\nabla|^s f\|_{L^p}.$$

We now apply the HLS inequality to obtain

$$\||\nabla|^{-s}g\|_{L^{p'}} \approx \||x|^{s-n} \star g\|_{L^{p'}} \lesssim \|g\|_{L^{r'}}$$

where

$$1 + \frac{1}{p'} = \frac{1}{r'} + \frac{n-s}{n} \quad \Rightarrow \quad \frac{1}{r} = \frac{1}{p} - \frac{s}{n}.$$

□

Using this, one can also prove the following general interpolation inequality.

Theorem 2.8 (Gagliardo-Nirenberg Interpolation inequality)**.** *Let $f \in \mathcal{S}(\mathbb{R}^n)$ then for each $p_0, p_1 \in (1,\infty)$ and $s > 0$, we have the following interpolation inequality for each $\theta \in [0,1]$*

$$\|f\|_{L^{p_\theta}} \le \|f\|_{L^{p_0}}^{1-\theta} \||\nabla|^s f\|_{L^{p_1}}^{\theta},$$

where

$$\frac{1}{p_\theta} = (1-\theta)\frac{1}{p_0} + \theta\left(\frac{1}{p_1} - \frac{s}{n}\right).$$

Proof. Left as an exercise. □

Remark 2.9. In the above Sobolev inequalities, we don't exactly recover the classical Sobolev inequalities for integer values of s. This is because of the innate difference between the operator $|\nabla|$ and the gradient operator ∇. While the behave similarly on the Fourier side. As operators, they behave quite differently as $|\nabla|$ is non-local and ∇ is local, meaning that $|\nabla|f(x)$ depends on all values of the function x, while $\nabla f(x)$ depends on only value of f in small neighborhood of the point evaluated. We will rectify this difference in the next section.

2.4 Riesz transform

As discussed in the previous section, we would like understand how to relate derivatives operators $|\nabla|f$ which are inherently non-local to ∇f, which is local to recover the classical Sobolev embedding theorems from the ones proven above. To do this, we note that the Fourier multiplier for ∂_j is given by $i\xi$, while the Fourier multiplier for $|\nabla|$ is given by $|\xi|$. At the level of the multiplier, we we can relate the two via

$$|\xi| = \sum_j \frac{-i\xi_j}{|\xi|} i\xi_j = \sum_j m_j(\xi) i\xi_j,$$

where we have defined

$$m_j(\xi) := -i\frac{\xi_j}{|\xi|},$$

called the *Riesz multiplier* (not to be confused with Riesz potential). The corresponding operator R_j defined for each $f \in \mathcal{S}(\mathbb{R}^n)$ and $j \in \{1,\dots,n\}$ by

$$\widehat{R_j f} = m_j(\xi) f$$

is called the *Riesz transform.* The vector of Riesz transforms we will denote by $R = (R_1, \ldots, R_n)$. It follows that we can write

$$|\nabla| = R \cdot \nabla = \sum_j R_j \partial_{x_j}.$$

Additionally, it is easy to see at the level of the multipliers that we also have $\nabla = R|\nabla|$. Note that the Riesz multiplier m_j is in L^∞ and therefore Lemma 1.16 implies that R_j is bounded from $L^2 \to L^2$. This immediately implies for $f \in \mathcal{S}(\mathbb{R}^n)$

$$\||\nabla|f\|_{L^2} \approx \|\nabla f\|_{L^2}.$$

However, as mentioned earlier, Lemma 1.16 does not apply generally in L^p for $p \neq 2$. The main issue being that the Fourier transform loses its isometric structure on L^p and, in light of Theorem 1.14, the lack of boundedness of the Fourier transform between L^p spaces.

One of the main goals of this section will be to show that the Riesz transform, along with a large class of similar Fourier multipliers, is bounded from L^p to L^p for $p \in (0, \infty)$ (see the Mikhlin multiplier Theorem **??** below).

Naturally, the Riesz transform can be written as a convolution with a distribution using the properties of the Fourier transform

$$R_j f = \mathcal{F}^{-1}(m_j \hat{f}) = \check{m}_j \star f, \quad f \in \mathcal{S}(\mathbb{R}^n),$$

where $\check{m}_j \in \mathcal{S}'(\mathbb{R}^n)$ denotes the inverse Fourier transform of m_j. In fact, using the explicit computation for the Riesz potential given in Proposition 2.2 we can easily compute $\check{m}_j$.

Lemma 2.10. *Let m_j be the Riesz multiplier, then the following formula holds in the sense of tempered distributions*

$$K(x) := \check{m}_j(x) = p.v. \frac{x_j}{c_n |x|^{n+1}},$$

where $c_n = \frac{\pi^{(n+1)/2}}{\Gamma((n+1)/2)} = \pi\omega_{n-1}$, where ω_{n-1} is the volume of the $n-1$ dimensional Euclidean unit ball, and p.v. denotes the principle value of the distribution, defined as the limit

$$\langle K, \phi \rangle = \lim_{\epsilon \to 0} \frac{1}{c_{1,n}} \langle x_j |x|^{-n-1} \mathbb{1}_{|x|>\epsilon}, \phi \rangle.$$

Proof. Recall from Proposition 2.2 that $\mathcal{F}^{-1}|\xi|^{-1} = \frac{1}{c_{1,n}}|x|^{1-n}$. Therefore, since $-i\xi_j$ is the Fourier multiplier of $-\partial_{x_j}$, we see that in the sense of distributions

$$\check{m}_j = \frac{-\partial_{x_j}|x|^{1-n}}{c_{1,n}}.$$

Computing this derivative away from zero (by choosing test functions supported in a neighborhood away from zero) gives

$$\check{m}_j(x) = c_n \frac{x_j}{|x|^{n+1}} \quad \text{on} \quad \mathbb{R}^n \backslash \{0\},$$

where we used that $\Gamma(1/2) = \sqrt{\pi}$ and $\frac{(n-1)}{2}\Gamma(\frac{n-1}{2}) = \Gamma(\frac{n+1}{2})$.

To see that the principle value exists, we note that $\frac{x_j}{|x|^{n+1}}$ is an odd function and therefore

$$\int_{B_\epsilon(0)^c} \frac{x_j}{|x|^{n+1}} \mathrm{d}x = 0.$$

It follows that for any $f \in \mathcal{S}(\mathbb{R}^n)$

$$\int_{B_\epsilon(0)^c} \frac{x_j}{|x|^{n+1}} f(x)\mathrm{d}x = \int_{B_\epsilon(0)^c} \frac{x_j}{|x|^{n+1}} (f(x) - f(0))\mathrm{d}x.$$

Since we can bound $|f(x) - f(0)| \leq |x| \|Df\|_{L^\infty}$, we see that

$$\left| \frac{x_j}{|x|^{n+1}} (f(x) - f(0)) \right| \lesssim |x|^{n-1}$$

which is integrable near zero. Therefore, the limit as $\epsilon \to 0$ exists by dominated convergence (the correct proof of this requires splitting the integral into $|x| \leq 1$ and $|x| \geq 1$ and using that f is rapidly decreasing). □

Therefore we see that the Riesz transform is defined as a convolution with the function $\frac{x_j}{|x|^{n+1}}$ interpreted in the principle value sense described in . Notably the function $\frac{x_j}{|x|^{n+1}} \leq \frac{1}{|x|^n}$ is not even locally absolutely integrable near 0 (although it is conditionally integrable as we saw above). This means that the Hardy/Littlewood/Sobolev inequality Theorem 2.3 doesn't apply in estimating the Riesz transform R_j since $\alpha = n$ is the boundary case. Nonetheless, such integrals can still be estimated.

2.5 The Calderón-Zygmund Theorem

In what follows, we consider a general class of kernels that mimic the properties we saw for the Riesz transform, these will be called the class of Calderón Zygmund potentials

Definition 2.11 (Calderón Zygmund potentials)**.** We say a function $K \in C^1(\mathbb{R}^n \backslash \{0\})$ is a Calderón Zygmund potential if the following properties hold:

1. $|K(x)| \lesssim \frac{1}{|x|^n}$
2. $|\nabla K(x)| \lesssim \frac{1}{|x|^{n+1}}$
3. For each $0 < R_1 < R_2 < \infty$,

$$\int_{R_1 \leq |x| \leq R_2} K(x)\mathrm{d}x = 0.$$

Remark 2.12. It is easy to see that the kernel $x_j/|x|^{n+1}$ associated to the Riesz transform is a Calderón Zygmund potential.

It is important to note that, following the exact same proof as in Lemma 2.10, properties 1 and 2 above imply that the convolution of K with $f \in \mathcal{S}(\mathbb{R}^n)$ exists in the principle values sense

$$K \star f(x) := p.v. \int_{\mathbb{R}^n} K(y) f(x-y)\mathrm{d}y = \lim_{\epsilon \to 0} \int_{\mathbb{R}^n} \mathbb{1}_{|y|>\epsilon} K(y) f(x-y)\mathrm{d}y.$$

Property 3 is a regularity condition, that is crucial for estimating the convolution. Typically, the regularity condition 3 is replaced by the much weaker smoothness condition (sometimes called Hörmander's condition)

$$\sup_{y \neq 0} \int_{|x|>2|y|} |K(x-y) - K(x)|\mathrm{d}x < \infty, \tag{2.6}$$

whose formulation will actually be more useful later.

Lemma 2.13. *The smoothness condition (2.6) follows from property 3.*

Proof. First note that property 3 implies

$$|K(x-y)-K(x)| \le |y|\int_0^1 |\nabla K(x-sy)|\mathrm{d}s \lesssim |y|\int_0^1 \frac{1}{|x-sy|^{n+1}}\mathrm{d}s \lesssim \frac{|y|}{(|x|-|y|)^{n+1}}$$

Therefore, when $|x|>2|y|$, we have

$$|K(x-y)-K(x)| \lesssim \frac{|y|}{|x|^{n+1}},$$

and so

$$\int_{|x|>2|y|} |K(x-y)-K(x)|\mathrm{d}x \lesssim \int_{|x|>2|y|} \frac{|y|}{|x|^{n+1}}\mathrm{d}x \lesssim |y|\frac{1}{|y|} = 1$$

□

The main theorem of this section is the following

Theorem 2.14. *Let K be a Calderón Zygmund potential and let $K \star f$ be the principle value convolution with $f \in \mathcal{S}(\mathbb{R}^n)$. Then the following estimates hold:*

1. *for every $\lambda > 0$*
$$|\{|K\star f| > \lambda\}| \lesssim \frac{\|f\|_{L^1}}{\lambda} \tag{2.7}$$
2. *for every $p \in (1,\infty)$,*
$$\|K\star f\|_{L^p} \lesssim \|f\|_{L^p}.$$

Moreover $K \star f$ can be extended to a bounded linear operator from $L^p \to L^p$ if $p \in (1,\infty)$.

The proof of this theorem will be broken up into a few steps. Our first step in the proof will be to prove an $L^2 \to L^2$ bound. After that, we will use this bound to prove the weak type estimate (2.7). This estimate is the most technical and requires a very special decomposition of functions, called the Calderón-Zygmund decomposition (see section 2.5.2). Finally, as in the our proof of the maximal function bound (2.2), we will use this weak type estimate to prove the strong type bound.

2.5.1 L^2 bound

Being a convolution, it is useful to see what the associated Fourier multiplier looks like. If we are in any sense similar to the Riesz transform, we should expect that the multiplier is a bounded function. If this is the case we know from Lemma 1.16 that $K \star f$ is bounded in the L^2 norm.

Lemma 2.15. *Let K be a Calderón Zygmund function, then $\hat{K} \in L^\infty$.*

Proof. We consider the Fourier transform $\hat{K}$ as a tempered distribution. Denote

$$\hat{K}_\epsilon(\xi) = \int_{\epsilon<|y|<\epsilon^{-1}} K(y)e^{-\mathrm{i}\xi\cdot y}\mathrm{d}y.$$

Then $\hat{K} = \lim_{\epsilon\to 0}\hat{K}_\epsilon$ in the sense of distributions, since for each $\phi \in \mathcal{S}(\mathbb{R}^n)$

$$\langle \hat{K},\phi\rangle = \langle K,\hat{\phi}\rangle = \lim_{\epsilon\to 0}\langle K\mathbb{1}_{\epsilon<|y|<\epsilon^{-1}},\hat{\phi}\rangle = \lim_{\epsilon\to 0}\langle \hat{K}_\epsilon,\phi\rangle.$$

To show that $\hat{K} \in L^\infty$ it suffices to show that

$$\langle \hat{K}, \phi \rangle \leq \|\phi\|_{L^1}$$

since this implies (by density of $\mathcal{S}(\mathbb{R}^n)$ in L^1) that $\hat{K}$ belongs to L^∞. Hence, we simply need to show that

$$\sup_{\epsilon>0} \|\hat{K}_\epsilon\|_{L^\infty} < \infty.$$

To see this, we divy up the integral over dyadic shells as we did in the proof of the Hardy-Litlewood-Sobolev inequality.

$$\hat{K}_\epsilon(\xi) = \sum_{\substack{r\in 2^{\mathbb{Z}} \\ \epsilon \leq r \leq \epsilon^{-1}}} \int_{r<|y|<2r} K(y) e^{\mathrm{i}\xi\cdot x} \mathrm{d}x. \tag{2.8}$$

Each term in the sum can be estimated differently depending on whether $r \leq R$ or $r \geq R$, for some R to be chosen. When $r \leq R$, we use the cancellation property to estimate

$$\begin{aligned}\left| \int_{r<|y|<2r} K(y) e^{\mathrm{i}\xi\cdot x} \mathrm{d}x \right| &= \left| \int_{r<|y|<2r} K(y)(e^{\mathrm{i}\xi\cdot x} - 1) \mathrm{d}x \right| \lesssim r|\xi| \int_{r\leq|x|\leq 2r} |K(x)| \mathrm{d}x \\ &\lesssim r|\xi| \int_{r\leq|x|\leq 2r} \frac{1}{|x|^n} \mathrm{d}x \lesssim r|\xi| r^n r^{-n} = r|\xi|.\end{aligned}$$

When $r \geq R$ we write $e^{\mathrm{i}\xi\cdot x}$ as $\frac{i\xi}{|\xi|^2} \cdot \nabla_x e^{\mathrm{i}\xi\cdot x}$ and integrate by parts to obtain

$$\begin{aligned}\left| \int_{r<|y|<2r} K(y) e^{\mathrm{i}\xi\cdot x} \mathrm{d}x \right| &= |\xi|^{-2} \left| \int_{r<|y|<2r} K(y) \xi \cdot \nabla e^{\mathrm{i}\xi\cdot x} \mathrm{d}x \right| \\ &\lesssim |\xi|^{-2} \left| \int_{r<|y|<2r} \xi \cdot \nabla K(y) e^{\mathrm{i}\xi\cdot x} \mathrm{d}x \right| + |\xi|^{-2} \left| \int_{\partial\{r<|y|<2r\}} \xi \cdot n(y) K(y) e^{\mathrm{i}\xi\cdot x} \mathrm{d}S(y) \right| \\ &\lesssim |\xi|^{-1} \left(\left| \int_{r<|y|<2r} |x|^{-n-1} \mathrm{d}x \right| + \left| \int_{\partial\{r<|y|<2r\}} |x|^{-n} \mathrm{d}S(y) \right| \right) \\ &\lesssim |\xi|^{-1} \left(r^{-n-1} r^n + r^{-n} r^{n-1} \right) \lesssim \frac{1}{|\xi| r}.\end{aligned}$$

Using properties of geometric series we have

$$\sum_{\substack{r\in 2^{\mathbb{Z}} \\ r\leq R}} r \lesssim R \quad \text{and} \quad \sum_{\substack{r\in 2^{\mathbb{Z}} \\ r\geq R}} \frac{1}{r} \lesssim \frac{1}{R}.$$

and therefore the sum in (2.8) can be bounded by a convergent series uniformly in ϵ

$$|\hat{K}_\epsilon(\xi)| \lesssim |\xi| R + \frac{1}{|\xi| R}.$$

Choosing $R = 1/|\xi|$ gives the bound $|\hat{K}_\epsilon(\xi)| \lesssim 1$, uniformly in ϵ which implies that $\hat{K} \in L^\infty$. ☐

As a corollary of Lemmas 2.15 and 1.16 we have

Corollary 2.16. *Let K be a Calderón-Zygmund potential, then for each $f \in \mathcal{S}(\mathbb{R}^n)$, $K \star f$ satisfies the L^2 bound*

$$\|K \star f\|_{L^2} \lesssim \|f\|_{L^2},$$

and can be extended to a bounded linear operator on L^2.

2.5.2 Calderón-Zygmund decomposition

To upgrade the L^2 bound to the weak type bound (2.7), we will need use of a certain decomposition, called (you guessed it) the Calderón-Zygmund decomposition. The decomposition is based on a fundamental covering Lemma using dyadic cubes that splits an integrable function up into regions where it is bounded by a certain number and regions where only the functions average over a certain set of disjoint cubes is controlled.

Lemma 2.17 (The covering Lemma). *Let $f \in L^1$ and $\lambda > 0$, then there exists a countable collection of disjoint cubes $\{Q_k\}$ such that*

1. *for each Q_k we have*
$$\lambda \leq \fint_{Q_k} |f(x)| \mathrm{d}x \leq 2^n \lambda,$$
2. *$|f(x)| \leq \lambda$ almost everywhere on $(\bigcup_k Q_k)^c$.*

Proof. Consider the unit cube $Q^0 = [0,1)^n$ and denote its translate by $m \in \mathbb{Z}^n$ as $Q^m = Q + m$ so that $\{Q^m\}$ form a partition of $\mathbb{R}^n$ into disjoint cubes. Additionally for each $n \in \mathbb{Z}$ we define the family of dyadic cubes $Q^{n,m} = 2^n Q^m$ of width 2^n so that for each fixed $j \in \mathbb{Z}$, $\{Q^{j,m}\}_m$ form a cubic partition of $\mathbb{R}^n$ into cubes of side length 2^j. We call j the dyadic scale of the cube $Q^{j,m}$.

Since f is integrable, we can always find an j large enough such that for each $m \in \mathbb{Z}^n$
$$\fint_{Q^{j,m}} |f(x)| \mathrm{d}x \leq \lambda.$$
Now subdivide each cube at scale j into 2^n sub cubes of scale $j-1$. We select a cube Q at scale $j-1$ if
$$\fint_Q |f(x)| \mathrm{d}x > \lambda \tag{2.9}$$
and call the collection of such cubes S^1. Of the remaining non-selected cubes, we divides further into cubes of scale $j-2$ and select the cubes that satisfy (2.9) and call the collection of these selected cubes $S^{(2)}$. We continue this procedure indefinitely.

The collection of all selected cubes $S = \bigcup_i S^{(i)}$ are clearly disjoint and by construction satisfy $\fint_Q |f(x)| \mathrm{d}x \geq \lambda$ for all $Q \in S$. However, since each selected cube Q is a subset of a non-selected cube Q' of twice the side length, which satisfies $\fint_{Q'} |f(x)| \mathrm{d}x \leq \lambda$, we see that
$$\lambda \leq \fint_Q |f(x)| \mathrm{d}x \leq \frac{1}{|Q|} \int_{Q'} |f(x)| \mathrm{d}x \leq 2^n \fint_{Q'} |f(x)| \mathrm{d}x \leq 2^n \lambda.$$

Now let $\{Q_k\}$ be an enumeration of the selected cubes. If $x \notin \bigcup_k Q_k$, then by construction there is a descending collection of non-selected cubes $\{Q_x^{(j)}\}$, $Q_x^{(j)} \supseteq Q_x^{(j+1)}$, each one half the side length of the previous and each containing x. It follows by the Lebesgue differentiation theorem that for almost every $x \in (\bigcup_k Q_k)^c$ we have
$$f(x) = \lim_{j \to \infty} \fint_{Q_x^{(j)}} |f(y)| \mathrm{d}y.$$
Since each of these cubes $\{Q_x^{(j)}\}$ is not selected, each average is bounded by λ and therefore the limit satisfies $|f| \leq \lambda$ a.e. on $(\bigcup_k Q_k)^c$.

□

Using the above Lemma for a given integrable gives rise to the so-called Calderón-Zygmund decomposition.

Definition 2.18 (Calderón-Zygmund decomposition)**.** Let $f \in L^1$ and $\lambda > 0$ and let $\{Q_j\}$ be the corresponding disjoint cubes from Lemma 2.17. Define

$$g(x) := \begin{cases} f(x) & \text{if } x \notin Q_k \\ ⨍_{Q_k} f(x)\mathrm{d}x & \text{if } x \in Q_k. \end{cases}$$

and let $b = f - g$. Then the decomposition $f = g + b$ is called the Calderón-Zygmund decomposition of f.

The functions g and b in the decomposition are often called the good and bad functions respectively due to the fact that g is controlled in L^∞ by λ, while b captures the unbounded parts of f.

Proposition 2.19. *The Calderón-Zygmund decomposition $f = g + b$, with cubes $\{Q_k\}$ has the following useful properties:*

1. $\|g\|_{L^\infty} \lesssim \lambda$
2. $\|g\|_{L^1} \leq \|f\|_{L^1}$
3. $b = 0$ *on* $(\bigcup_k Q_k)^c$
4. $\int_{Q_k} b \,\mathrm{d}x = 0$
5. $\|b\|_{L^1} \lesssim \|f\|_{L^1}$
6. $|\bigcup_k Q_k| \leq \frac{1}{\lambda}\|f\|_{L^1}$

Proof. There properties are easy to verify and left as an exercise. □

2.5.3 Proof of Theorem 2.14

We are now equipped to prove the Calderón-Zygmund Theorem 2.14.

Proof. Let $f \in S(\mathbb{R}^n)$ and let $f = g + b$ be the Calderón-Zygmund decomposition for $\lambda > 0$ with $\{Q_k\}$ the associated family of disjoint cubes. Write

$$|\{|K \star f| > \lambda\}| \leq |\{|K \star g| > \lambda/2\}| + |\{|K \star b| > \lambda/2\}|.$$

We first estimate the good part by Chebyshev, L^2 boundedness and the L^∞ bound on g,

$$|\{|K \star g| > \lambda/2\}| \lesssim \frac{1}{\lambda^2}\int_{\mathbb{R}^n} |K \star g|^2 \mathrm{d}x \lesssim \frac{1}{\lambda^2}\int_{\mathbb{R}^n} |g|^2 \mathrm{d}x \leq \frac{1}{\lambda}\|g\|_{L^1} = \frac{1}{\lambda}\|f\|_{L^1}$$

To estimate the bad term, we first remove a set of order $\frac{1}{\lambda}\|f\|_{L^1}$. Let y_k denote the center of the cube Q_k and let $Q_k^* = 2\sqrt{n}Q_k$ a scaled up version. Moreover we have

$$\left|\bigcup_k Q_k^*\right| \leq \sum_k |Q_k^*| = (2\sqrt{n}\ell(Q_k))^n \left|\bigcup_k Q_k\right| \lesssim_n \frac{1}{\lambda}\|f\|_{L^1}.$$

Therefore

$$|\{|K \star b| > \lambda/2\}| \leq |\{x \notin \cup_k Q_k^* \,:\, |K \star b(x)| > \lambda/2\}| + \frac{1}{\lambda}\|f\|_{L^1}.$$

The remaining term we estimate via Chebyshev

$$|\{x \notin \cup_k Q_k^* : |K \star b(x)| > \lambda/2\}| \lesssim \frac{1}{\lambda} \int_{\bigcap_k (Q_k^*)^c} |K \star b|(x) \mathrm{d}x.$$

Using that b is supported on $\bigcup_k Q_k$ and has mean zero on each Q_k, we can write the convolution above as

$$(K \star b)(x) = \sum_k \int_{Q_k} (K(x-y) - K(x-y_k)) b(y) \mathrm{d}y.$$

Thus,

$$\begin{aligned} \int_{\bigcap_k (Q_k^*)^c} |K \star b|(x) \mathrm{d}x &\leq \sum_k \int_{(Q_k^*)^c} \int_{Q_k} |K(x-y) - K(x-y_k)||b(y)| \mathrm{d}y \mathrm{d}x \\ &= \sum_k \int_{Q_k} |b(y)| \int_{(Q_k^*)^c} |K(x-y) - K(x-y_k)| \mathrm{d}x \mathrm{d}y. \end{aligned}$$

Note that Q_k and Q_k^* have the property that for each $y \in Q_k$ and $x \notin Q_k^*$ one has $|x-y_k| \geq 2|y-y_k|$ since, by construction, one can fit an annulus whose outer radius is twice as large as its inner radius into the set $Q_k^* \backslash Q_k$. Thus for $y \in Q_k$, denoting $y' = y - y_k$ and $x' = x - y_k$ we have

$$\int_{(Q_k^*)^c} |K(x-y) - K(x-y_k)| \mathrm{d}x \leq \int_{|x'| \geq 2|y'|} |K(x'-y') - K(x')| \mathrm{d}x' \lesssim 1,$$

by the smoothness condition (2.6). It follows that

$$\int_{\bigcap_k (Q_k^*)^c} |K \star b|(x) \mathrm{d}x \lesssim \sum_k \int_{Q_k} |b(x)| \mathrm{d}x \lesssim \|f\|_{L^1},$$

which completes the proof of the weak type inequality.

The proof of the L^p bound now follows from the weak type estimate exactly as in step 2 in the proof of the Hardy-Littlewood maximal inequality Theorem 2.4. □

Corollary 2.20. *The Riesz transform extends to a bounded linear operator on L^p for $p \in (0, \infty)$.*

References

[1] Loukas Grafakos. *Classical Fourier Analysis.* Springer New York.

[2] Lars Hörmander. *The Analysis of Linear Partial Differential Operators I.* Springer Berlin Heidelberg.

[3] Camil Muscalu and Wilhelm Schlag. *Classical and Multilinear Harmonic Analysis Vol. I*, volume 1. Cambridge University Press, January 2013.