LECTURE NOTES 3 FOR 254A

TERENCE TAO

1. Product estimates

In last week's notes we introduced the Sobolev spaces $W^{s,p}(\mathbf{R}^n)$ in terms of Littlewood-Paley operators. This week we study the question of product estimates for these spaces, asking such questions as: if f lives in one Sobolev space W^{s_1,p_1} and g lives in another Sobolev space W^{s_2,p_2}, what Sobolev spaces $W^{s,p}$ does the product fg live in? More quantitatively, for which s, s_1, s_2, p, p_1, p_2 do we have the bilinear estimate

$$\|fg\|_{W^{s,p}} \lesssim \|f\|_{W^{s_1,p_1}} \|g\|_{W^{s_2,p_2}}. \tag{1}$$

As one might expect, the general solution to this problem is a little messy (although it can be solved completely). We shall address only a special case of this problem, namely the question of which Sobolev spaces $W^{s,p}$ are closed under multiplication. In other words, we ask for which s, p do we have the estimate

$$\|fg\|_{W^{s,p}} \lesssim \|f\|_{W^{s,p}} \|g\|_{W^{s,p}}.$$

If $W^{s,p}$ enjoyed such an estimate, then it would be a *Banach algebra*, which gives it several nice properties. For instance, if $F(z)$ is an entire function, then the non-linear map $f \mapsto F(f)$ would then be bounded on $W^{s,p}$ (just use the Taylor expansion of F).

Some spaces are clearly closed under multiplication: L^∞, for instance. Another example is C^1, the class of functions which have a continuous and bounded derivative, with norm

$$\|f\|_{C^1} = \|f\|_\infty + \|\nabla f\|_\infty.$$

One can easily see this space is closed under multiplication once one uses the Leibnitz rule $\nabla(fg) = (\nabla f)g + f(\nabla g)$.

On the other hand, a space such as L^2 is not closed under multiplication (the product of two L^2 functions is only in L^1). There is a general rule of thumb, which says that a Banach space X should only be closed under multiplication if its elements are bounded (i.e. if the X norm controls the L^∞ norm). Heuristically, the reason is that if g is not in L^∞, then the map $f \mapsto fg$ should not be bounded on any Banach space X (just test this operator on a function f supported where g is large).

So the question then arises as to when the $W^{s,p}$ norm controls the L^∞ norm. The Sobolev embedding theorem (see previous notes) says that this occurs when

$s > n/p$. Conversely, when $s \leq n/p$, $W^{s,p}$ cannot control L^∞. To see this, let ψ be a non-negative bump function adapted to $1 \leq |\xi| \leq 2$ and consider the function

$$\frac{1}{N^{1/p}} \sum_{k=1}^{N} 2^{-sk} 2^{nk/p} \hat{\psi}(2^k x).$$

This function has an $W^{s,p}$ norm of $O(1)$ (why?) but it's value at 0 goes to ∞ as $N \to \infty$ (why?).

Indeed, $W^{s,p}$ is not closed under multiplication when $s \leq n/p$ (see Exercise 1). On the other hand, we have

Proposition 1.1. *If $1 < p < \infty$ and $s > n/p$, then*

$$\|fg\|_{W^{s,p}(\mathbf{R}^n)} \lesssim \|f\|_{W^{s,p}(\mathbf{R}^n)} \|g\|_{W^{s,p}(\mathbf{R}^n)}$$

for all functions f, g for which the right-hand side makes sense.

This result by itself is not particularly interesting, however the method of proof is definitely worth learning about, as it can be used to handle a very wide class of bilinear and multilinear estimates.

Proof We shall prove this using the *Littlewood-Paley trichotomy*, which is a basic way one can analyze bilinear expressions such as the pointwise product $(f, g) \mapsto fg$.

We can split the left-hand side using Littlewood-Paley projections as

$$\|fg\|_{W^{s,p}} \sim \|P_{\leq 0}(fg)\|_p + \|(\sum_{k>0} |2^{sk} P_k(fg)|^2)^{1/2}\|_p.$$

The low frequency component is easily estimated by Hölder and Sobolev:

$$\|P_{\leq 0}(fg)\|_p \lesssim \|fg\|_p \lesssim \|f\|_p \|g\|_\infty \lesssim \|f\|_{W^{s,p}} \|g\|_{W^{s,p}}.$$

It's the high frequency component which is interesting. More precisely, one is interested in understanding the quantity $P_k(fg)$ - the k^{th} frequency component of k.

One can split f, g into Littlewood-Paley pieces too, and write

$$P_k(fg) = \sum_{k',k'' \in \mathbf{Z}} P_k((P_{k'} f)(P_{k''} g)).$$

The summand should be thought of as the contribution to the k^{th} frequency piece of fg coming from the k'^{th} frequency piece of f and the k''^{th} frequency piece of g.

This seems like an awful double sum, however most of the terms are zero. Recall that $P_{k'} f$ has Fourier support in $2^{k'-1} \leq |\xi| \leq 2^{k'+1}$ and $P_{k''} g$ has Fourier support in $2^{k''-1} \leq |\xi| \leq 2^{k''+1}$, so that $(P_{k'} f)(P_{k''} g)$ has Fourier support in the sum of these two annuli. This sum needs to intersect the annulus $2^{k-1} \leq |\xi| \leq 2^{k+1}$ in order for the summand to be non-zero.

This has the following consequences. If $k' < k-5$, then the summand vanishes unless $k-3 \leq k'' \leq k+3$. Similarly, if $k'' < k-5$, then the summand vanishes unless $k-3 \leq k' \leq k+3$. If $k' > k-5$, then the summand vanishes unless $k'-3 \leq k'' \leq k'+3$. Putting this all together, we get

$$\begin{aligned} P_k(fg) = & P_k[(P_{<k-5}f)(P_{k-3\leq \cdot \leq k+3}g) \\ & + (P_{k-3\leq \cdot \leq k+3}f)(P_{<k-5}g) \\ & + (P_{k-5\leq \cdot \leq k+5}f)(P_{k-5\leq \cdot \leq k+5}g) \\ & + \sum_{k',k''>k+5:|k'-k''|\leq 3} (P_{k'}f)(P_{k''}g)]. \end{aligned} \tag{2}$$

In other words, there are four distinct interactions of f and g which give a non-zero contribution to $P_k(fg)$:

- (Low-high interactions): f has frequency $\ll 2^k$, while g has frequency $\sim 2^k$.
- (High-low interactions): f has frequency $\sim 2^k$, while g has frequency $\ll 2^k$.
- (Low-low interactions): Both f and g have frequency $\sim 2^k$.
- (High-high interactions): Both f and g have frequency $\sim 2^{k'}$ for some $2^{k'} \gg 2^k$.

This splitting is widely used, but does not seem to have a standardized name; I call it the *Littlewood-Paley trichotomy*[1].

Let's consider the low-high interaction:

$$\|(\sum_{k>0} |2^{sk} P_k((P_{<k-5}f)(P_{k-3\leq \cdot \leq k+3}g))|^2)^{1/2}\|_p.$$

We would like to throw away the P_k projection. This is accomplished by the following variant of the Littlewood-Paley inequality[2]:

Lemma 1.2. *For all $1 < p < \infty$, we have*

$$\|(\sum_{k>0} |P_k f_k|^2)^{1/2}\|_p \lesssim \|(\sum_{k>0} |f_k|^2)^{1/2}\|_p.$$

This lemma is consistent with a general philosophy that one can always throw away the projections P_k and $P_{<k}$ whenever one wishes (cf. the left-hand side of the cheap Littlewood-Paley inequality).

Proof One can rewrite this estimate as

$$\|Tf\|_p \lesssim \|f\|_p$$

where $f = (f_k)_{k\in \mathbf{Z}}$ is a vector-valued function and T is the operator

$$(Tf)_k = P_k f_k.$$

[1] Okay, so this is a tetrachotomy as stated. Often the low-low interaction is subsumed into one of the other three interactions.

[2] There is a much stronger (and quite useful!) version of this estimate known as the *Fefferman-Stein maximal inequality*, but we will not need it in this course.

This operator is a Calderón-Zygmund operator whose kernel is an infinite matrix with entries

$$K_{jk}(x,y) = \delta_{jk} 2^{nk} \hat{\psi}(2^k(x-y)).$$

From Plancherel we can easily check that T is bounded on L^2, and the kernel obeys the right bounds, and so the L^p estimate follows. ■

Using this lemma we can bound the previous by

$$\|(\sum_{k>0} |2^{sk}(P_{<k-5}f)(P_{k-3\leq \cdot \leq k+3}g))|^2)^{1/2}\|_p.$$

We would now like to throw away the $P_{<k-5}$ projection. We can't quite do that, but what we can say is that

$$|P_{<k-5}f(x)| \lesssim Mf(x).$$

where M is the *Hardy-Littlewood maximal operator*

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

This is an example of a general principle, which is that averages of f are almost always dominated by the maximal function (which is one reason why the maximal function is so useful). To see this, we write

$$\begin{aligned}
|P_{<k-5}f(x)| &= |\int f(y) 2^{n(k-5)} \hat{\phi}(2^{k-5}(x-y)) \, dy| \\
&\lesssim \int |f(y)| 2^{nk} (1 + 2^k|x-y|)^{-100n} \, dy \\
&\lesssim 2^{nk} \int_{B(x,2^{-k})} |f(y)| \, dy + \sum_{j>0} 2^{nk} 2^{-100nj} \int_{B(x,2^{-k+j})} |f(y)| \, dy \\
&\lesssim Mf(x) + \sum_{j>0} 2^{nj} 2^{-100nj} Mf(x) \\
&\lesssim Mf(x).
\end{aligned}$$

Returning to the low-high interactions, we can now bound this contribution by

$$\|(\sum_{k>0} |2^{sk}(Mf)(P_{k-3\leq \cdot \leq k+3}g))|^2)^{1/2}\|_p.$$

We can take the Mf out:

$$\||Mf|(\sum_{k>0} |2^{sk}(P_{k-3\leq \cdot \leq k+3}g))|^2)^{1/2}\|_p.$$

By Sobolev embedding,

$$\|f\|_\infty \lesssim \|f\|_{W^{s,p}}.$$

Clearly[3] $\|Mf\|_\infty \lesssim \|f\|_\infty$, so we can bound the previous by

$$\|f\|_{W^{s,p}} \sim \|(\sum_{k>0} |2^{sk}(P_{k-3\leq \cdot \leq k+3}f)|^2)^{1/2}\|_p.$$

[3]More generally, we have $\|Mf\|_p \lesssim \|f\|_p$ for all $1 < p \leq \infty$. See Appendix II.

The operator $P_{k-3\leq\cdot\leq k+3}$ is a mild variant of P_k, indeed it is clear that

$$(\sum_{k>0} |2^{sk}(P_{k-3\leq\cdot\leq k+3}g))|^2)^{1/2} \lesssim (\sum_{k>0} |2^{sk}(P_k g))|^2)^{1/2}.$$

By definition of $W^{s,p}$ we have thus bounded the low-high interaction by $\|f\|_{W^{s,p}}\|g\|_{W^{s,p}}$ as desired. The high-low and low-low interactions are similar, so let us now look at the high-high interaction

$$\|(\sum_{k>0} |2^{sk}P_k(\sum_{k',k''>k+5:|k'-k''|\leq 3} (P_{k'}f)(P_{k''}g)]))|^2)^{1/2}\|_p.$$

We write $k' = k+a$, $k'' = k+b$:

$$\|(\sum_{k>0} |2^{sk}P_k(\sum_{a,b>5:|a-b|\leq 3} (P_{k+a}f)(P_{k+b}g)]))|^2)^{1/2}\|_p.$$

What's going to happen is that the larger a and b get, the smaller this contribution is going to get (it is difficult for two high frequencies to get together and cancel to form a low frequency, especially when the function is as smooth as $W^{s,p}$ (when a function is smooth the high frequency component is small). In general, the only time the high-high interaction becomes a big problem is when the functions are very rough (e.g. negative order Sobolev spaces)). To exploit this we use the triangle inequality to isolate the contributions of the individual a and b:

$$\sum_{a,b>5:|a-b|\leq 3} \|(\sum_{k>0} |2^{sk}P_k((P_{k+a}f)(P_{k+b}g)]))|^2)^{1/2}\|_p.$$

Again, we throw away the P_k, and estimate $P_{k+a}f$ by Mf:

$$\sum_{a,b>5:|a-b|\leq 3} \|Mf(\sum_{k>0} |2^{sk}(P_{k+b}g)]))|^2)^{1/2}\|_p.$$

Observe that

$$(\sum_{k>0} |2^{sk}(P_{k+b}g)|^2)^{1/2} \lesssim 2^{-sb}(\sum_{k>0} |2^{sk}(P_k g)|^2)^{1/2}$$

so by the previous arguments we have

$$\sum_{a,b>5:|a-b|\leq 3} 2^{-sb}\|f\|_{W^{s,p}}\|g\|_{W^{s,p}}$$

which is OK. ■

A less rigorous proof of the above result proceeds by using

Heuristic 1.3 (Fractional Leibnitz rule). *If D is a differentiation operator (possibly of fractional order) then*

$$D(fg) \lessapprox (Df)g + f(Dg).$$

Pretending this rule works for the moment, we apply it with the operator D defined by

$$\widehat{Df}(\xi) := \langle\xi\rangle^s \hat{f}(\xi).$$

Then we have

$$\|fg\|_{W^{s,p}} \sim \|D(fg)\|_p.$$

Applying the Leibnitz rule we obtain

$$\|D(fg)\|_p \lesssim \|Df\|_p \|g\|_\infty + \|f\|_\infty \|Dg\|_p.$$

Using Sobolev embedding $\|f\|_\infty \lesssim \|f\|_{W^{s,p}}$ we thus obtain the desired bound

$$\|fg\|_{W^{s,p}} \lesssim \|f\|_{W^{s,p}} \|g\|_{W^{s,p}}.$$

The fractional Leibnitz rule is not quite rigorously correct, in the same way that the heuristic $\nabla P_k f \sim 2^k P_k f$ is not quite correct, however it is close enough that any Sobolev space result one "proves" with it is quite likely to be true. (For instance, one can use it to prove most of the estimates of the form (1) which are actually true.)

To see why this rule works, let us try it for the operator D defined above. Heuristically we expect

$$DP_k f \sim \langle 2^k \rangle^s P_k f.$$

Inserting this into the formula (2) we obtain heuristically that

$$\begin{aligned} P_k D(fg) \sim & P_k[\langle 2^k\rangle^s (P_{<k-5} f)(P_{k-3 \leq \cdot \leq k+3} g) \\ & + \langle 2^k\rangle^s (P_{k-3 \leq \cdot \leq k+3} f)(P_{<k-5} g) \\ & + \langle 2^k\rangle^s (P_{k-5 \leq \cdot \leq k+5} f)(P_{k-5 \leq \cdot \leq k+5} g) \\ & + \langle 2^k\rangle^s \sum_{k',k''>k+5: |k'-k''| \leq 3} (P_{k'} f)(P_{k''} g)]. \end{aligned}$$

If we throw away some P_k stuff, the first factor looks like $f(P_k Dg)$, while the second factor looks like $(P_k Df)g$. The third factor looks like either one. The fourth factor looks like (for instance)

$$\sum_{k'>k+5} 2^{-s(k-k')} (P_{k'} Df) g.$$

If one sums all these heuristic estimates in k one obtains the Leibnitz rule.

Another way to view the rule is using the Fourier transform. The Fourier transform of $D(fg)$ at ξ is given by

$$\widehat{D(fg)}(\xi) = \int_{\xi = \xi_1 + \xi_2} \langle \xi \rangle^s \hat f(\xi_1) \hat g(\xi_2)\, d\xi_1$$

while the Fourier transform of $(Df)g + f(Dg)$ is given by

$$\widehat{(Df)g + f(Dg)}(\xi) = \int_{\xi = \xi_1 + \xi_2} (\langle \xi_1 \rangle^s + \langle \xi_2 \rangle^s) \hat f(\xi_1) \hat g(\xi_2)\, d\xi_1.$$

The key point is the "triangle inequality"

$$\langle \xi \rangle^s \lesssim \langle \xi_1 \rangle^s + \langle \xi_2 \rangle^s$$

(check the "low-high" case $\langle \xi_1 \rangle \ll \langle \xi_2 \rangle$, "high-low" case $\langle \xi_1 \rangle \gg \langle \xi_2 \rangle$, and "high-high" case $\langle \xi_1 \rangle \sim \langle \xi_2 \rangle$ separately). In particular, if f and g have a non-negative Fourier transform, then the Fourier transform of $D(fg)$ is pointwise dominated by the Fourier transform of $(Df)g + f(Dg)$. This quantitative version of the Leibnitz rule is useful for L^2 type estimates, since one can then use Plancherel to obtain bounds like

$$\|D(fg)\|_2 \lesssim \|(Df)g\|_2 + \|f(Dg)\|_2$$

at least when f, g have non-negative Fourier transform. (This condition is not particularly onerous in L^2, since any L^2 function f can be decomposed as a finite linear combination of functions with non-negative Fourier transform and whose L^2 norm is bounded by that of f, by Plancherel).

2. Multilinear operators

The pointwise product operator $(f, g) \mapsto fg$ is a bilinear operator which can be written using the Fourier transform as

$$\widehat{fg}(\xi) = \int_{\xi=\xi_1+\xi_2} \hat{f}(\xi_1)\hat{g}(\xi_2)\, d\xi_1.$$

Some other examples of bilinear operators were given above. Note that they all have the form

$$\widehat{T(f,g)}(\xi) = \int_{\xi=\xi_1+\xi_2} m(\xi_1,\xi_2)\hat{f}(\xi_1)\hat{g}(\xi_2)\, d\xi_1.$$

More generally, we could consider multi-linear operators of the form

$$T(\widehat{f_1, \ldots}, f_n)(\xi) = \int_{\xi=\xi_1+\ldots+\xi_n} m(\xi_1, \ldots, \xi_n)\hat{f}_1(\xi_1)\ldots\hat{g}(\xi_2)\, d\xi_1 \ldots d\xi_{n-1}. \tag{3}$$

Such an operator $T = T_m$ is called a *multi-linear Fourier multiplier with symbol* m, and generalizes the notion of a linear Fourier multiplier (which is the $n = 1$ case). For example, on the line $\mathbf{R}$, the operator

$$(f, g, h) \mapsto f''g'h + fg'h'' + (fgh)''$$

is such an operator with symbol

$$(2\pi i\xi_1)^2(2\pi i\xi_2) + (2\pi i\xi_2)(2\pi i\xi_3)^2 + 2\pi i(\xi_1 + \xi_2 + \xi_3)^2.$$

Thus any combination of derivatives and products can be represented as one of these operators - which means that these operators come up all the time in PDE, especially in non-linear PDE. On the other hand, there are more exotic operators which do not have an obvious representation in terms of derivatives, e.g. $m(\xi_1, \xi_2, \xi_3) = \mathrm{sgn}(\xi_1 - \xi_2)$.

These operators are all translation invariant in the following sense. Let $x_0 \in \mathbf{R}^n$, and let f_{x_0} denote the translation $f_{x_0}(x) := f(x - x_0)$ of f by x_0. Then

$$T(f_1, \ldots, f_n)_{x_0} = T((f_1)_{x_0}, \ldots, (f_n)_{x_0}).$$

This is because translation by x_0 in physical space corresponds to modulation by $\exp(2\pi i\xi \cdot x_0)$, and because $\exp(2\pi i\xi \cdot \xi_0)$ is the product of the $\exp(2\pi i\xi_j \cdot \xi_0)$ for $j = 1, \ldots, n$.

Conversely, any multi-linear translation operator T must be of the form (3) (at least, if one is willing to accept symbols m which are distributions or worse). To see this, suppose that T is translation invariant, and suppose we test[4] T on exponentials

[4]We shall ignore any issues as to whether T can actually be defined on such functions, and whether the integrals that we shall write down actually converge.

$f_j(x) := \exp(2\pi i x \cdot \xi_j)$ for $j = 1, \ldots, n$ and $\xi_j \in \mathbf{R}^n$. Then we have the modulated translation invariance $(f_j)_{x_0} = \exp(2\pi i x_0 \cdot \xi_j) f_j$, which implies that

$$T(f_1, \ldots, f_n)_{x_0} = \exp(2\pi i x_0 \cdot (\xi_1 + \ldots + \xi_n)) T(f_1, \ldots, f_n)$$

for all x_0, which implies that

$$T(f_1, \ldots, f_n)(x) = \exp(2\pi i x \cdot (\xi_1 + \ldots + \xi_n)) m(\xi_1, \ldots, \xi_n)$$

for some $\xi_1, \ldots, \xi_n)$. If one then writes a general function f_j as an average of exponentials using the Fourier inversion formula and uses linearity, one obtains (3).

Many multi-linear questions about Sobolev spaces and similar function spaces eventually reduce to the question of whether a certain operator T maps some Lebesgue spaces to another:

$$T : L^{p_1} \times \ldots \times L^{p_n} \to L^p$$

or in other words that

$$\|T(f_1, \ldots, f_n)\|_p \lesssim \|f_1\|_{p_1} \ldots \|f_n\|_{p_n}. \tag{4}$$

For instance, the estimate (1) can be written as

$$\|D^s(fg)\|_p \lesssim \|D^{s_1} f\|_{p_1} \|D^{s_2} g\|_{p_2}$$

where D^s is the Fourier multiplier with symbol $\langle \xi \rangle^s$. This estimate is equivalent to

$$\|D^s((D^{-s_1} f D^{-s_2} g)\|_p \lesssim \|D^{s_1} f\|_{p_1} \|D^{s_2} g\|_{p_2}$$

or in other words that

$$T : L^{p_1} \times L^{p_2} \to L^p$$

where T is the bilinear Fourier multiplier with symbol

$$m(\xi_1, \xi_2) := \frac{\langle \xi_1 + \xi_2 \rangle^s}{\langle \xi_1 \rangle^{-s_1} \langle \xi_2 \rangle^{-s_2}}.$$

It would therefore be nice to be able to characterize the conditions on the symbol m so that estimates of the form (4) are satisfied. Unfortunately this task seems well beyond our current technology; even in the linear case $n = 1$, in which there has been fifty years of intensive work on the multiplier problem, we still only have a partial understanding of which multipliers are bounded on L^p spaces, especially when these multipliers have singularities on large sets. In the multilinear case we are finally beginning to get a reasonably good hold of the L^2 case, but our results for the general L^p problem are far from complete.

An important special case occurs when m is a *symbol of order 0* in the sense that

$$|\nabla^j m(\xi)| \lesssim |\xi|^j$$

for all $j \geq 0$, with the implicit constant depending on j. A typical example of such a multiplier is $m \equiv 1$, in which case T is just the pointwise product operator $T(f_1, \ldots, f_n) = f_1 \ldots f_n$. The L^p estimates for this operator are given by Hölder's inequality:

$$\|f_1 \ldots f_n\|_p \lesssim \|f_1\|_{p_1} \ldots \|f_n\|_{p_n} \text{ whenever } \frac{1}{p} = \frac{1}{p_1} + \ldots + \frac{1}{p_n}.$$

The same estimate holds for more general multipliers of order 0, provided that we stay away from the dangerous spaces L^1 and L^∞:

Theorem 2.1 (Coifman-Meyer multiplier theorem). *If $1 < p_j < \infty$ for $j = 1, \ldots, n$, and T is a multi-linear operator whose symbol m is a symbol of order 0, then* (4) *holds whenever*

$$\frac{1}{p} = \frac{1}{p_1} + \ldots + \frac{1}{p_n}.$$

This is the generalization of the Hörmander-Mihlin multiplier theorem to higher linearities.

We shall prove a special case of Theorem 2.1 in the next section. For now, we give one sample application. Recall from (2) that we can split the product fg into four pieces

$$\begin{aligned} fg = &\sum_k P_k[(P_{<k-5}f)(P_{k-3\leq\cdot\leq k+3}g)] \\ &+ \sum_k P_k[(P_{k-3\leq\cdot\leq k+3}f)(P_{<k-5}g)] \\ &+ \sum_k P_k[(P_{k-5\leq\cdot\leq k+5}f)(P_{k-5\leq\cdot\leq k+5}g) \\ &+ \sum_k P_k[\sum_{k',k''>k+5:|k'-k''|\leq 3}(P_{k'}f)(P_{k''}g)]. \end{aligned} \tag{5}$$

These four pieces are examples of *paraproducts* - in other words, components of the full product fg. The first paraproduct only allows the low frequencies of f to multiply the high frequencies of g, but not vice versa, while the other three paraproducts have similar types of behavior (in which all but one kind of interaction is suppressed). Paraproducts are generally denoted by the symbol π, thus the first paraproduct above might be called $\pi_1(f,g)$. Paraproducts are useful for separating a product into pieces which behave differently with respect to derivatives. For instance, the Leibnitz rule for the paraproduct $\pi_1(f,g)$ is simpler than that for the full product, indeed we essentially have $D\pi_1(f,g) \sim \pi_1(f,Dg)$ (why?). Similar rules exist for π_2, etc. Thus these paraproducts can be used to analyze such rules as the Leibnitz rule more accurately.

These paraproducts are multi-linear operators of the form discussed above (see Exercise 2). Thus they have the same type of L^p estimates as the usual pointwise product does, except at the endpoints L^1 and L^∞ (which are always problematic).

We will give another example, the div-curl lemma, in the next week's notes.

3. Proof of Theorem 2.1

We will prove a special case of Theorem 2.1, in which one only deals with a bilinear operator and in which all exponents stay above L^1. The general case is more technical and will not be pursued here. In other words, we shall only prove

Theorem 3.1. *If $1 < p, q, r < \infty$ is such that $1/p = 1/q + 1/r$, and T is a bi-linear operator whose symbol m is a symbol of order 0, then*

$$\|T(f,g)\|_p \lesssim \|f\|_q \|g\|_r.$$

The first task is to split the multiplier $m(\xi_1, \xi_2)$ of T into low-high, high-low, and high-high components.

We can partition unity as

$$1 = \chi_1(\xi_1, \xi_2) + \chi_2(\xi_1, \xi_2) + \chi_3(\xi_1, \xi_2)$$

where χ_1 is supported on the region $|\xi_1| \leq |\xi_2|/4$ and equals 1 on the region $|\xi_1| \leq |\xi_2|/8$, χ_2 is supported on the region $|\xi_1| \geq 4|\xi_2|$ and equals 1 on the region $|\xi_1| \geq 8|\xi_2|$, and χ_3 is supported on the region $|\xi_2|/8 \leq |\xi_1| \leq 8|\xi_2|$, and such that we have the bounds

$$|\nabla^j \chi_i(\xi_1, \xi_2)| \lesssim |\xi|^{-j}$$

for all $j \geq 0$ and $i = 1, 2, 3$. We can then partition

$$m = m\chi_1 + m\chi_2 + m\chi_3,$$

and observe that each $m\chi_i$ is also a symbol of order 0 for $i = 1, 2, 3$. Thus it suffices to prove the theorem for each of the $m\chi_i$ separately, i.e. we have to show that

$$\| \int e^{2\pi i x(\xi_1+\xi_2)} m\chi_i(\xi_1, \xi_2) \hat{f}(\xi_1) \hat{g}(\xi_2) \; d\xi_1 d\xi_2 \|_{L^p_x} \lesssim \|f\|_q \|g\|_r$$

for $i = 1, 2, 3$.

We begin with $m\chi_3$, which is the easiest of the three. This symbol is supported on the region $|\xi_1| \sim |\xi_2|$. We may therefore do a Littlewood-Paley like decomposition, writing

$$m\chi_3(\xi_1, \xi_2) = \sum_{k \in \mathbf{Z}} m_k(\xi_1, \xi_2)$$

where m_k is a bump function adapted to the region $|\xi_1| \sim |\xi_2| \sim 2^k$. We therefore need to control

$$\| \sum_k \int e^{2\pi i x(\xi_1+\xi_2)} m_k(\xi_1, \xi_2) \hat{f}(\xi_1) \hat{g}(\xi_2) \; d\xi_1 d\xi_2 \|_{L^p_x}. \tag{6}$$

It would be convenient at this point if we could factor $m_k(\xi_1, \xi_2)$ as a tensor product

$$m_k(\xi_1, \xi_2) = m'_k(\xi_1) m''_k(\xi_2)$$

because we could then factor the previous expression as

$$\| \sum_k (T_{m'_k} f)(T_{m''_k} g) \|_{L^p_x},$$

where $T_{m'_k}$ is the Fourier multiplier with symbol m'_k, and similarly for m''_k.

Of course, not all functions m_k can be factored as tensor products. However, we can do the next best thing, which is to write m_k as a rapidly decreasing linear combination of tensor products:

Lemma 3.2. *We can write*

$$m_k(\xi_1, \xi_2) = \sum_{j,j' \in \mathbf{Z}^n} c_{j,j'} m'_{k,j,j'}(\xi_1) m''_{k,j,j'}(\xi_2)$$

where for each $j, j' \in \mathbf{Z}^n$, the functions $m'_{k,j,j'}$ and $m''_{k,j,j'}$ are bump functions adapted to $|\xi| \sim 2^k$ uniformly in j, j', and the co-efficients $c_{j,j'}$ obey the bounds $|c_{j,j'}| \lesssim (1 + |j| + |j'|)^{-M}$ for any $M > 0$.

This Lemma turns out to be quite useful in many other contexts; it asserts that all bump functions are essentially tensor products of lower-dimensional bump functions.

Proof We may rescale $k = 0$.

On the support of m_0, the variables ξ_1, ξ_2 range in some annulus $A = \{|\xi| \sim 2^1\}$. Let $\chi(\xi)$ be a bump function which equals 1 on A and is adapted to a thickening of A. We then have

$$m_0(\xi_1, \xi_2) = \chi(\xi_1)\chi(\xi_2) m_0(\xi_1, \xi_2).$$

The function m_0 can be thought of as a smooth function on the $2n$-dimensional cube $[-C, C]^n \times [-C, C]^n$, or as a smooth function on the corresponding torus $(\mathbf{R}/2C\mathbf{Z})^n \times (\mathbf{R}/2C\mathbf{Z})^n$, where C is a large constant. By the Fourier inversion formula for the torus we thus have

$$m_0(\xi_1, \xi_2) = \sum_{j,j' \in \mathbf{Z}^n} e^{2\pi i j \cdot \xi_1 / C} e^{2\pi i j' \cdot \xi_2 / C} a_{j,j'}$$

where $a_{j,j'}$ is a rapidly decreasing sequence in j, j'. We write $a_{j,j'} = c^3_{j,j'}$. The claim then follows with

$$m'_{0,j,j'}(\xi_1) := c_{j,j'} e^{2\pi i j \cdot \xi_1 / C} \chi(\xi_1)$$

and

$$m''_{0,j,j'}(\xi_2) := c_{j,j'} e^{2\pi i j' \cdot \xi_2 / C} \chi(\xi_2).$$

Note that the modulations $e^{2\pi i j \cdot \xi_1 / C}$, $e^{2\pi i j' \cdot \xi_2 / C}$ may make the derivatives $\nabla^s m'$, $\nabla^s m''$ larger by a polynomial factor in j and j', but this is more than compensated by the rapidly decreasing factor $c_{j,j'}$. ∎

Because of this lemma we may write (6) as

$$\| \sum_{j,j'} c_{j,j'} \sum_k (T_{m'_{k,j,j'}} f)(T_{m''_{k,j,j'}} g) \|_{L^p_x}.$$

To bound this by $\|f\|_q \|g\|_r$, it suffices by the rapid decay of $c_{j,j'}$ to show that

$$\| \sum_k (T_{m'_{k,j,j'}} f)(T_{m''_{k,j,j'}} g) \|_{L^p_x} \lesssim \|f\|_q \|g\|_r$$

uniformly in j, j'.

Fix j, j'. The symbols $m'_{k,j,j'}$ and $m''_{k,j,j'}$ are bump functions adapted to the annulus $|\xi| \sim 2^k$, so we shall suggestively rename $T_{m'_{k,j,j'}}$ and $T_{m''_{k,j,j'}}$ as P'_k and P''_k

respectively. We then estimate the previous using Cauchy-Schwarz as

$$\|(\sum_k |P'_k f|^2)^{1/2}(\sum_k |P''_k g|^2)^{1/2}\|_p$$

which then splits using Hölder as

$$\|(\sum_k |P'_k f|^2)^{1/2}\|_q\|(\sum_k |P''_k g|^2)^{1/2}\|_r$$

which by the Littlewood-Paley inequality is bounded by $\|f\|_q\|g\|_r$. This shows that the $m\chi_3$ component of the paraproduct is OK.

To handle the $m\chi_1$ component we use a duality trick. We need to prove that

$$\|\int e^{2\pi i x(\xi_1+\xi_2)} m\chi_1(\xi_1,\xi_2)\hat{f}(\xi_1)\hat{g}(\xi_2)\ d\xi_1 d\xi_2\|_{L^p_x} \lesssim \|f\|_q\|g\|_r.$$

By duality this is the same as showing

$$|\langle\int e^{2\pi i x(\xi_1+\xi_2)} m\chi_1(\xi_1,\xi_2)\hat{f}(\xi_1)\hat{g}(\xi_2)\ d\xi_1 d\xi_2, h\rangle| \lesssim \|f\|_q\|g\|_r\|h\|_{p'}$$

for all f, g, h. By Plancherel the left-hand side is

$$|\int \delta(\xi_1+\xi_2+\xi_3) m\chi_1(\xi_1,\xi_2)\hat{f}(\xi_1)\hat{g}(\xi_2)\hat{h}(\xi_3)|$$

which is equal to

$$|\langle\int e^{2\pi i x(\xi_1+\xi_3)} m\chi_1(\xi_1,-\xi_1-\xi_3)\hat{f}(\xi_1)\hat{h}(\xi_3)\ d\xi_1 d\xi_3, g\rangle|.$$

Thus it suffices to show that

$$\|\int e^{2\pi i x(\xi_1+\xi_3)} m\chi_1(\xi_1,-\xi_1-\xi_3)\hat{f}(\xi_1)\hat{h}(\xi_3)\ d\xi_1 d\xi_3\|_{r'} \lesssim \|f\|_q\|h\|_{p'}.$$

Now the observation is that $m\chi_1(\xi_1,-\xi_1-\xi_3)$ behaves almost exactly like $m\chi_3(\xi_1,\xi_3)$ (it obeys the same derivative estimates, and is supported on the region $|\xi_1| \sim |\xi_3|$). So this estimate is OK by the same arguments used to handle $m\chi_3$. (Note that $1/p = 1/q + 1/r$ iff $1/r' = 1/q + 1/p'$).

A similar trick deals with $m\chi_2$, and Theorem 3.1 is proved.

4. Appendix I: The Marcinkiewicz interpolation theorem

In this appendix we prove the Marcinkiewicz interpolation theorem, which states that weak-type estimates can be interpolated to become strong-type estimates. In fact we shall improve this statement slightly, replacing weak-type estimates by the even weaker notion of a *restricted weak-type estimate.*

Let X and Y be measure spaces. Recall that if T is a linear operator from functions of X to functions of Y, we say that T is of strong-type (p,q) if

$$\|Tf\|_q \lesssim \|f\|_p \text{ for all } f.$$

We say that T is of weak-type (p,q) if

$$|\{|Tf| \geq \lambda\}| \lesssim \|f\|_p^q \lambda^{-q} \text{ for all } f, \lambda > 0;$$

Clearly weak-type is weaker than strong-type. An even weaker estimate obtains if we restrict f to be a characteristic function:

$$|\{|T\chi_E| \geq \lambda\}| \lesssim |E|^{q/p}\lambda^{-q} \text{ for all } E, \lambda > 0; \tag{7}$$

in this case we say T is of restricted weak-type (p, q).

It is convenient to rephrase (7) in a more symmetric form.

Lemma 4.1. *Suppose* $1 < p, q < \infty$. *Then* (7) *holds if and only if one has*

$$|\langle T\chi_E, \chi_F\rangle| \lesssim |E|^{1/p}|F|^{1/q'} \tag{8}$$

for all sets E, F.

As a comparison, note that the strong-type estimate is equivalent to

$$|\langle Tf, g\rangle| \leq \|f\|_p\|g\|_{q'}.$$

Proof For our purposes, we only need the implication (7) $\implies$ (8); I leave the opposite implication as an exercise. (Hint: apply (8) to the set $F = \{Re(T\chi_E) > \lambda\}$).

We have

$$|\langle T\chi_E, \chi_F\rangle| \leq \int_F |T\chi_E(x)|\ dx.$$

We rewrite this as

$$\int_F \int_0^\infty \chi_{|T\chi_E|>\lambda}(x)\ d\lambda\ dx.$$

Interchanging the integrals, and then doing the x integration, this becomes

$$\int_0^\infty |\{x \in F : |T\chi_E(x)| > \lambda\}|\ d\lambda.$$

We have two estimates for the integrand. The first is just $|F|$. The second is $O(|E|^{q/p}\lambda^{-p})$. Thus we can estimate this integral by

$$O(\int_0^\infty \min(|F|, |E|^{q/p}\lambda^{-q})\ d\lambda).$$

The first term in the minimum is smaller when $\lambda \leq |E|^{1/p}|F|^{-1/q}$, and the second term is smaller otherwise. Breaking up the integral accordingly, and doing some freshman calculus, we eventually get (8). ■

Restricted weak-type estimates are clearly weaker than strong type estimates; however for interpolation purposes they are equally good.

Lemma 4.2 (Marcinkiewicz interpolation). *Suppose* $1 < p_0 < q_0 < \infty$, $1 < p_1 < q_1 < \infty$, $p_0 < p_1$, $q_0 < q_1$, *and and* T *is of restricted weak-type* (p_0, q_0) *and* (p_1, q_1). *Then* T *is of strong-type* (p_θ, q_θ) *for any* $0 < \theta < 1$, *where* $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$ *and similarly for the* q*s.*

Proof Fix θ, and write p, q for p_θ, q_θ. We have to show that

$$|\langle Tf, g\rangle| \lesssim \|f\|_p \|g\|_{q'}.$$

We may assume that f, g are non-negative, because any complex function is the linear combination of at four non-negative functions.

By re-arranging X and Y, we may assume that X and Y are the positive real line $(0, \infty)$ with the standard measure, and f, g are monotone decreasing functions. By a limiting argument we can assume that f, g are test functions.

We now write f as a linear combination of characteristic functions:

$$f = \int_0^\infty f(\lambda)\delta_\lambda \, d\lambda = -\int_0^\infty f'(\lambda)\chi_{[0,\lambda]} \, d\lambda$$

and similarly for g. By linearity, we thus have

$$\langle Tf, g\rangle = \int_0^\infty \int_0^\infty f'(\lambda) g'(\beta) \langle T\chi_{[0,\lambda]}, \chi_{[0,\beta]}\rangle \, d\lambda d\beta.$$

Fromm the previous lemma and our assumptions we have

$$|\langle T\chi_{[0,\lambda]}, \chi_{[0,\beta]}\rangle| \lesssim \lambda^{1/p_i}\beta^{1/{q'}_i}$$

for $i = 1, 2$. Thus we have

$$|\langle Tf, g\rangle| \lesssim \int_0^\infty \int_0^\infty f'(\lambda) g'(\beta) \min_{i=1,2}(\lambda^{1/p_i}\beta^{1/{q'}_i}) \, d\lambda d\beta.$$

Integrating by parts, this is[5]

$$C\int_0^\infty \int_0^\infty f(\lambda) g(\beta) \min_{i=1,2}(\lambda^{1/p_i - 1}\beta^{1/{q'}_i - 1}) \, d\lambda d\beta.$$

Let a, b be such that

$$1/{q'}_i = -a/p_i + ab \text{ for } i = 1, 2;$$

note that this implies that

$$1/q = -a/p + ab. \tag{9}$$

The integral now becomes

$$C\int_0^\infty \int_0^\infty f(\lambda) g(\beta) \min_{i=1,2}(\lambda/\beta^a)^{1/p_i}\beta^{ab}\frac{d\lambda}{\lambda}\frac{d\beta}{\beta}.$$

Making the change of variables $\eta = \beta^a$, this becomes

$$C\int_0^\infty \int_0^\infty f(\lambda) g(\eta^{1/a}) \min_{i=1,2}(\lambda/\eta)^{1/p_i}\eta^{b}\frac{d\lambda}{\lambda}\frac{d\eta}{\eta}.$$

Now making the change of variables $\gamma = \lambda/\eta$, this becomes

$$C\int_0^\infty \int_0^\infty f(\eta\gamma) g(\eta^{1/a}) \min_{i=1,2}\gamma^{1/p_i}\eta^{b}\frac{d\gamma}{\gamma}\frac{d\eta}{\eta}.$$

Interchanging the integrals, this is

$$C\int_0^\infty \min_{i=1,2}\gamma^{1/p_i} \int_0^\infty f(\eta\gamma) g(\eta^{1/a})\eta^{b-1} \, d\eta \frac{d\gamma}{\gamma}.$$

[5]Note that the discontinuity caused by the minimum does not cause a difficulty.

By Hölder, the inner integral is bounded by

$$(\int_0^\infty f(\eta\gamma)^p\ d\eta)^{1/p}(\int_0^\infty (g(\eta^{1/a})\eta^{b-1})^{p'}\ d\eta)^{1/p'}.$$

A tedious computation using (9) shows that this simplifies to

$$C\gamma^{-1/p}\|f\|_p\|g\|_{q'}.$$

Putting all this together, we get

$$|\langle Tf,g\rangle| \lesssim \|f\|_p\|g\|_{q'}\int_0^\infty \gamma^{-1/p}\min_{i=1,2}\gamma^{1/p_i}\ \frac{d\gamma}{\gamma}.$$

Since $p_0 < p < p_1$, we see that the integral converges, and we are done. ■

One can also obtain this theorem at an endpoint $p_0 = 1$, providing one replaces restricted weak-type by just weak-type. We leave this as an exercise to the reader.

5. Appendix II: The Hardy-Littlewood maximal inequality

In the main text we introduced the Hardy-Littlewood maximal operator

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

It is clear that $\|Mf\|_\infty \lesssim \|f\|_\infty$. However, more is true. Indeed, we have $\|Mf\|_p \lesssim \|f\|_p$ for all $1 < p \leq \infty$. To prove this, it suffices by Marcinkiewicz interpolation[6] to show that M is weak-type $(1,1)$:

$$|\{|Mf(x)| \geq \lambda\}| \lesssim \frac{\|f\|_1}{\lambda}.$$

This inequality is known as the *Hardy-Littlewood maximal inequality.*

To show this we first introduce the *dyadic Hardy-Littlewood maximal operator*

$$M_\Delta f(x) = \sup_{Q:x\in Q}\frac{1}{|Q|}\int_Q|f(y)|\ dy$$

where Q ranges over all dyadic cubes which contain x. We first show the *dyadic Hardy-Littlewood maximal inequality*

$$|\{|M_\Delta f(x)| \geq \lambda\}| \lesssim \frac{\|f\|_1}{\lambda}.$$

To do this we repeat the machinery of the previous week's notes. Call a dyadic cube Q *heavy* if $\int_Q |f(y)|\ dy \geq \lambda|Q|$, and call a heavy cube *maximal* if it is not contained

[6]Strictly speaking, M is not a linear operator, and so the version of Marcinkiewicz interpolation stated above does not apply. However, there is a *linearization* trick which gets around this. Specifically, one lets $r(x)$ be an arbitrary function, and then one defines the operator $M_{r(x)}$ by

$$M_{r(x)}f(x) := \frac{1}{|B(x,r(x))|}\int_{B(x,r(x))}f(y)\ dy.$$

This operator is linear in f, so if this operator is weak-type (1,1) then it is bounded on L^p. Then one takes suprema over all choices of $r(x)$ to obtain the L^p boundedness of M, for non-negative f at least. The general case then follows easily from the non-negative case.

in any other heavy cube. As in last week's notes, the maximal heavy cubes Q are all disjoint and are such that

$$\sum_Q |Q| \lesssim \frac{\|f\|_1}{\lambda}.$$

On the other hand, it is clear that the set $\{|M_\Delta f(x)| \geq \lambda\}$ is exactly equal to $\bigcup_Q Q$, so the claim follows.

Now to pass from the dyadic maximal inequality to the non-dyadic maximal inequality. To do this we use the $1/3$ *translation trick*, which appears to be first due to Michael Christ. For simplicity we illustrate the method in dimension $n = 1$.

By a limiting argument we may assume that f is supported on a compact set. By rescaling we may assume that f is supported on the unit ball. In this case the only balls which matter for the definition of M are those for which $r \lesssim 1$.

Define a $1/3$-shifted dyadic interval to be an interval of the form $I + 1/3$, where I is a dyadic interval. We can then define a $1/3$-shifted dyadic Hardy-Littlewood maximal operator $M_{\Delta,1/3}$ in the obvious manner, and it must also satisfy the weak-type inequality.

The key observation is that every interval $B(x,r)$ can be contained inside either a dyadic interval I, or a $1/3$ shifted interval $I + 1/3$, such that $r < |I| < 100r$. (Basically, one first tries to use a dyadic interval of size about $10r$ or so. This works unless an endpoint of dyadic interval lies inside $B(x,r)$. In this case we shift by $1/3$.) Because of this, we see that

$$Mf(x) \lesssim M_\Delta f(x) + M_{\Delta,1/3} f(x)$$

and the non-dyadic Hardy-Littlewood maximal inequality follows.

6. Homework for week 4 - due Friday, February 9

- Let $n \geq 1$, $1 < p < \infty$, and $0 < s \leq n/p$. Find a function f such that f is in $W^{s,p}(\mathbf{R}^n)$, but that f^2 is not in $W^{s,p}(\mathbf{R}^n)$.
- Show that the operator $\pi_1(f,g)$ defined by (5) is a bilinear Fourier multiplier whose symbol $m(\xi_1,\xi_2)$ equals 1 when $|\xi_1| \ll |\xi_2|$, equals 0 when $|\xi_1| \gg |\xi_2|$, and is a symbol of order 0.
- Let $s > 0$, and let D be the Fourier multiplier
$$\widehat{Df}(\xi) = (1 + |\xi|^2)^{s/2} \hat{f}(\xi).$$
Prove that
$$\|D(fg)\|_p \lesssim \|Df\|_q \|g\|_r + \|f\|_q \|Dg\|_r$$
whenever $1 < p, q, r < \infty$ is such that $1/p = 1/q + 1/r$. Explain why this is consistent with the fractional Leibnitz rule heuristic.

  (Note: one can either prove the above estimate directly using Littlewood-Paley theory, or one can break up fg into paraproducts and apply the Coifman-Meyer theorem. In the former case you may find the Hardy-Littlewood maximal inequality useful.)

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

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