LECTURE NOTES 2 FOR 254A

TERENCE TAO

1. Littlewood-Paley projections

This week we will introduce the *Littlewood-Paley decomposition*, which is a very basic way to carve up the phase plane. After describing its basic properties, we shall apply it to a class of function spaces known as *Sobolev spaces*. During this week and the next we shall prove such results as the Sobolev embedding and trace lemmas, the Gagliardo-Nirenberg inequality, and the multiplication laws for Sobolev spaces.

There is a certain amount of flexibility in how one sets up the Littlewood-Paley decomposition[1] on $\mathbf{R}^n$, but one standard way is as follows. Let $\phi(\xi)$ be a real radial bump function supported on $\{\xi \in \mathbf{R}^n : |\xi| \leq 2\}$ which equals 1 on $\{\xi \in \mathbf{R}^n : |\xi| \leq 1$. Let $\psi(t)$ be the function

$$\psi(\xi) := \phi(\xi) - \phi(2\xi).$$

Thus ψ is a bump function supported on the annulus $\{1/2 \leq |\xi| \leq 2\}$. By construction we have

$$\sum_{k \in \mathbf{Z}} \psi(\xi/2^k) = 1$$

for all $\xi \neq 0$. Thus we can partition unity into the functions $\psi(\xi/2^k)$ for integers k, each of which is supported on an annulus of the form $|\xi| \sim 2^k$.

We now define the *Littlewood-Paley projection operators*[2] P_k, $P_{\leq k}$ by

$$\widehat{P_k f}(\xi) = \psi(\xi/2^k)\hat{f}(\xi)$$
$$\widehat{P_{\leq k} f}(\xi) = \phi(\xi/2^k)\hat{f}(\xi).$$

Informally, P_k is a frequency projection to the annulus $\{|\xi| \sim 2^k\}$, while $P_{\leq k}$ is a frequency projection to the ball $\{|\xi| \lesssim 2^k\}$. We shall sometimes write $P_{<k}$ for $P_{\leq k-1}$.

Observe that $P_k = P_{\leq k} - P_{<k}$. Also, if f is an L^2 function then $P_{\leq k} f \to 0$ in L^2 as $k \to -\infty$, and $P_{\leq k} f \to f$ in L^2 as $k \to +\infty$ (this is an easy consequence of

[1] The reader may take $n = 1$ if desired (which is convenient for e.g. drawing phase planes); there is almost no difference between one-dimensional Littlewood-Paley theory and the higher-dimensional theory.

[2] Strictly speaking, these are not quite projections; even though they are self adjoint, they do not quite square to themselves, because we chose ϕ to be a smooth cutoff rather than a rough one. However, the operator $P_k P_k$ is of the same form as P_k, and similarly for $P_{\leq k}$, and so it is still quite reasonable to think of these operators as (smoothed out) projection operators.

Parseval's theorem). By telescoping the series, we thus have the *Littlewood-Paley decomposition*

$$f = \sum_{k \in \mathbf{Z}} P_k f$$

for all $f \in L^2$, where the summation is in the L^2 sense[3]

This decomposition takes a single function and writes it as a superposition of a countably infinite family of functions $P_k f$, each one of which has frequency of magnitude roughly 2^k. Lower values of k represent low frequency components of f; higher values represent high frequency components.

Indeed, from the basic facts about the Fourier transform we have the representation

$$P_{\leq k} f(x) = f * (2^{nk} \hat{\phi}(2^k \cdot)) = \int f(x + 2^{-k} y) \hat{\phi}(y)\ dy. \quad (1)$$

Note that $\hat{\phi}(y)$ is a Schwartz function with total mass $\int \hat{\phi}(y)\ dy = \phi(0) = 1$. Thus the function $P_{\leq k} f$ is an average of f localized to physical scales $\lesssim 2^{-k}$. In particular, we expect $P_{\leq k} f$ to be essentially constant at scales $\ll 2^{-k}$; this is consistent with the uncertainty principle.

What does a single function $P_k f$ look like? Since $P_k f = P_{\leq k+2} P_k f$, we see from (1) that we have the self-reproducing formula

$$P_k f(x) = \int P_k f(x + 2^{-k-2} y) \hat{\phi}(y)\ dy. \quad (2)$$

Thus $P_k f$ is essentially constant at physical scales $\ll 2^{-k}$. On the other hand we have $P_{\leq k-2} P_k f = 0$, so from (1) we have

$$\int P_k f(x + 2^{-k+2} y) \hat{\phi}(y)\ dy = 0 \text{ for all } x \in \mathbf{R}^n.$$

This roughly asserts that $P_k f$ has mean zero at scales $\lesssim 2^{-k+2}$: To reconcile these two properties, we see that on each ball of radius $O(2^{-k})$, the function $P_k f$ is smooth at scales $\ll 2^{-k}$ and contains about $O(1)$ oscillations.

Another indication of oscillation is given by the mean zero property

$$\int P_k f(x)\ dx = \widehat{P_k f}(0) = 0.$$

In fact all moments of $P_k f$ vanish:

$$\int (2\pi i x)^\alpha P_k f(x)\ dx = (\partial_\xi^\alpha \widehat{P_k f})(0) = 0.$$

Later on we shall compare this informal vision of affairs with the rigorous situation of the *Haar system*, which is kind of a Walsh analogue of the Littlewood-Paley decomposition.

[3]Actually, this decomposition works for just about any locally integrable function which has some decay at infinity, and one usually has all the convergence properties of the summation that one needs. In many applications one can make the *a priori* assumption that f is Schwartz, in which case the convergence is (for instance) uniform. However, if f does not decay, then this formula fails. For instance, if $f \equiv 1$, then all the projections $P_k f$ vanish.

The Littlewood-Paley pieces $P_k f$ of f react very well with derivatives. Roughly speaking, we have

$$\nabla P_k f \approx 2^k P_k f. \tag{3}$$

Informally this makes sense, since ∇ is multiplication by $2\pi i\xi$ on the frequency side, and $|\xi| \sim 2^k$ on the support of $P_k f$.

A more precise statement is as follows.

Lemma 1.1. *Let k be an integer, and let f be a function with Fourier support in the annulus $\{2^{k-1} \leq |\xi| \leq 2^{k+1}\}$. Then we have*

$$\|\nabla f\|_p \sim 2^k \|f\|_p$$

for all $1 \leq p \leq \infty$. In particular, we have $\|\nabla P_k f\|_p \sim 2^k \|P_k f\|_p$ for arbitrary f.

Proof We have the analogue of (2):

$$f(x) = \int f(x + 2^{-k-2}y)\hat{\phi}(y)\ dy.$$

Differentiating this we have

$$\nabla f(x) = \int \nabla_x f(x + 2^{-k-2}y)\hat{\phi}(y)\ dy.$$

Integrating by parts, this becomes

$$\nabla f(x) = 2^{k+2} \int f(x + 2^{-k-2}y)\nabla\hat{\phi}(y)\ dy.$$

Since $\nabla\hat{\phi}$ is a bump function, we see from Minkowski's inequality that

$$\|\nabla f\|_p \lesssim 2^k \|f\|_p$$

To obtain the reverse implication, we need to invert ∇. Observe that

$$\phi(\xi/2^{k+2})\widehat{\partial_{x_j} f}(\xi) = 2\pi i \xi_j \hat{f}(\xi)$$

and hence that

$$\hat{f}(\xi) = \sum_{j=1}^{n} \phi(\xi/2^{k+2}) \frac{\xi_j}{2\pi i|\xi|^2} \widehat{\partial_{x_j} f}(\xi).$$

Inverting the Fourier transform, we obtain a way of recovering f from ∇f:

$$f = 2^{-k} \sum_{j=1}^{n} f * K_{k,j} \partial_{x_j} f \tag{4}$$

where

$$\begin{aligned} K_{k,j}(x) &:= 2^k \int \phi(\xi/2^{k+2}) \frac{\xi_j}{2\pi i|\xi|^2} e^{2\pi i x\cdot\xi}\ d\xi \\ &= 2^{nk} \int \phi(\xi/4) \frac{\xi_j}{2\pi i|\xi|^2} e^{2\pi i 2^k x\cdot\xi}\ d\xi. \end{aligned}$$

Replacing everything by absolute values we obtain the bound

$$|K_{k,j}(x)| \lesssim 1.$$

By repeated integration by parts we also obtain the bounds

$$|K_{k,j}(x)| \lesssim |2^k x|^{-s}$$

for all $s > 0$, with the constants depending on s (cf. last week's notes). Thus $K_{k,j}$ is an approximation to the identity at scales 2^{-k}. Thus by Minkowski's inequality again we may obtain

$$\|\nabla f\|_p \gtrsim 2^k \|f\|_p$$

and we are done. ∎

From (3) we have the heuristic relationship

$$\nabla \sim \sum_k 2^k P_k; \tag{5}$$

in other words, one can morally split a derivative as a linear combination of Littlewood-Paley operators. This heuristic asserts that the effect of a derivative on a function f is to accentuate the high frequencies and diminish the low frequencies, since each frequency component $P_k f$ is essentially multiplied by its own frequency 2^k. One can apply similar heuristics to other differentiation operators such as the Laplacian Δ, as well as integration operators such as the inverse Laplacian Δ^{-1}[4].

We now study how the Littlewood-Paley pieces $P_k f$ (or $P_{\leq k} f$) of a function are related to the function itself. Specifically, we are interested in how the L^p behavior of the $P_k f$ relate to the L^p behavior of f.

From (1) and Minkowski's inequality we have (for $1 \leq p \leq \infty$)

$$\|P_{\leq k} f\|_p \leq \int \|f(x + 2^{-k} y)\|_p |\hat{\phi}(y)| \; dy \lesssim \|f\|_p.$$

Thus $P_{\leq k} f$ does not get any bigger than f itself as measured in L^p (or indeed in any translation invariant Banach space). This makes sense, since $P_{\leq k} f$ is an average of f. Similarly $\|P_k f\|_p \lesssim \|f\|_p$. On the other hand, we have $f = \sum_k P_k f$. So by the triangle inequality we thus have the *cheap Littlewood-Paley inequality*

$$\sup_k \|P_k f\|_p \lesssim \|f\|_p \lesssim \sum_k \|P_k f\|_p. \tag{6}$$

This inequality was ridiculously easy to prove, yet it is already quite powerful. Generally it is adequate for one's purposes as long as one has an epsilon number of derivatives to spare somewhere[5]. This comment will seem less cryptic as time goes by, but I'll illustrate one application right away:

[4]However, to get both sides of the analogue of (3) for such operators one needs to assume that the symbol of these operators are *elliptic*. For instance, for a second-order constant co-efficient operator such as $\sum_{j,j'} c_{j,j'} \partial_{x_j} \partial_{x_{j'}}$, one can obtain both sides of the analogue of (3) provided one assumes the ellipticity assumption $|\sum_{j,j'} c_{j,j'} \xi_j \xi_{j'}| \sim |\xi|^2$ for all $\xi \in \mathbf{R}^n$.

[5]It is often said that an inequality such as the cheap Littlewood-Paley inequality is "sharp up to a logarithm". To explain this, suppose that for some reason we only needed to consider frequencies in the range $a \lesssim |\xi| \lesssim A$ for some $0 < a < A < \infty$. For instance a and A could be the lowest and highest physically attainable frequency in the universe. Then the number of dyadic scales k which are relevant is only $O(\log_2(A/a))$. Thus the left and right sides of (6) differ by at most a factor of $O(\log(A/a))$. In many applications in real life, or those in which one can lose an epsilon of derivatives (which corresponds to something like $(A/a)^\varepsilon$, this logarithmic loss is quite acceptable. This principle has been stated rather succinctly by Coifman as "As a first approximation, $\log \infty = O(1)$".

Lemma 1.2 (Non-endpoint Sobolev embedding)**.** *Let $1 \leq p < q \leq \infty$ be such that*

$$\frac{1}{p} - \frac{1}{n} > \frac{1}{q}.$$

Then we have

$$\|f\|_{L^q(\mathbf{R}^n)} \lesssim \|f\|_{L^p(\mathbf{R}^n)} + \|\nabla f\|_{L^p(\mathbf{R}^n)} \tag{7}$$

for all functions f in $\mathbf{R}^n$ for which the right-hand side is finite. (The implicit constant depends on p, q, n).

Proof By the usual limiting arguments we may assume that f is Schwartz. Let's denote the right-hand side of (7) by X. From (6) we have

$$\|P_k f\|_p \lesssim X$$

for all k. We also have

$$\|\nabla P_k f\|_p \lesssim X.$$

By Lemma 1.1 we thus have

$$\|P_k f\|_p \lesssim 2^{-k} X.$$

Combining the two we thus have

$$\|P_k f\|_p \lesssim \min(1, 2^{-k}) X.$$

Note how the first component of X is good at controlling low frequencies $k \leq 0$, while the second component of X is good at controlling high frequencies $k \geq 0$.

Now $P_k f$ has Fourier support in $|\xi| \sim 2^k$. Since $q > p$, we can then apply Bernstein's inequality (from last week's notes) to obtain

$$\|P_k f\|_q \lesssim 2^{(\frac{1}{p} - \frac{1}{q})nk} \min(1, 2^{-k}) X.$$

Applying the cheap Littlewood-Paley inequality (6) again we obtain

$$\|f\|_q \lesssim \sum_k 2^{(\frac{1}{p} - \frac{1}{q})nk} \min(1, 2^{-k}) X.$$

If you take a plot of $2^{(\frac{1}{p} - \frac{1}{q})nk} \min(1, 2^{-k})$ as a function of k, it reaches a maximum of 1 at $k = 0$ and decays exponentially as $k \to \pm\infty$ because of our assumptions on p, q. So we can sum this and obtain the desired conclusion. ∎

The above proof uses a very useful strategy in proving any sort of estimates involving derivatives: first prove it for a single Littlewood-Paley piece $P_k f$ (i.e. assume all frequencies ξ are of magnitude $\sim 2^k$). If you can't even prove it for a single piece, there is very little chance you can prove it for the whole function f. Once you see what's going on for a single piece, add things up (either by (6), or by more sophisticated estimates). This method works surprisingly well in a large number of situations (see e.g. the exercises for this week).

The Sobolev embedding theorem also holds at the endpoint $\frac{1}{p} - \frac{1}{n} = \frac{1}{q}$ (as long as $p \neq 1$ and $q \neq \infty$), and in this case we don't need the $\|f\|_{L^p}$ component:

Lemma 1.3 (Endpoint Sobolev embedding). *Let $1 < p < q < \infty$ be such that*
$$\frac{1}{p} - \frac{1}{n} = \frac{1}{q}.$$
Then we have
$$\|f\|_{L^q(\mathbf{R}^n)} \lesssim \|\nabla f\|_{L^p(\mathbf{R}^n)} \tag{8}$$
for all functions f in $\mathbf{R}^n$ for which the right-hand side is finite. (The implicit constant depends on p, q, n).

Proof We first prove the weak-type estimate
$$\|f\|_{L^{q,\infty}(\mathbf{R}^n)} \lesssim \|\nabla f\|_{L^p(\mathbf{R}^n)} \tag{9}$$
and show how to bootstrap this to the strong-type estimate at the end. By scaling we may take $\|\nabla f\|_p = 1$.

We wish to prove that
$$|\{|f| > \lambda\}| \lesssim \lambda^{-q}$$
for all $\lambda > 0$. We follow the Littlewood-Paley philosophy, and begin with the problem of estimating
$$|\{|P_k f| > \lambda\}|$$
for some integer k. First of all, by Chebyshev we have
$$|\{|P_k f| > \lambda\}| \lesssim \frac{\|P_k f\|_p^p}{\lambda^p} \sim \frac{(2^{-k}\|\nabla P_k f\|_p)^p}{\lambda^p} \lesssim 2^{-kp}\lambda^{-p}.$$
On the other hand, by Bernstein's inequality we also have
$$\|P_k f\|_\infty \lesssim 2^{nk/p}\|P_k f\|_p \lesssim 2^{nk/p}2^{-k}. \tag{10}$$
Thus $\{|P_k f| > \lambda\}$ vanishes unless $\lambda \lesssim 2^{nk/p}2^{-k}$. But then we have
$$\begin{aligned} 2^{-kp}\lambda^{-p} &\lesssim 2^{-kp}\lambda^{-q}(2^{nk/p}2^{-k})^{q-p} \\ &= \lambda^{-q} \end{aligned}$$
by the definition[6] of q. Thus we can prove the weak-type inequality for a single Littlewood-Paley piece $P_k f$.

To prove the weak-type inequality for the entire function f, we note in the above argument that for a fixed λ there is a critical value k_λ of k, defined (up to an $O(1)$ error) by
$$\lambda \sim 2^{nk_\lambda/p}2^{-k_\lambda}.$$
From our hypotheses we have $p < n$, so that the right-hand side of (10) is increasing in k. We can therefore sum and obtain
$$\|P_{\leq k_\lambda - C} f\|_\infty < \lambda/2 \tag{11}$$
for some sufficiently large constant C. On the other hand, we have
$$\|P_k f\|_p \lesssim 2^{-k}$$

[6] It may seem like magic that the powers of k magically cancelled, but this is to be expected by scaling or dimensional analysis considerations. If one gives physical space the units of *length*, so that frequency space has the units of *length*$^{-1}$, then the relationship $1/q = 1/p - 1/n$ is forced by dimensional analysis.

so by summing we obtain

$$\|P_{>k_\lambda - C} f\|_p \lesssim 2^{-k_\lambda}$$

and so by Chebyshev and the previous computation

$$|\{|P_{>k_\lambda - C} f| \geq \lambda/2\}| \lesssim 2^{-k_\lambda p} \lambda^{-p} \sim \lambda^{-q}.$$

Combining this with (**??**) we obtain the weak-type estimate.

Now to get the strong-type estimate from the weak-type estimate. In principle this is immediate from the Marcinkiewicz interpolation theorem, but the ∇ on the right makes things a little tricky. Morally one should be able to rewrite (9) as

$$\|\nabla^{-1} g\|_{L^{q,\infty}(\mathbf{R}^n)} \lesssim \|g\|_{L^p(\mathbf{R}^n)}$$

where $g = \nabla f$, in which case the Marcinkiewicz interpolation theorem does apply, however there is a catch because ∇ is not surjective. To get around this we need to extend the ∇^{-1} operator to all n-tuples of functions (not just those which are gradients). This can be done e.g. by the operator $\Delta^{-1}\nabla\cdot$, which inverts ∇. By repeating the above arguments we see that

$$\|\Delta^{-1}\nabla \cdot g\|_{L^{q,\infty}(\mathbf{R}^n)} \lesssim \|g\|_{L^p(\mathbf{R}^n)}$$

for all n-tuples of Schwartz functions g, and so by interpolation we can bootstrap the weak-type estimate to a strong type estimate, and we are done. ∎

As the name suggests, the cheap Littlewood-Paley inequality is not the sharpest statement one can make connecting the L^p norms of $P_k f$ with those of f. One can already see this when $p = 2$, in which case we have

$$\|f\|_2 \sim (\sum_k \|P_k f\|_2^2)^{1/2} \tag{12}$$

(why is this stronger than (6)?). To see this, square both sides and take Plancherel to obtain

$$\int |\hat{f}(\xi)|^2 \, d\xi \sim \sum_k \int |\psi(\xi/2^k)|^2 |\hat{f}(\xi)|^2 \, d\xi.$$

Observe that for each $\xi \neq 0$ there are only two values of $\psi(\xi/2^k)$ which does not vanish, and that these two add up to 1. The claim follows.

Another way to rewrite (12) is

$$\|f\|_2 \sim \|(\sum_k |P_k f|_2^2)^{1/2}\|_2. \tag{13}$$

The quantity $(\sum_k |P_k f|_2^2)^{1/2}$ is known as the *Littlewood-Paley square function* and is sometimes denoted $|Sf|$. I'll reserve the notation Sf for the vector-valued function

$$Sf := (P_k f)_{k \in \mathbf{Z}};$$

thus $|Sf|$ is the magnitude of Sf in the l^2 norm.

We can now give the true *Littlewood-Paley inequality*:

Proposition 1.4 (Littlewood-Paley inequality). *For any $1 < p < \infty$, we have*

$$\|Sf\|_p \sim \|f\|_p$$

with the implicit constant depending on p.

In order to prove Proposition 1.4 we will need to quickly review some Calderón-Zygmund theory.

Definition 1.5. A *Calderón-Zygmund operator* T is a linear operator[7] on $\mathbf{R}^n$ of the form

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

for some (possibly distributional, possibly matrix valued) kernel K which obeys the bounds

$$|K(x,y)| \lesssim \frac{1}{|x-y|^n} \tag{14}$$

and

$$|\nabla K(x,y)| \lesssim \frac{1}{|x-y|^{n+1}} \tag{15}$$

for all $x \neq y$. We also require that T be bounded on L^2.

The function $1/|x|^n$ is just barely non-integrable on $\mathbf{R}^n$, so that T just barely fails to automatically be bounded on L^2 (or in fact on L^p, $1 \leq p \leq \infty$). Indeed, if $1/|x|^n$ were integrable, then we would have

$$\|Tf\|_p \lesssim \| \int |f(x-y)|/|y|^n\ dy\|_p \lesssim \int \|f\|_p/|y|^n\ dy \lesssim \|f\|_p.$$

Because $1/|x|^n$ is just barely singular, an operator such as T is often referred to as a *singular integral.*

The classic example of a Calderón-Zygmund kernel is the *Hilbert transform*

$$Hf(x) := p.v. \int \frac{1}{x-y} f(y)\ dy$$

on $\mathbf{R}$. This operator is a Fourier multiplier with symbol $-\pi i \operatorname{sgn}(\xi)$ (exercise!), so is bounded on L^2 by Plancherel.

A basic result of Calderón and Zygmund is that

Lemma 1.6. *If T is a Calderón-Zygmund operator, then T is bounded on all L^p, $1 < p < \infty$.*

We defer the proof of this lemma to an Appendix, and return to the proof of Proposition 1.4. We observe that S is a vector-valued Calderón-Zygmund operator with vector-valued kernel

$$K(x,y) := (2^{nk} \hat{\psi}(2^k(x-y)))_{k \in \mathbf{Z}}.$$

[7] There are more general notions of a Calderón-Zygmund kernel, but we shall not need them in this course.

Indeed, the decay properties (14), (15) are easily verified since $\hat{\psi}$ is a Schwartz function. The L^2 boundedness properties come from (13). From Lemma 1.6 we thus have one half of Proposition 1.4, namely that

$$\|Sf\|_p \lesssim \|f\|_p$$

for all $1 < p < \infty$. By duality we also have

$$\|S^* f\|_p \lesssim \|f\|_p.$$

By untangling this, we see that

$$\|\sum_k P_k f_k\|_p \lesssim \|(\sum_k |f_k|^2)^{1/2}\|_p.$$

Similarly we have

$$\|\sum_k \tilde{P}_k f_k\|_p \lesssim \|(\sum_k |f_k|^2)^{1/2}\|_p$$

where $\tilde{P}_k := P_{\leq k+2} - P_{\leq k-2}$.

We apply this with $f_k := P_k f$. Since $\tilde{P}_k P_k = P_k$, we obtain

$$\|f\|_p \lesssim \|(\sum_k |P_k f|^2)^{1/2}\|_p$$

which is the other half of Proposition 1.4.

The Littlewood-Paley inequality can be viewed as an assertion of the form

$$|\sum_k P_k f| \approx (\sum_k |P_k f|^2)^{1/2}.$$

Note that such a claim would be false if all the $P_k f$ were scalar multiples of each other. The point is that the $P_k f$ behave very differently (each $P_k f$ oscillates at a frequency $\sim 2^k$). There is a general principle which asserts that if a collection of functions $f_1, \ldots, f_k$ oscillate at very different rates (so that the phase of one function f_i is essentially independent of the phase of any other function), then $|\sum_i f_i| \approx (\sum_i |f_i|^2)^{1/2}$. A typical example of this principle appears in *Khinchin's inequality*, which we will not discuss in this course. The principle is also related to the *law of large numbers* or *central limit theorem* from probability theory.

Some easy consequences of the above theory are as follows. Firstly, if for each k, the operator $\tilde{P}_k$ was some smooth frequency localization operator[8], then

$$\|(\sum_k |\tilde{P}_k f|^2)^{1/2}\|_p \lesssim \|f\|_p \tag{16}$$

for all $1 < p < \infty$. A dual statement is that

$$\|\sum_k \tilde{P}_k f_k\|_p \lesssim \|(\sum_k |f_k|^2)^{1/2}\|_p \tag{17}$$

for all $1 < p < \infty$ and arbitrary functions f_k. (In other words, when adding functions of widely different frequencies, they add as if they square-summed). We leave these statements as exercises. One quick consequence is

[8] In other words, $\tilde{P}_k$ is given by a bump function adapted to the annulus $\{|\xi| \sim 2^k\}$. More informally, $\tilde{P}_k$ looks like P_k.

Proposition 1.7 (Hörmander-Mikhlin multiplier theorem). *Let $m(\xi)$ be a multiplier such that*

$$|\nabla^j m(\xi)| \lesssim |\xi|^j \tag{18}$$

for all $j \geq 0$ (with the implicit constant depending on j). Let T_m be the Fourier multiplier with symbol m:

$$\widehat{T_m f}(\xi) = m(\xi) f(\xi)$$

then T_m is bounded on L^p for all $1 < p < \infty$.

Proof We have

$$T_m = \sum_{k,k'} P_k T_m P_{k'} = \sum_k \tilde{P}_k \tilde{\tilde{P}}_k$$

where $\tilde{P}_k := P_k T_m$ and $\tilde{\tilde{P}}_k := \sum_{k-2<k'<k+2} P_k$. From (18) we see that $\tilde{P}_k$ and $\tilde{\tilde{P}}_k$ are smooth frequency localization operators to the annulus $\{|\xi| \sim 2^k\}$. The claim then follows by composing (17) (for $\tilde{P}_k$) with (16) (for $\tilde{\tilde{P}}_k$). ■

A typical consequence of the Hörmander-Mikhlin multiplier theorem is the estimate

$$\|\partial_{x_i}\partial_{x_j} f\|_p \lesssim \|\Delta f\|_p$$

for all $1 < p < \infty$, where $\Delta = \sum_{i=1}^n \partial_{x_i}^2$ is the Laplacian on $\mathbf{R}^n$. This follows because

$$\widehat{\partial_{x_i}\partial_{x_j} f}(\xi) = \frac{\xi_i \xi_j}{|\xi|^2} \widehat{\Delta f}(\xi)$$

and the symbol $m(\xi) := (\xi_i \xi_j)/|\xi|^2$ satisfies the conditions (18). This particular estimate is an example of the principle known as *elliptic regularity*: once one controls an elliptic operator[9] such as Δ, one automatically controls all other derivative operators of the same strength.

2. Sobolev spaces

Littlewood-Paley theory is especially good for dealing with spaces which combine L^p type norms with derivatives; this is basically because of the heuristic decomposition (5). The most commonly used family of such spaces are the *Sobolev spaces*[10] $W^{s,p}$. (There are also Besov spaces, Hölder spaces, and Triebel-Lizorkin spaces, and then there are closely related spaces such as Hardy spaces and BMO (although strictly speaking, these spaces do not contain derivatives). We'll discuss them briefly at the end of these notes).

[9] This begs the question as to what it means for an operator to be elliptic. A circular (but not entirely facetious) answer is that an operator is elliptic if it has some sort of elliptic regularity estimate. A slightly better answer would be that the symbol of the operator (when thought of as a Fourier multiplier) is positive definite (so that its level sets are ellipsoids, at least in the constant co-efficient second order case).

[10] These spaces are also called L^p_s, $H^{s,p}$, or $W^{p,s}$ in the literature. When $p = 2$ these spaces are often just called H^s. Unfortunately no one individual has had enough clout to standardize the notation.

The spaces $W^{s,p}$ are defined for $s \in \mathbf{R}$ and $1 < p < \infty$. (One can also define them for $p = \infty$ or $0 < p \leq 1$ but some technical details come up that I'd rather not discuss). When $s \geq 0$ is a non-negative integer, the spaces can be easily defined by

$$\|f\|_{W^{s,p}} := \sum_{j=0}^{s} \|\nabla^j f\|_p.$$

In other words, $W^{s,p}$ are those functions such that f and the first s derivatives of f are in L^p. The exponent p thus measures integrability while s measures regularity. When $s = 0$ we of course have $W^{0,p} = L^p$.

These spaces are ubiquitous in the study of PDE, both linear and non-linear. For instance, many elliptic PDE arise from minimizing functionals such as

$$\int |f(x)|^2 + |\nabla f(x)|^2 \, dx,$$

which is essentially the $W^{1,2}$ norm.

Before we define these spaces for non-integer s, we first try to rewrite the above definition in terms of Littlewood-Paley operators P_k rather than derivatives ∇. We begin by observing that

Lemma 2.1. *For any $j \geq 0$ and $1 < p < \infty$, we have*

$$\|\nabla^j f\|_p \sim \| \sum_k |2^{jk} P_k f|^2)^{1/2}\|_p.$$

Note that this lemma is consistent with Proposition 1.4 and (3).

Proof By a little Fourier analysis we may write

$$2^{jk} P_k f = \tilde{P}_k \nabla^j f$$

where $\tilde{P}_k$ is an operator similar to P_k (i.e. a Fourier multiplier whose symbol is a bump function adapted to $|\xi| \sim 2^k$). By (16) and Proposition 1.4 we thus have

$$\| \sum_k |2^{jk} P_k f|^2)^{1/2}\|_p = \| \sum_k |\tilde{P}_k \nabla^j f|^2)^{1/2}\|_p \lesssim \|\nabla^j f\|_p$$

which is one direction of the implication. To go the other way we use a variant identity

$$P_k \nabla^j f = 2^{jk} \tilde{\tilde{P}}_k P_k f$$

where $\tilde{\tilde{P}}_k$ is another operator similar to P_k. By (17) we have

$$\|\nabla^j f\|_p = \| \sum_k 2^{jk} \tilde{\tilde{P}}_k P_k f\|_p \lesssim \|(\sum_k |2^{jk} P_k f|^2)^{1/2}\|_p$$

as desired. ∎

From this lemma we easily see that

$$\|f\|_{W^{s,p}} \sim \| \sum_k |(1 + 2^k)^s P_k f|^2)^{1/2}\|_p.$$

Note that the right-hand side makes sense for all s, not just integer s. Because of this, it is natural to use the right-hand side to *define* the Sobolev space $W^{s,p}$ for all s. (Of course, because we have a $\sim$ rather than an equality, this norm is only defined up to a constant. This, however, is good enough in practice; in most applications one is prepared to lose constant factors all over the place.)

Fractional Sobolev spaces are related to the notion of *fractional derivatives*. Recall that

$$\widehat{\nabla f}(\xi) = 2\pi i \xi \hat{f}(\xi).$$

Because of this, it is natural to define the operator $|\nabla|$ by

$$\widehat{|\nabla| f}(\xi) := 2\pi|\xi|\hat{f}(\xi)$$

and more generally $|\nabla|^s$ for arbitrary $s \in \mathbf{R}$ by

$$\widehat{|\nabla|^s f}(\xi) := (2\pi|\xi|)^s \hat{f}(\xi).$$

These operators are not immediately intuitive. When $s = 2$ we have $|\nabla|^2 = -\Delta$. More generally we can use the functional calculus of the Laplacian Δ as $|\nabla|^s = \sqrt{-\Delta}^s$ (exercise!). For $-n < s < 0$, $|\nabla|^s$ is known as an *fractional integration operator*, and has the convolution kernel representation

$$|\nabla|^{-s} f(x) = C_{n,s} \int \frac{f(y)dy}{|x-y|^{n-s}}$$

for some constant $C_{n,s}$ depending on n, s (exercise! Note that this representation is consistent with dilations and rotations). This kernel is sometimes known as a *Riesz potential*, although this is confusing as Riesz potentials are distinct from Riesz transforms.

In analogy to (3) we have the heuristic

$$|\nabla|^s P_k f \approx 2^{ks} P_k f;$$

in other words, $|\nabla|^s$ accentuates the k^{th} frequency piece of f by a factor of 2^{ks}.

One can make this heuristic rigorous in several ways. For instance, by repeating the proof of Lemma 2.1 we have

$$\| |\nabla|^s f\|_p \sim \| \sum_k |2^{sk} P_k f|^2)^{1/2}\|_p.$$

As a consequence we see that

$$\|f\|_{W^{s,p}} \sim \|f\|_p + \||\nabla|^s f\|_p$$

for all $s \geq 0$. (When $s < 0$ the formula is a little different; in fact we have

$$\|f\|_{W^{s,p}} \sim \inf_{f=g+h} \|g\|_p + \||\nabla|^s h\|_p$$

(exercise!). Also, for any $s \in \mathbf{R}$ and $\sigma > 0$: we see that the operator $|\nabla|^\sigma$ removes σ degrees of regularity:

$$\||\nabla|^\sigma f\|_{W^{s-\sigma,p}} \lesssim \|f\|_{W^{s,p}}$$

(exercise!). Indeed, if D^k is any constant-coefficient operator of order $k > 0$, then

$$\|D^k f\|_{W^{s-k,p}} \lesssim \|f\|_{W^{s,p}}.$$

Because of these facts, Sobolev spaces are very good at tracking the regularity of various components of a PDE.

One can connect the Sobolev spaces of a given integrability exponent p more tightly by introducing modified fractional differentiation operators $\langle\nabla\rangle^s$ by

$$\widehat{\langle\nabla\rangle^s f}(\xi) := \langle 2\pi\xi\rangle^s \hat{f}(\xi)$$

where the *Japanese bracket* $\langle x\rangle$ is defined by $\langle x\rangle := (1+|x|^2)^{1/2}$. In terms of the Laplacian, we have $\langle\nabla\rangle^s = \sqrt{1-\Delta}^s$. The operator $\langle\nabla\rangle^s$ is essentially the same as $|\nabla|^s$ for high frequencies $|\xi| \gtrsim 1$, but behaves like the identity for low frequencies $|\xi| \lesssim 1$ (unlike $|\nabla|^s$, which damps low frequencies when $s > 0$ and accentuates them for $s < 0$). When s is negative, the operators $\langle\nabla\rangle^s$ are known as *smoothing operators*, and the kernels are known as *Bessel potentials*. (Despite the name, these potentials are only tenuously connected with Bessel functions).

The reader may verify using Littlewood-Paley theory that one has the estimates

$$\|f\|_{W^{s,p}} \sim \|\langle\nabla\rangle^s f\|_p.$$

In other words, the Sobolev space $W^{s,p}$ is just the images of L^p under the operator $\langle\nabla\rangle^{-s}$. As a consequence, any two Sobolev spaces $W^{s,p}$ and $W^{s',p}$ are isomorphic via the map $\langle\nabla\rangle^{s'-s}$. (One can now use this definition to define Sobolev spaces for $p = 1$ or $p = \infty$). As another corollary we see that the spaces $W^{s,p}$ and $W^{-s,p'}$ are dual for $1 < p < \infty$, where p' is defined by $1/p + 1/p' = 1$.

When $p = 2$ we can thus use Plancherel's theorem to obtain a very convenient Fourier representation of a Sobolev space:

$$\|f\|_{W^{s,2}} \sim (\int (\langle\xi\rangle^s |f(\xi)|)^2 \, d\xi)^{1/2}.$$

Because of this, the $p = 2$ Sobolev spaces can be manipulated directly without recourse to Littlewood-Paley theory. However, once one has the Littlewood-Paley toolbox in one's possession, then the $p \neq 2$ spaces are not much more complicated than the $p = 2$ spaces.

The Sobolev embedding theorems in Lemma 1.2 and Lemma 1.3 can be generalized to fractional spaces; indeed, these embedding theorems are one of the major reasons why we care about fractional Sobolev spaces in the first place. We shall simply state the theorems here, as they are proven in an almost identical fashion to their integer analogues.

Lemma 2.2 (Non-endpoint Sobolev embedding). *Let $1 \leq p < q < \infty$ and $s' < s$ be such that*

$$\frac{1}{p} - \frac{s-s'}{n} > \frac{1}{q}.$$

Then we have

$$\|f\|_{W^{q,s}(\mathbf{R}^n)} \lesssim \|f\|_{L^{p,s'}(\mathbf{R}^n)} \tag{19}$$

for all functions f in $\mathbf{R}^n$ for which the right-hand side is finite. (The implicit constant depends on p, q, n, s, s').

Lemma 2.3 (Endpoint Sobolev embedding). *Let $1 < p < q < \infty$ and $s' < s$ be such that*

$$\frac{1}{p} - \frac{s-s'}{n} = \frac{1}{q}.$$

Then we have

$$\|f\|_{W^{q,s}(\mathbf{R}^n)} \lesssim \|\nabla f\|_{L^p(\mathbf{R}^n)} \quad (20)$$

for all functions f in $\mathbf{R}^n$ for which the right-hand side is finite. (The implicit constant depends on p, q, n, s, s').

One should think of these Sobolev embedding theorems as tools which allow one to trade in regularity[11] for a gain in integrability. (One cannot reverse this and trade in integrability for regularity; even if a function is in the best possible integrability class, namely L^∞, there is no reason why it should have any differentiability properties whatsoever.)

In next week's notes we shall start to study the multilinear and non-linear properties of Sobolev spaces. For instance, we shall see how the Sobolev class of f and g control the Sobolev class of fg.

3. Other spaces

The Sobolev spaces $W^{s,p}$ are good examples of spaces with regularity s and integrability p, however, there are a couple other spaces with similar exponents s, p which are useful in applications. We will not use them here, but just give a very quick tour of them.

The Sobolev spaces $W^{s,p}$ can be generalized to the *Triebel-Lizorkin spaces*[12] $F_q^{s,p}$, defined for $s \in \mathbf{R}$, $0 < p < \infty$, and $1 \leq q \leq \infty$ by

$$\|f\|_{F_q^{s,p}} := \|(\sum_q |(1+2^k)^s P_k f|^q)^{1/q}\|_p.$$

Thus $F_2^{s,p} = W^{s,p}$ when $1 < p < \infty$. When $0 < p \leq 1$ the space $F_2^{0,p}$ is equal to the *Hardy space* H^p, which plays an important role in complex analysis, and is in many ways a preferable space to L^p in this range of exponents. The exponent q only influences things by a logarithm or so (see the footnote to (6)) and is of minor importance compared to s and p. Thus if logarithms are unimportant then we may pretend that $F_q^{s,p} \approx W^{s,p}$.

[11] Remember that, ultimately, all these Sobolev embedding theorems are based on Bernstein's inequality, which allows one to exploit frequency localization to gain integrability. Regularity can be considered as a form of frequency localization (the smoother a function is, the faster its Fourier transform decays in frequency). The physical space analogue of this is *Hölder's inequality*: if a function is compactly supported in physical space, then higher L^p norms control lower L^q norms. More generally, one use Hölder to trade in decay for a lowering of integrability.

[12] As with the Sobolev spaces, these spaces unfortunately do not have a standardized notation.

In light of (6) it is natural to also consider the *Besov spaces* $B^{s,p}_q$ defined by

$$\|f\|_{B^{s,p}_q} := (\sum_q |(1+2^k)^s \|P_k f\|_p|^q)^{1/q}.$$

Thus, for instance, (6) can be rewritten as

$$\|f\|_{B^{0,p}_\infty} \lesssim \|f\|_p \lesssim \|f\|_{B^{0,p}_1}.$$

These spaces are handy for a number of applications, for instance one can often recover endpoint Sobolev inequalities for Besov spaces at the endpoints $p = 1$ or $q = \infty$, which are not available for ordinary Sobolev spaces. However, the differences are again only logarithmic, and if logarithms are unimportant then $B^{s,p}_q \approx W^{s,p}$.

When s is not an integer, the spaces $B^{s,p}_q$ are equivalent to the more classical *Hölder spaces* $\Lambda^{s,p}_q$. An example of this is given in the exercises.

One space which is important, but not listed in these families, is the space BMO of functions of bounded mean oscillation. In other words, BMO consists of those functions f for which the norm

$$\|f\|_{BMO(\mathbf{R}^n)} := \sup_B \frac{1}{B} \int_{|B|} |f - f_B|$$

is finite, where B ranges over all balls in $\mathbf{R}^n$ and $f_B := \frac{1}{|B|}\int_B f$ is the mean of f on B. This space often turns up as a substitute for L^∞ (we don't ask that the function be bounded everywhere, but only that its deviation from its mean stays bounded on the average). Clearly $\|f\|_{BMO} \lesssim \|f\|_\infty$, but the reverse inequality fails in general (e.g. $f(x) := \log|x|$ is in BMO but not L^∞. However, we have

Lemma 3.1. *If f has frequency support on a single annulus $\{|\xi| \sim 2^k\}$ then we have the relationship*

$$\|f\|_{BMO} \sim \|f\|_\infty.$$

Proof By scaling we may take $k = 0$. It suffices to show that $|f(x)| \lesssim \|f\|_{BMO}$ for all $x \in \mathbf{R}^n$.

Fix x. By repeating the proof of (2) we have the reproducing formula

$$f(x) = \int f(x-y) 2^{nC} \hat\phi(2^{-C} y)\, dy$$

for sufficiently large C. We also have

$$0 = \int f(x-y) 2^{-nC} \hat\phi(2^C y)\, dy,$$

so that

$$f(x) = \int f(x-y)(2^{nC}\hat\phi(2^{-C}y) - 2^{-nC}\hat\phi(2^C y))\, dy.$$

Let B be the ball $B := \{y : |x-y| \leq 2^{C'}\}$, where C' is a constant to be chosen later. Clearly we have

$$0 = \int f_B (2^{nC}\hat\phi(2^{-C}y) - 2^{-nC}\hat\phi(2^C y))\, dy.$$

Thus

$$f(x) = \int (f(x-y) - f_B)(2^{nC}\hat{\phi}(2^{-C}y) - 2^{-nC}\hat{\phi}(2^C y))\ dy.$$

We split the integral into the regions $y \in B$ and $y \notin B$. On the region $y \in B$ the integral is $O(\|f\|_{BMO})$, where the implicit constant depends on C and C'. On the region $y \notin B$, we can estimate the integral by

$$2\|f\|_\infty \int_{|y|>C'} (2^{nC}\hat{\phi}(2^{-C}y) - 2^{-nC}\hat{\phi}(2^C y))\ dy.$$

If C' is sufficiently large depending on C, this integral is $< 1/4$. Thus we have

$$|f(x)| \leq O(\|f\|_{BMO}) + \frac{1}{2}\|f\|_\infty.$$

If we take suprema over x, the claim follows. ∎

Unfortunately BMO does not have as nice a Littlewood-Paley representation as the other spaces listed here (even though it does turn out to be the dual of $H^1 = F_2^{0,1}$).

Before we finish this quick tour of function spaces, we should mention that these "inhomogeneous" spaces have "homogeneous" or "pure" counterparts, in which only one order of derivative is used. For instance, the homogeneous Sobolev space $\dot{W}^{s,p}$ is defined by

$$\|f\|_{\dot{W}^{s,p}} \sim \|\sum_k |(2^k)^s P_k f|^2)^{1/2}\|_p$$

instead of

$$\|f\|_{W^{s,p}} \sim \|\sum_k |(1+2^k)^s P_k f|^2)^{1/2}\|_p.$$

Similarly one can define $\dot{F}_q^{s,p}$ and $\dot{B}_q^{s,p}$. These homogenous spaces can be a little harder to work with (for instance, one has far fewer Sobolev embeddings for these spaces) but they scale much more nicely, and there are many situations in PDE in which the homogeneous norm is the only norm which one actually has control on. For high frequencies $|\xi| \gtrsim 1$ there is essentially no distinction between the homogeneous and inhomogeneous spaces, but the low frequencies $|\xi| \lesssim 1$ are measured quite differently (one or the other space is stronger, depending on the sign of s).

4. Appendix: Proof of Lemma 1.6

To prove Lemma 1.6 it suffices by duality to check the case $1 < p < 2$. (Individual Calderón-Zygmund operators need not be self-adjoint, but the *class* of Calderón-Zygmund operators is self-adjoint). By the Marcinkiewicz interpolation theorem it suffices to show that T is weak-type (1,1), i.e. that

$$|\{|Tf| > \lambda\}| \lesssim \frac{\|f\|_1}{\lambda}. \tag{21}$$

To prove this we use a basic philosophy of Calderón-Zygmund theory, namely that, on the average, Tf is only large when the Hardy-Littlewood maximal function

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

is large (here $B(x,r)$ denotes the ball of radius r centered at x). From the definition, Mf is clearly larger than f (consider the limit as $r \to 0$), however the difference between Mf and f is really not that large, and as a first approximation one can assume $Mf \approx f$. For instance, from the Hardy-Littlewood maximal inequality, M is bounded on L^p for all $1 < p \leq \infty$, and is also weak-type (1,1)[13].

We now return to the proof of (21). The idea is to split f into a "good" part g, which is in L^2 and thus amenable to the L^2 boundedness assumption of T, and a "bad" part $\sum_I b_I$, which is not in L^2 but is well localized to a collection I of intervals, and also has a mean zero condition which can be exploited via the smoothness condition (15). This decomposition is known as the *Calderón-Zygmund decomposition.*

Fix f (which we can take to be Schwartz) and λ. To avoid irrelevant technicalities we shall only prove the one-dimensional case $n = 1$. Since f is Schwartz, the set $\{Mf > \lambda\}$ is the union of a finite number of disjoint intervals $\{I\}$. From the previous heuristic these intervals are where we expect Tf to be large.

On each of these intervals we have

$$\int_I |f| \sim \lambda |I|, \tag{22}$$

so we have

$$\lambda \sum_I |I| \lesssim \|f\|_1. \tag{23}$$

We split $f = g + \sum_I b_I$, where

$$g := f\chi_{\mathbf{R} - \bigcup_I I} + \sum_I f_I \chi_I$$

and

$$b_I := (f - f_I)\chi_I,$$

and $f_I := \frac{1}{I}\int_I f$ is the mean of f on I. Thus b_I is the mean-zero part of f on I, and g is everything else.

We shall estimate the left-hand side of (21) by

$$|\{|Tg| \geq \lambda/2\}| + |\{|\sum_I Tb_I| \geq \lambda/2\}|.$$

The first term we can estimate using Chebyshev's inequality by

$$\lesssim \lambda^{-2}\|Tg\|_2^2 \lesssim \lambda^{-2}\|g\|_2^2 \lesssim \lambda^{-2}\|g\|_1\|g\|_\infty.$$

[13]This is an easy exercise in one dimension, by a modification of the arguments below. In general dimension it is a consequence of the Vitali covering lemma. See e.g. E. M. Stein, *Singular integrals*, Princeton University Press.

From (22) we see that $\|b_I\|_1 \lesssim \lambda|I|$. Since $f = g + \sum_I b_I$, we thus see from (23) that $\|g\|_1 \lesssim \|f\|_1$. Also, from (22) and the construction of the I we see that $\|g\|_\infty \lesssim \lambda$. Thus this term is acceptable.

Now to control the second term $|\{|\sum_I Tb_I| \geq \lambda/2\}|$. We expect[14] the function $\sum_I Tb_I$ to be pretty nasty in the *exceptional set* $\bigcup_I 2I$, but fortunately this set has small size by (23) and we can discard it. The function b_I is supported in I and has mean zero, so by (15) and the mean-value theorem we have the estimates

$$\begin{aligned}
|Tb_I(x)| &= |\int_I b_I(y)K(x,y)\ dy| \\
&= |\int_I b_I(y)(K(x,y) - K(x,y_I))\ dy| \\
&\lesssim \int_I |b_I||y-y_I|/|x-y_I|^2\ dy \\
&\lesssim |I|\|b_I\|_1/\mathrm{dist}(x,I)^2 \\
&\lesssim \lambda|I|^2/\mathrm{dist}(x,I)^2
\end{aligned}$$

whenever $x \notin 2I$. Thus, outside of the exceptional set we have

$$\|Tb_I\|_1 \lesssim \lambda|I|$$

so by summing using (23) we have

$$\|\sum_I Tb_I\|_1 \lesssim \|f\|_1$$

and so this part is also OK by Chebyshev's inequality. This proves (21), and so Calderón-Zygmund operators are bounded in L^p, $1 < p < \infty$.

5. Homework for week 2 - due Friday, January 26

In all these problems, and $n \geq 1$ is an integer, and P_k denotes some family of Littlewood-Paley operators on $\mathbf{R}^n$. k ranges over the integers $\mathbf{Z}$ unless otherwise indicated. Implicit constants are allowed to depend on the parameters p, n, α, q, θ.

- Let $n > 1$, and let $\mathbf{R}^{n-1}$ be a hyperplane in $\mathbf{R}^n$. Show that if $1 \leq p \leq \infty$ and g is an L^p function on $\mathbf{R}^n$ with Fourier support in the ball $|\xi| \lesssim 2^k$ for some integer k, then
$$\|g\|_{L^p(\mathbf{R}^{n-1})} \lesssim 2^{k/p}\|g\|_{L^p(\mathbf{R}^n)}.$$
From this, conclude the *Sobolev trace lemma*: whenever $1 < p < \infty$, $s \in \mathbf{R}$, and $\sigma > 1/p$, we have
$$\|f\|_{W^{p,s-\sigma}(\mathbf{R}^{n-1})} \lesssim \|f\|_{W^{p,s}(\mathbf{R}^n)}.$$

[14]This argument can be viewed using phase plane heuristics. A Calderón-Zygmund operator such as T, when applied to a function of frequency $|\xi| \sim 2^k$, only spreads out the function by about 2^{-k} or so (why?). Since the function b_I is supported on I and has mean zero, it mostly consists of frequencies $|\xi| \gtrsim |I|^{-1}$ (why?). Thus T only spreads b_I by about $O(|I|)$ or so.

- Prove the *Gagliardo-Nirenberg inequality*: If $1 < p < q < \infty$ is such that $1/q = 1/p + \theta/n$ for some $0 < \theta < 1$, then we have
$$\|f\|_q \lesssim \|f\|_p^{1-\theta} \|\nabla f\|_p^\theta$$
for all Schwartz functions f on $\mathbf{R}^n$. (Note: it is tempting to apply Lemma 1.3, but this only works when $p < n$ (why?).)
- Prove the *Poincare inequality*: if $1 \leq p \leq \infty$, and f is a Schwartz function supported on the unit ball $\{|x| \leq 1\}$, then[15]
$$\|f\|_p \lesssim \|\nabla f\|_p.$$
(Hint: From Lemma 1.1 one can control $\|P_{>-C} f\|_p$ well for any large constant C. To get from this partial result to control of all of $\|f\|_p$, try to prove an inequality of the form
$$\|f\|_p \leq C\|P_{>-C} f\|_p + \frac{1}{2}\|f\|_p.$$
One way to do this is to use the identity $f = \tilde{\chi} P_{>-C} f + \sum_{k \leq -C} \tilde{\chi} P_k f$, where $\tilde{\chi}$ is a bump function adapted to $\{|x| \leq 2\}$ which equals 1 on $\{|x| \leq 1\}$.)
  Given an informal explanation using the uncertainty principle as to why $P_{>-C} f$ should be the "dominant" component of f.
- Let $0 < \alpha < 1$. The *Hölder class* $\Lambda_\alpha(\mathbf{R}^n)$ is the space of all functions f on $\mathbf{R}^n$ for which the norm
$$\|f\|_{\Lambda_\alpha} := \|f\|_\infty + \sup_{x,y \in \mathbf{R}^n : 0 < |x-y| < 1} \frac{|f(x) - f(y)|}{|x-y|^\alpha}$$
is finite. (Functions in Λ_α are called *Hölder continuous of order* α). These spaces are especially useful in elliptic PDE.
  (a) Prove that
$$\|f\|_{\Lambda_\alpha} \sim \|P_{\leq 0} f\|_\infty + \sup_{k>0} 2^{k\alpha} \|P_k f\|_\infty.$$
with the implicit constant depending on n, α. (Thus we have the heuristic $\Lambda_\alpha \approx W^{\alpha,\infty}$.
  (b) Prove the following variant of the Sobolev embedding theorem: if $n < p < \infty$ and $0 < \alpha < 1$ are such that $n/p < 1 - \alpha$, then
$$\|f\|_{\Lambda_\alpha} \lesssim \|\nabla f\|_p.$$
  (c) The *Zygmund class* Λ_* is the space of all functions f whose norm
$$\|f\|_{\Lambda_*} := \|f\|_\infty + \sup_{x,h \in \mathbf{R}^n : 0 < |h| < 1} \frac{|f(x+h) - 2f(x) + f(x-h)|}{|h|}$$
is finite. Show that
$$\|f\|_{\Lambda_*} \sim \|P_{\leq 0} f\|_\infty + \sup_k 2^k \|P_k f\|_\infty.$$
(Thus $\Lambda_* \approx W^{1,\infty}$).

[15]This particular inequality can be proven by more elementary means, such as the fundamental theorem of calculus. However, the *method* of proof sketched here is far more flexible, and extends to much more general situations.

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

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