LECTURE NOTES 6 FOR 247B

TERENCE TAO

1. Paraproducts - introduction

In the previous quarter we have focused primarily on linear (or sublinear) operators, which take one function f as input and return a function Tf as output. In this set of notes we shall consider some examples of *multilinear* (or more specifically *bilinear*) and *nonlinear* operators. Of course there are an infinite number of such operators, but we shall focus on operators related to the two model examples of such operators, the *pointwise product* operator

$$(f, g) \mapsto fg$$

and a *pointwise nonlinear operator*

$$f \mapsto F(f)$$

where $F : \mathbf{C} \to \mathbf{C}$ is a specific function (e.g. a power-type function $F(z) := |z|^{p-1}z$). These two operators and their variants, and their behaviour on Sobolev spaces, are particularly relevant in the theory of nonlinear PDE. Two model questions are as follows:

- For which values of $s, p, s_1, p_1, s_2, p_2, d$ is it true that whenever $f_1 \in W^{s_1,p_1}(\mathbf{R}^d)$ and $f_2 \in W^{s_2,p_2}(\mathbf{R}^d)$, one has $f_1f_2 \in W^{s,p}(\mathbf{R}^d)$?
- Let $F : \mathbf{C} \to \mathbf{C}$ be given. For which values of s, p, t, q is it true that whenever $f \in W^{s,p}(\mathbf{R}^d)$, one has $F(f) \in W^{t,q}(\mathbf{R}^d)$?

These types of questions turn out to be most easily answered by Littlewood-Paley decomposition, breaking up expressions such as the pointwise product into component pieces known as *paraproducts*. (The analogous decomposition for nonlinear functions $u \mapsto F(u)$ is *Bony's linearisation formula*.) The theory of such decompositions is known as the *paradifferential calculus*[1]. The term "paraproduct" is somewhat vaguely defined, but loosely speaking, paraproducts tend to be restricted versions of products in which only certain types of frequency interactions are permitted.

[1]This is the theory of multilinear constant coefficient differential operators and their generalisations. In contrast, the *pseudodifferential calculus* is the theory of linear variable coefficient differential operators and their generalisations. One could formulate a "parapseudodifferential calculus" encompassing multilinear variable coefficient differential operators, but this gets messy and not particularly enlightening.

Before we plunge into the details, let us give some informal discussion to motivate why some sort of decomposition is necessary. Let us prove the following specific assertion:

Proposition 1.1. *If $f, g \in \mathcal{S}(\mathbf{R}^3)$, then*
$$\|fg\|_{W^{1,3/2}(\mathbf{R}^3)} \lesssim \|f\|_{W^{1,2}(\mathbf{R}^3)} \|g\|_{W^{1,2}(\mathbf{R}^3)}. \tag{1}$$

This proposition is only stated for Schwartz functions, but it is not hard to use this estimate to then show that the same estimate is also true for all $f \in W^{1,2}(\mathbf{R}^3)$ and $g \in W^{1,2}(\mathbf{R}^3)$ by density arguments; we leave the details as an exercise to the reader.

To prove this proposition, let us normalise
$$\|f\|_{W^{1,2}(\mathbf{R}^3)} = \|g\|_{W^{1,2}(\mathbf{R}^3)}.$$
From Sobolev embedding we conclude
$$\|f\|_{L^2(\mathbf{R}^3)}, \|\nabla f\|_{L^2(\mathbf{R}^3)}, \|f\|_{L^6(\mathbf{R}^3)}, \|g\|_{L^2(\mathbf{R}^3)}, \|\nabla g\|_{L^2(\mathbf{R}^3)}, \|g\|_{L^6(\mathbf{R}^3)} \lesssim 1. \tag{2}$$
From Hölder's inequality this already gives
$$\|fg\|_{L^{3/2}(\mathbf{R}^3)} \leq \|f\|_{L^2(\mathbf{R}^3)} \|g\|_{L^6(\mathbf{R}^3)} \lesssim 1.$$
Since
$$\|fg\|_{W^{1,3/2}(\mathbf{R}^3)} \sim \|fg\|_{L^3(\mathbf{R}^3)} + \|\nabla(fg)\|_{L^3(\mathbf{R}^3)}$$
we see that it will now suffice to show that
$$\|\nabla(fg)\|_{L^{3/2}(\mathbf{R}^3)} \lesssim 1.$$
Comparing this estimate with (2) we observe that the derivative ∇ is on the outside of the product fg here, whereas in (2) the derivative is applied to f and g separately. But of course we can relate one to the other by the *Leibnitz rule* (or *product rule*)
$$\nabla(fg) = (\nabla f)g + f(\nabla g). \tag{3}$$
From Hölder we have
$$\|(\nabla f)g\|_{L^{3/2}(\mathbf{R}^3)} \leq \|\nabla f\|_{L^2(\mathbf{R}^3)} \|g\|_{L^6(\mathbf{R}^3)} \lesssim 1$$
and
$$\|f(\nabla g)\|_{L^{3/2}(\mathbf{R}^3)} \leq \|f\|_{L^6(\mathbf{R}^3)} \|\nabla g\|_{L^2(\mathbf{R}^3)} \lesssim 1$$
and so the claim follows from the triangle inequality. Note how the two terms $(\nabla f)g$ and $f(\nabla g)$, while similar, had to be treated in slightly different ways, and so $\nabla(fg)$ could not be treated directly by a single application of Hölder.

It is also instructive to work through (1) with a specific (but slightly informal) example. Let f and g be non-trivial bump functions of height 1 adapted to balls of radius $1/N$ and $1/M$ respectively for some $N, M \geq 1$ (the cases $N \leq 1$ or $M \leq 1$ require a slightly different treatment, in which the lower order terms in the Sobolev norms become more important, but we leave this to the reader). Then $\|f\|_{L^2(\mathbf{R}^3)} \sim N^{-3/2}$ and $\|g\|_{L^2(\mathbf{R}^3)} \sim M^{-3/2}$. The function f has "wavelength" $\sim 1/N$ and thus "frequency" $\sim N$; more concretely, the Fourier transform of f (which is a modulated Schwartz function of height N^{-3} adapted to the ball $B(0, N)$)

is concentrated on frequencies of magnitude comparable to N. Because of this, we expect ∇f to "behave like" Nf, and indeed ∇f is a bump function of height N adapted to the same ball of radius $1/N$. In particular it is not hard to see that $\|\nabla f\|_{L^2(\mathbf{R}^3)} \sim N^{-1/2}$, and so $\|f\|_{W^{1,2}(\mathbf{R}^3)} \sim N^{-1/2}$. Similarly $\|g\|_{W^{1,2}(\mathbf{R}^3)} \sim M^{-1/2}$. Observe that this is consistent with Sobolev embedding, for instance we have $\|f\|_{L^6(\mathbf{R}^3)} \sim N^{-1/2}$ and $\|g\|_{L^6(\mathbf{R}^3)} \sim M^{-1/2}$.

Now let us look at the product fg. At worst this will be a bump of height 1 adapted to a ball of radius $1/\max(N,M)$ (this is when the balls supporting f, g are nested; otherwise fg is either smaller than this, or vanishes entirely). Thus the "wavelength" here is $1/\max(N,M)$ and the "frequency" is $\max(N,M)$. This is tied to the fundamental observation that the frequency of a product is the sum of the frequencies of the factors:

$$e^{2\pi i\xi_1\cdot x} \times e^{2\pi i\xi_2\cdot x} = e^{2\pi i(\xi_1+\xi_2)\cdot x}$$

or in terms of Fourier transforms

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

So we expect to have

$$\|fg\|_{L^{3/2}(\mathbf{R}^3)} \lesssim \max(N,M)^{-2}$$

and

$$\|\nabla(fg)\|_{L^{3/2}(\mathbf{R}^3)} \lesssim \max(N,M)^{-1}$$

and so the inequality (1) becomes

$$\max(N,M)^{-1} \lesssim N^{-1/2}M^{-1/2}.$$

But this is easily verified by splitting into two cases $N \geq M$ and $N \leq M$ and recalling that $N, M \geq 1$. This splitting of cases is the analogue of the decomposition (3); notice in our specific example that when $N \geq M$ the term $(\nabla f)g$ dominates, whereas when $N \leq M$ the term $f(\nabla g)$ dominates. This example also shows that (1) is rarely sharp; even in the model case of bump functions, we only expect equality when the bumps have coincident support and with wavelength $\ll 1$.

Pretending for the moment that ∇ is invertible, the decomposition (3) can be recast as

$$fg = \nabla^{-1}((\nabla f)g) + \nabla^{-1}(f(\nabla g)). \tag{4}$$

The two expressions $\nabla^{-1}((\nabla f)g)$ and $\nabla^{-1}(f(\nabla g))$ are thus two components of the product fg, and are thus model examples of *paraproducts*. Roughly speaking, $\nabla^{-1}((\nabla f)g)$ is the portion of fg which favours the "high-low" interactions when a high-frequency component of f is multiplied with a low-frequency component of g, whereas $\nabla^{-1}(f(\nabla g))$ is the portion which favours the "low-high" interaction when a low-frequency component of f is multiplied with a high-frequency component of g. (The situation is more subtle with the "high-high" interactions in which high frequency components of f and g multiply and cancel in phase to create a low frequency contribution to fg. In this case $\nabla^{-1}((\nabla f)g)$ and $\nabla^{-1}(f(\nabla g))$ are quite large compared to fg but have opposite sign, and so largely cancel each other. Thus in this case, these two expressions are not really behaving like paraproducts, as they are *worse* than the initial product.)

To deal with more general Sobolev spaces we need to find extensions of (3) to other differential operators, for instance can we decompose $\langle\nabla\rangle^s(fg)$ in a similar fashion? For ∇^2 we can use the Leibnitz rule twice to give

$$\nabla^2(fg) = (\nabla^2 f)g + f(\nabla^2 g) + 2\nabla f\nabla g.$$

A useful heuristic from PDE is that the worst terms always are the highest order terms (ones involving the highest order of differentiation), so we arrive at the heuristic

$$\nabla^2(fg) \approx (\nabla^2 f)g + f(\nabla^2 g).$$

More generally we see

$$\nabla^k(fg) \approx (\nabla^k f)g + f(\nabla^k g)$$

for integer $k \geq 0$. Extrapolating from this, we could conjecture some sort of *fractional Leibnitz rule*

$$D(fg) \approx (Df)g + f(Dg)$$

for any positive order differential or pseudodifferential operator such as $\langle\nabla\rangle^s$ for $s \geq 0$. Of course this conjecture is not well formulated at present because the $\approx$ symbol is undefined. Nevertheless we shall see shortly that a version of this rule can indeed be made rigorous.

2. Coifman-Meyer multipliers

Just as linear Fourier multipliers are a good framework with which to study constant co-efficient differential operators and related objects, *multilinear Fourier multipliers* are a good framework to study (translation-invariant) multilinear operators. For simplicity we discuss only the bilinear case, although the multilinear case is quite similar. The starting point is the product formula

$$fg(x) = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} e^{2\pi i x\cdot(\xi_1+\xi_2)}\hat f(\xi_1)\hat g(\xi_2)\ d\xi_1 d\xi_2.$$

Inspired by this, we define the *bilinear multiplier* T_m for any (locally integrable, tempered) function $m : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{C}$ and Schwartz $f, g \in \mathcal{S}(\mathbf{R}^d)$ by the formula

$$T_m(f,g)(x) := \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} m(\xi_1,\xi_2)e^{2\pi i x\cdot(\xi_1+\xi_2)}\hat f(\xi_1)\hat g(\xi_2)\ d\xi_1 d\xi_2$$

or equivalently

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

Formally, T_m is the unique bilinear operator such that

$$T_m(e_{\xi_1}, e_{\xi_2}) = m(\xi_1,\xi_2)e_{\xi_1+\xi_2}$$

for all frequencies $\xi_1, \xi_2 \in \mathbf{R}^d$, where we use e_ξ to denote the character (or plane wave) $e_\xi(x) := e^{2\pi i x\cdot\xi}$. Thus bilinear Fourier multipliers multiply plane waves together (thus adding their frequencies), but also modulate their amplitude by a symbol $m(\xi_1,\xi_2)$.

In analogy with the linear case, we refer to m as the *symbol* of T_m. For instance, we have $fg = T_1(f,g)$, $(\nabla f)g = T_{i\xi_1}(f,g)$, $f(\nabla g) = T_{i\xi_2}(f,g)$, and $\nabla(fg) =$

$T_{i(\xi_1+\xi_2)}(f,g)$, where we extend the notation to vector valued m in the obvious manner. Note that as the map $m \mapsto T_m$ is linear in m, this gives a Fourier-analytic proof of the Leibnitz rule (3). The composition calculus is more complicated in the bilinear case than in the linear one, since there is not really a good notion of a composition of two bilinear operators (except perhaps to form a trilinear operator). Nevertheless, we have the useful identities

$$\begin{aligned} T_m(a(D)f,g) &= T_{a(\xi_1)m}(f,g) \\ T_m(f,a(D)g) &= T_{a(\xi_2)m}(f,g) \\ a(D)T_m(f,g) &= T_{a(\xi_1+\xi_2)m}(f,g) \end{aligned}$$

for all "reasonable" m, a (e.g. polynomial growth and smooth will suffice) and Schwartz f, g. Thus linear Fourier multipliers can easily be absorbed into bilinear Fourier multipliers.

For linear Fourier multipliers $a(D)$, we have the simple adjoint relationship $a(D)^* = \overline{a}(D)$. The situation is slightly more complicated for bilinear operators, because it turns out there are two adjoint operators (or more precisely, *transpose* operators). Indeed, for any Schwartz f, g, h and reasonable m (e.g. polynomial growth), a simple application of Fubini and Parseval shows that

$$\begin{aligned} \int_{\mathbf{R}^d} T_m(f,g)h &= \int_{\mathbf{R}^d} T_{m'}(g,h)f \\ &= \int_{\mathbf{R}^d} T_{m''}(h,f)g \\ &= \int_{\xi_1+\xi_2+\xi_3=0} m(\xi_1,\xi_2)\hat{f}(\xi_1)\hat{g}(\xi_2)\hat{h}(\xi_3)\ d\xi_1 d\xi_2 \end{aligned}$$

where

$$m'(\xi_2,\xi_3) := m(-\xi_2-\xi_3,\xi_2); \quad m''(\xi_3,\xi_1) = m(\xi_1,-\xi_1-\xi_3).$$

Similarly we have $T_m(f,g) = T_{m^t}(g,f)$, where $m^t(\xi_2,\xi_1) := m(\xi_1,\xi_2)$.

Recall the *Hörmander-Mikhlin multiplier theorem*, which established (among other things) the L^p boundedness properties of linear Fourier multipliers $m(D)$ provided that m obeyed the symbol estimates

$$|\nabla^j m(\xi)| \lesssim_{j,d} |\xi|^{-j}$$

for all $j \geq 0$ and $\xi \neq 0$ (actually we only needed this for finitely many j, namely $j = 0, 1, \ldots, d+2$). There is an analogue for bilinear Fourier multipliers (and indeed for multilinear multipliers), called the *Coifman-Meyer multiplier theorem.* It is not quite as universally applicable as its linear counterpart - we will have to also establish a more "dyadic" variant of this theorem in order to obtain satisfactory applications - but it is still highly useful, and captures the essential flavour of paradifferential calculus, in particular the decomposition of interactions into high-high, high-low, and low-high frequency interactions.

Definition 2.1 (Coifman-Meyer multipliers)**.** A *Coifman-Meyer symbol* is a function $m : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{C}$ obeying the estimates

$$|\nabla_{\xi_1}^{j_1} \nabla_{\xi_2}^{j_2} m(\xi_1,\xi_2)| \lesssim_{j_1,j_2,d} (|\xi_1|+|\xi_2|)^{-j_1-j_2}$$

for all[2] $j_1, j_2 \geq 0$. The corresponding operator T_m is a *Coifman-Meyer multiplier.*

- If we have $|\xi_1| \sim |\xi_2|$ on the support of m, we say that T_m is a *high-high Coifman-Meyer paraproduct* and denote T_m also by π_{hh} (thus different occurrences of π_{hh} can refer to different multipliers, analogously to the $O()$ notation).
- If we have $|\xi_1 + \xi_2| \sim |\xi_2|$ on the support of m, we say that T_m is a *low-high Coifman-Meyer paraproduct* and denote T_m also by π_{lh}.
- If we have $|\xi_1 + \xi_2| \sim |\xi_1|$ on the support of m, we say that T_m is a *high-low Coifman-Meyer paraproduct* and denote T_m also by π_{hl}.

Observe from the Leibnitz rule that the product of two Coifman-Meyer symbols is still a Coifman-Meyer symbol (and thus we also obtain analogous product properties for Coifman-Meyer paraproduct symbols). This fact is not as fundamental to the theory as the corresponding fact for linear symbols, because multiplication of bilinear symbols is not automatically related to any composition operation, but is still a handy fact to know nevertheless.

Let us give three model examples of paraproducts. All three involve bump functions $\psi_j(\xi)$ adapted to an annulus $|\xi| \sim 2^j$, and $\psi_{<j}(\xi)$ adapted to an annulus $|\xi| \lesssim 2^j$, and all sums are over the integers unless otherwise noted.

Example 2.2 (High-high product). *The operator $\pi_{hh}(f,g) := \sum_j (\psi_j(D)f)(\psi_j(D)g)$ is a high-high Coifman-Meyer paraproduct with symbol $m(\xi_1, \xi_2) = \sum_j \psi_j(\xi_1)\psi_j(\xi_2)$. Informally, this paraproduct multiplies high ($\sim 2^j$) frequencies of f with high frequencies of g to produce comparable or lower frequencies ($\lesssim 2^j$) in the output.*

Example 2.3 (Low-high product). *If the constant C is sufficiently large, the operator $\pi_{lh}(f,g) := \sum_j (\psi_{<j-C}(D)f)(\psi_j(D)g)$ will be a low-high Coifman-Meyer paraproduct with symbol $m(\xi_1, \xi_2) = \sum_j \psi_{<j-C}(\xi_1)\psi_j(\xi_2)$ (in particular we have $|\xi_1| \lesssim 2^{-C}|\xi_2|$, which is what ensures $|\xi_1 + \xi_2| \sim |\xi_2|$ if C is large enough. Informally, this paraproduct multiplies low ($\ll 2^j$) frequencies of f with high frequencies of g to produce high frequencies ($\lesssim 2^j$) in the output.*

Example 2.4 (High-low product). *If the constant C is sufficiently large, the operator $\pi_{hl}(f,g) := \sum_j (\psi_j(D)f)(\psi_{<j-C}(D)g)$ will be a high-low Coifman-Meyer paraproduct with symbol $m(\xi_1, \xi_2) = \sum_j \psi_j(\xi_1)\psi_{<j-C}(\xi_2)$.*

Example 2.5 (Non-Coifman-Meyer paraproducts). *In general, expressions such as $\sum_j \sum_k c_{j,k}(\psi_j(D)f)(\psi_k(D)f)$, where $c_{j,k}$ are bounded constants are* not *Coifman-Meyer multipliers (in contrast ot the linear situation, in which $\sum_j c_j \psi_j(D)$ is a Hörmander-Mikhlin multiplier). Indeed, for certain extremely pathological choices of $c_{j,k}$, such expressions do not obey the expected L^p estimates (a result of Grafakos and Kalton). Related to this is the fact that the Coifman-Meyer multiplier class is not closed under composition with Hörmander-Mikhlin multipliers; if T_m is a Coifman-Meyer multiplier and $a(D)$ is a Hörmander-Mikhlin multiplier then $T_m(a(D)f, g)$, $T_m(f, a(D)g)$, and $a(D)T_m(f,g)$ are in general not Coifman-Meyer multipliers. It is because of these facts that Coifman-Meyer multipliers do not, by themselves,*

[2] Actually, in practice only a finite number of j_1, j_2 depending on d are needed.

form a fully satifactory class of multipliers. Nevertheless we shall see later that it is possible to redress this failing to a large extent by introducing a class of "residual paraproducts" to absorb certain error terms.

Example 2.6 (Leibnitz rule). *Let's work in one dimension for simplicity. The two expressions $\nabla^{-1}((\nabla f)g$ and $\nabla^{-1}(f(\nabla g))$ are (formally) bilinear multipliers with symbols $\xi_1/(\xi_1+\xi_2)$ and $\xi_2/(\xi_1+\xi_2)$ respectively. These are certainly not Coifman-Meyer multipliers - they are not even locally integrable - but they behave somewhat like high-low and low-high paraproducts, in that the former is concentrated in the region $|\xi_1| \gtrsim |\xi_2|$ and the latter in the region $|\xi_2| \gtrsim |\xi_1|$.*

We remark that the transposes of a Coifman-Meyer multiplier, as defined above, are still Coifman-Meyer multipliers. These transpose operations can convert low-high paraproducts to high-high, etc.; we leave the precise description of the permutations as an exercise to the reader.

One pleasant fact about Coifman-Meyer multipliers is that they can always be decomposed into paraproducts:

Lemma 2.7 (Bony's paraproduct decomposition). *Let T_m be a Coifman-Meyer paraproduct. Then we have the decomposition*

$$T_m(f,g) = \pi_{hh}(f,g) + \pi_{hl}(f,g) + \pi_{lh}(f,g)$$

for some high-high, high-low, and low-high Coifman-Meyer paraproducts $\pi_{hh}, \pi_{hl}, \pi_{lh}$ respectively (which of course will depend on m. In particular, the pointwise product has such a decomposition:

$$fg = \pi_{hh}(f,g) + \pi_{hl}(f,g) + \pi_{lh}(f,g).$$

Proof There are several ways to perform such a decomposition. One such way involves a Littlewood-Paley decomposition $1 = \sum_j \psi_j$, where each ψ_j is a bump function adapted to the annulus $\{\xi : 2^{j-1} \leq |\xi| \leq 2^{j+1}\}$. The telescoping sums $\psi_{<j} := \sum_{k<j} \psi_k$ are then bump functions adapted to the ball $\{\xi : |\xi| \leq 2^{j+1}\}$. We then have the partition of unity

$$\begin{aligned}
1 &= \sum_j \sum_k \psi_j(\xi_1)\psi_k(\xi_2) \\
&= \sum_j \sum_{k<j-5} \psi_j(\xi_1)\psi_k(\xi_2) + \sum_j \sum_{|k-j|\leq 5} \psi_j(\xi_1)\psi_k(\xi_2) + \sum_j \sum_{k>j+5} \psi_j(\xi_1)\psi_k(\xi_2) \\
&= \sum_j \psi_j(\xi_1)\psi_{<j-5}(\xi_2) + \sum_j \sum_{|k-j|\leq 5} \psi_j(\xi_1)\psi_k(\xi_2) + \sum_k \psi_{<k-5}(\xi_1)\psi_k(\xi_2).
\end{aligned}$$

The three expressions on the right can be easily verified to be high-low, high-high, and low-high Coifman-Meyer symbols respectively. Multiplying both sides by $m(\xi_1, \xi_2)$ we obtain the result. ■

By the triangle inequality, we thus see that to estimate Coifman-Meyer multipliers it suffices to treat each type of paraproduct separately. This we now do.

3. Paraproduct estimates

The pointwise product $(f,g) \mapsto fg$ obeys the Hölder inequality

$$\|fg\|_{L^r(\mathbf{R}^d)} \leq \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}$$

whenever $0 < p,q,r \leq \infty$ and $1/p + 1/q = 1/r$. A useful and pleasant fact is that paraproducts $\pi_{hh}(f,g)$, $\pi_{hl}(f,g)$, $\pi_{lh}(f,g)$ obey analogues of this Hölder inequality for large ranges of p,q,r (though not quite as large as for the Hölder inequality itself). We illustrate this fact first with some model examples and then steadily build up to more general theorems, culminating in the Coifman-Meyer multiplier theorem.

We first observe a simple lemma which will be very useful in the computations which follow. It is a formalisation of the heuristic that band-limited functions (e.g. functions whose Fourier transform are localised to the ball $B(0,2^k)$) are "essentially" constant at scale 2^{-k}. Here, we shall use the Hardy-Littlewood maximal inequality to formalise the modifier "essentially".

Lemma 3.1 (Local constancy of band-limited functions)**.** *Let $j \in \mathbf{Z}$, and let $\psi_{\leq j}$ be a bump function adapted to a ball $\{|\xi| \lesssim 2^j\}$. Then we have*

$$|\psi_{\leq j}(D) f(y)| \lesssim_d \langle 2^j (y-x) \rangle^d Mf(x)$$

and more generally

$$|\nabla^k \psi_{\leq j}(D) f(y)| \lesssim_{k,d} 2^{jk} \langle 2^j (y-x) \rangle^d Mf(x) \tag{5}$$

for all $x,y \in \mathbf{R}^d$ and $k \geq 0$, where M is the Hardy-Littlewood maximal function. In particular, if f itself has Fourier transform supported in $B(0,2^j)$, then

$$|f(y)| \lesssim_d \langle 2^j (y-x) \rangle^d Mf(x)$$

and more generally

$$|\nabla^k f(y)| \lesssim_{k,d} 2^{jk} \langle 2^j (y-x) \rangle^d Mf(x).$$

Proof It suffices to prove (5). We can translate so that $x = 0$, and then rescale so that $j = 0$. Expressing $\nabla^k \psi_{\leq 0}(D)$ in physical space, we reduce to showing that

$$\Big|\int_{\mathbf{R}^d} \nabla^k \check{\psi}_{\leq 0}(y-z) f(z)\ dz\Big| \lesssim_{k,d} \langle y \rangle^d Mf(0).$$

We use the pointwise bound

$$\nabla^k \check{\psi}_{\leq 0}(y-z) = O_{k,d}(\langle y-z \rangle^{-100d})$$

and reduce to showing that

$$\int_{\mathbf{R}^d} \langle y-z \rangle^{-100d} |f(z)|\ dz \lesssim_d \langle y \rangle^d Mf(0).$$

For the region $\langle z \rangle \lesssim \langle y \rangle$ this follows by estimating $\langle y-z \rangle^{-100d}$ crudely by $O(1)$. For the region $\langle z \rangle \gg \langle y \rangle$ we estimate $\langle y-z \rangle^{-100d}$ by $\langle z \rangle^{-100d}$ and use dyadic decomposition in $|z|$. ∎

Now let us first consider the high-high paraproduct

$$\pi_{hh}(f,g) := \sum_j (\psi_j(D)f)(\psi_j(D)g)$$

in Example 2.2. From Cauchy-Schwarz we see that

$$|\pi_{hh}(f,g)| \leq (Sf)(Sg)$$

where S is the square function

$$Sf := (\sum_j |\psi_j(D)f|^2)^{1/2}.$$

Thus for any $0 < p, q, r \leq \infty$ with $1/p + 1/q = 1/r$, the ordinary Hölder inequality gives

$$\|\pi_{hh}(f,g)\|_{L^r(\mathbf{R}^d)} \leq \|Sf\|_{L^p(\mathbf{R}^d)} \|Sg\|_{L^q(\mathbf{R}^d)}.$$

If we also have $1 < p, q < \infty$, we thus see from the Littlewood-Paley inequality $\|Sf\|_{L^p(\mathbf{R}^d)} \sim_{d,p} \|f\|_{L^p(\mathbf{R}^d)}$ (and similarly for $\|Sg\|_{L^q(\mathbf{R}^d)}$) that

$$\|\pi_{hh}(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,d} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$

Thus this paraproduct enjoys a Hölder type inequality, as long as we impose the additional requirements $1 < p, q < \infty$.

Now let us consider the low-high paraprodyct

$$\pi_{lh}(f,g) := \sum_j (\psi_{<j-C}(D)f)(\psi_j(D)g).$$

If C is large enough, the product $(\psi_{<j-C}(D)f)(\psi_j(D)g)$ has Fourier transform supported in frequencies of magnitude $\sim 2^j$. So we can write

$$(\psi_{<j-C}(D)f)(\psi_j(D)g) = \tilde{\psi}_j(D)[(\psi_{<j-C}(D)f)(\psi_j(D)g)]$$

where $\tilde{\psi}_j$ is a bump function adapted to an annulus $\{|\xi| \sim 2^j\}$. Now suppose that we have $1 < p, q, r < \infty$. Then by the Littlewood-Paley inequality

$$\| \sum_j \tilde{\psi}_j(D) f_j \|_{L^r(\mathbf{R}^d)} \lesssim_{d,r} \|(\sum_j |f_j|^2)^{1/2}\|_{L^r(\mathbf{R}^d)}$$

we see that

$$\|\pi_{lh}(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim_{d,r} \|(\sum_j |(\psi_{<j-C}(D)f)(\psi_j(D)g)|^2)^{1/2}\|_{L^r(\mathbf{R}^d)}.$$

To deal with this, we observe from Lemma 3.1 that we have the pointwise bound

$$\psi_{<j-C}(D)f(x) \lesssim Mf(x)$$

for all $j \in \mathbf{Z}$ and $x \in \mathbf{R}^d$. Using this bound, we see that

$$(\sum_j |(\psi_{<j-C}(D)f)(\psi_j(D)g)|^2)^{1/2} \lesssim (Mf)(Sg)$$

and so on taking L^r norms and using Hölder, the Hardy-Littlewood inequality, and the Littlewood-Paley inequality we obtain

$$\|\pi_{lh}(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$

Note that in this case we also get the endpoint $p = \infty$, since the maximal function M is trivially bounded here.

A similar argument shows that the paraproduct π_{hl} in Example 2.4 obeys the estimate

$$\|\pi_{hl}(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}\|g\|_{L^q(\mathbf{R}^d)}.$$

whenever $1 < p, q, r < \infty$ (and one also gets the endpoint $q = \infty$).

We thus see that these types of paraproducts can be estimated by means of the Hölder, Hardy-Littlewood, and Littlewood-Paley inequalities. Now we establish the same estimates for more general multipliers.

Lemma 3.2 (High-high paraproducts)**.** *Let π_{hh} be a high-high paraproduct. Then we have*

$$\|\pi_{hh}(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,r,d} \|f\|_{L^p(\mathbf{R}^d)}\|g\|_{L^q(\mathbf{R}^d)}$$

whenever $1 < p, q < \infty$, $f, g \in \mathcal{S}(\mathbf{R}^d)$, and $1/r = 1/p + 1/q$.

Proof We allow all implied constants to depend on p, q, r, d. The strategy is to decompose the paraproduct so that it resembles Example 2.2.

We use a Littlewood-Paley decomposition $1 = \sum_j \psi_j^2$ to split

$$\pi_{hh}(f,g) = \sum_j \sum_k \pi_{hh}(\psi_j(D)\psi_j(D)f, \psi_k(D)\psi_k(D)g).$$

The operator $\pi_{hh}(\psi_j(D)f, \psi_k(D)g)$ vanishes unless $j = k + O(1)$, in which case it is bilinear multiplier whose symbol m_{jk} is a bump function adapted to the domain $\{\xi_1, \xi_2 = O(2^j)\}$. Thus by the triangle inequality

$$|\pi_{hh}(f,g)| \leq \sum_{j,k:j=k+O(1)} |T_{m_{jk}}(\psi_j(D)f, \psi_k(D)g)|.$$

To proceed further we need to do something about the symbol m_{jk}. We shall use Fourier decomposition on a cube in $\mathbf{R}^d \times \mathbf{R}^d$ of sidelength $C2^j$ for some sufficiently large C to write

$$m_{jk}(\xi_1, \xi_2) = \sum_{n_1,n_2 \in \mathbf{Z}^d} c_{n_1,n_2} e^{2\pi i(n_1\cdot\xi_1 + n_2\cdot\xi_2)/C2^j}$$

on the support of $\psi_j(\xi_1)\psi_k(\xi_2)$, where the Fourier coefficients c_{n_1,n_2} are rapidly decreasing, for instance

$$c_{n_1,n_2} \lesssim (1 + |n_1| + |n_2|)^{-100d}.$$

Applying this we see that

$$T_{m_{jk}}(\psi_j(D)f, \psi_k(D)g)(x) = \sum_{n_1,n_2\in\mathbf{Z}^d} c_{n_1,n_2}\psi_j(D)f(x - n_1/C2^j)\psi_k(D)g(x - n_2/C2^j)$$

and thus by the triangle inequality

$$|\pi_{hh}(f,g)(x)| \leq \sum_{j,k:j=k+O(1)} \sum_{n_1,n_2\in\mathbf{Z}^d} (1+|n_1|+|n_2|)^{-100d} |\psi_j(D)f(x-n_1/C2^j)||\psi_k(D)f(x-n_2/C2^j)|.$$

To deal with the shifts by $n_1/C2^j$ and $n_2/C2^j$ we use the reproducing formula

$$\psi_j(D)f = \tilde\psi_j(D)\psi_j(D)f$$

for some suitable bump function $\tilde{\psi}_j$ adapted to the ball $B(0, 2^{j+3})$ (say). From Lemma 3.1 we have

$$|\psi_j(D)f(x - n_1/C2^j)| \lesssim (1 + |n_1|)^d M(\psi_j(D)f)(x). \tag{6}$$

Similarly (recalling that $k = j + O(1)$)

$$|\psi_k(D)f(x - n_2/C2^j)| \lesssim (1 + |n_2|)^d M(\psi_k(D)g)(x).$$

Thus we have

$$|\pi_{hh}(f, g)| \lesssim \sum_{j,k:j=k+O(1)} \sum_{n_1,n_2 \in \mathbf{Z}^d} (1+|n_1|+|n_2|)^{-100d}(1+|n_1|)^d(1+|n_2|)^d |M\psi_j(D)f(x)||M\psi_k(D)f(x)|.$$

We can perform the n_1, n_2 sum to obtain

$$|\pi_{hh}(f, g)| \lesssim \sum_{j,k:j=k+O(1)} |M\psi_j(D)f||M\psi_k(D)f|.$$

By Schur's test or Young's inequality (or Cauchy-Schwartz) we conclude

$$|\pi_{hh}(f, g)| \lesssim (\sum_j |M\psi_j(D)f|^2)^{1/2} (\sum_k |M\psi_k(D)f|^2)^{1/2}.$$

Taking L^r norms and using Hölder we obtain

$$\|\pi_{hh}(f, g)\|_{L^r(\mathbf{R}^d)} \lesssim \|(\sum_j |M\psi_j(D)f|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \|(\sum_k |M\psi_k(D)f|^2)^{1/2}\|_{L^q(\mathbf{R}^d)}.$$

By the Fefferman-Stein maximal inequality followed by the Littlewood-Paley inequality we have

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

and similarly for g. The claim follows. ∎

Lemma 3.3 (Low-high paraproducts)**.** *Let π_{lh} be a low-high paraproduct. Then we have*

$$\|\pi_{lh}(f, g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,r,d} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}$$

whenever $1 < p \leq \infty$, $1 < q, r < \infty$, $f, g \in \mathcal{S}(\mathbf{R}^d)$, and $1/r = 1/p + 1/q$.

Proof We allow all implied constants to depend on p, q, r, d. We apply a similar (but not identical) strategy to the previous proof. Performing a Littlewood-Paley decomposition to g alone, we obtain

$$\pi_{lh}(f, g) = \sum_j \pi_{lh}(f, \psi_j(D)\psi_j(D)g).$$

The low-high nature of the paraproduct ensures that we may replace f by $\psi_{<j-C}(D)\psi_{<j-C}(D)f$ for some sufficiently large C. Thus we can write

$$\pi_{lh}(f, g) = \sum_j T_{m_j}(\psi_{<j-C}(D)f, \psi_j(D)g)$$

where $m_j(\xi_1,\xi_2) := m(\xi_1,\xi_2)\psi_{<j-C}(\xi_1)\psi_j(\xi_2)$. The Coifman-Meyer bounds on m imply that m_j is a bump function adapted to the region $\{\xi_1,\xi_2 = O(2^j)\}$. We can perform a Fourier decomposition as before and conclude

$$\pi_{lh}(f,g)(x) = \sum_j \sum_{n_1,n_2\in\mathbf{Z}^d} c_{n_1,n_2}\psi_{<j-C}(D)f(x-n_1/C2^j)\psi_j(D)g(x-n_2/C2^j).$$

Taking absolute values and using Lemma 3.1 as before we obtain

$$|\pi_{lh}(f,g)| \lesssim \sum_j \sum_{n_1,n_2\in\mathbf{Z}^d} (1+|n_1|+|n_2|)^{-100d}(1+|n_1|)^d(1+|n_2|)^d M(\psi_{<j-C}(D)f)M(\psi_j(D)g).$$

We can perform the n_1,n_2 summations, and also apply Lemma 3.1 to conclude

$$|\pi_{lh}(f,g)| \lesssim \sum_j M(Mf)M(\psi_j(D)g).$$

If we then take L^r norms and use Hölder, the Hardy-Littlewood inequality, the Fefferman-Stein inequality, and the Littlewood-Paley inequality we obtain the claim. ■

There is of course an analogus result for high-low paraproducts with the roles of p and q reversed. Combining all these results with Lemma 2.7 on the common domain of p,q,r we obtain

Corollary 3.4 (Coifman-Meyer multiplier theorem, easy case)**.** *Let T_m be a Coifman-Meyer multiplier. Then*

$$\|T_m(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,r,d} \|f\|_{L^p(\mathbf{R}^d)}\|g\|_{L^q(\mathbf{R}^d)}$$

whenever $1<p,q,r<\infty$ and $\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$.

4. The BMO theory

Corollary 3.4 is not fully satisfactory because it misses the endpoints when p,q,r equal 1 or ∞. For individual paraproducts, we already saw that we could obtain some of these endpoints. Now we extend the theory to these (more difficult) endpoints. It turns out that for some L^∞ endpoints we can in fact use the BMO norm instead. We begin with a model case.

Proposition 4.1. *If π_{lh} is the low-high paraproduct*

$$\pi_{lh}(f,g) = \sum_j (\psi_{<j-C}(D)f)(\psi_j(D)g)$$

for C sufficiently small, then for all Schwartz f,g we have

$$\|\pi_{lh}(f,g)\|_{L^2(\mathbf{R}^d)} \lesssim_d \|f\|_{L^2(\mathbf{R}^d)}\|g\|_{\mathrm{BMO}(\mathbf{R}^d)}.$$

Proof We normalise $\|g\|_{\mathrm{BMO}(\mathbf{R}^d)} = \|f\|_{L^2(\mathbf{R}^d)} = 1$, and allow all implicit constants to depend on d. Observe that $(\psi_{<j-C}(D)f)(\psi_j(D)g)$ has Fourier support in the

annulus $\{\xi \in \mathbf{R}^d : |\xi| \sim 2^j\}$. These annuli have only finite overlap, thus by Plancherel we have

$$\|\pi_{lh}(f,g)\|_{L^2(\mathbf{R}^d)}^2 \lesssim \sum_j \|(\psi_{<j-C}(D)f)(\psi_j(D)g)\|_{L^2(\mathbf{R}^d)}^2$$

so it suffices to show that

$$\sum_j \int_{\mathbf{R}^d} |\psi_{<j-C}(D)f|^2 |\psi_j(D)g|^2 \lesssim 1.$$

We shall dyadically decompose the $\psi_{<j-C}(D)f$ factor. Since

$$|\psi_{<j-C}(f)|^2 \lesssim \sum_k 2^{2k} 1_{|\psi_{<j-C}(D)f| \geq 2^k}$$

we have

$$\sum_j \int_{\mathbf{R}^d} |\psi_{<j-C}(D)f|^2 |\psi_j(D)g|^2 \lesssim \sum_k 2^{2k} \int_{\mathbf{R}^d} \sum_{j:|\psi_{<j-C}(D)f| \geq 2^k} |\psi_j(D)g|^2.$$

From Lemma 3.1 we see that if $|\psi_{<j-C}(D)f| \geq 2^k$ then $Mf > \varepsilon 2^k$ for some sufficiently small absolute constant $\varepsilon > 0$ (we will choose this constant later). Thus we reduce to showing that

$$\sum_k 2^{2k} \int_{Mf > \varepsilon 2^k} \sum_{j:|\psi_{<j-C}(D)f| \geq 2^k} |\psi_j(D)g|^2 \lesssim 1.$$

Now observe from the Hardy-Littlewood maximal inequality that

$$\sum_k 2^{2k} |\{Mf \geq \varepsilon^2 2^k\}| \lesssim_\varepsilon \int_{\mathbf{R}^d} Mf^2 \lesssim \int_{\mathbf{R}^d} f^2 \lesssim 1.$$

Thus it will suffice to show that

$$\int_{Mf > \varepsilon 2^k} \sum_{j:|\psi_{<j-C}(D)f| \geq 2^k} |\psi_j(D)g|^2 \lesssim_\varepsilon |\{Mf \geq \varepsilon^2 2^k\}|$$

for all k. By dividing f by 2^k we can assume $k = 0$. (This destroys the L^2 normalisation of f, but we will no longer need this normalisation), thus we wish to show

$$\int_{Mf > \varepsilon} \sum_{j:|\psi_{<j-C}(D)f| \geq 1} |\psi_j(D)g|^2 \lesssim_\varepsilon |\{Mf \geq \varepsilon^2\}|$$

for all f and for a sufficiently small absolute constant $\varepsilon > 0$.

By monotone convergence we can replace the set $\{Mf \geq \varepsilon\}$ by an arbitrary compact set E inside this set, as long as our bounds are of course uniform in E. By the usual Vitali covering lemma argument we can cover E by finitely many balls

$$E \subset 3B_1 \cup \ldots \cup 3B_N$$

where the balls $B_1, \ldots, B_N$ are disjoint with

$$⨍_{B_i} |f| = \varepsilon$$

and such that

$$⨍_{tB_i} |f| \leq \varepsilon \text{ for all } t > 1.$$

In particular we see (if ε is small enough) that

$$Mf > \varepsilon^2 \text{ on } B_i$$

and in particular

$$\sum_i |B_i| \leq |\{Mf > \varepsilon^2\}|.$$

Thus it will suffice to show that

$$\int_{3B_i} \sum_{j:|\psi_{<j-C}(D)f|\geq 1} |\psi_j(D)g|^2 \lesssim_\varepsilon |B_i|.$$

By translation we may centre B_i at the origin, thus $B_i = B(0, r_i)$. By replacing $f(x), g(x)$ with $f(r_i x), g(r_i x)$ we may then set $r_i = 1$. We now wish to show that

$$\int_{B(0,3)} \sum_{j:|\psi_{<j-C}(D)f|\geq 1} |\psi_j(D)g|^2 \lesssim_\varepsilon 1.$$

Now by hypothesis we know that

$$\int_{B(0,t)} |f| \lesssim \varepsilon t^d$$

for all $t \geq 1$. From this and the kernel bounds on $\psi_{<j-C}(D)$ and dyadic decomposition one can easily verify that

$$|\psi_{<j-C}(D)f(x)| \lesssim \varepsilon$$

in the low frequency case $j \leq 0$. Thus (if ε is small enough) we can restrict to $j \geq 0$, and it suffices to show that

$$\int_{B(0,3)} \sum_{j\geq 0} |\psi_j(D)g|^2 \lesssim_\varepsilon 1. \tag{7}$$

We would like to say that this estimate is invariant under subtraction of a constant from g (since the BMO norm is insensitive to such changes), which will allow us to normalise $⨍_{B(0,1)} g = 0$. Unfortunately this interferes with the hypothesis that g is Schwartz, which is necessary to define $\psi_j(D)g$ properly. This is fixable by a number of soft methods. Here is one: firstly we can restrict j to a finite range $0 \leq j \leq J$, so long as our final bounds are uniform in J. Now we replace g by $g - ⨍_{B(0,1)} g\phi(x/R)$, where R is a really large radius (larger than anything depending on g and J) and ϕ is a bump function adapted to $B(0,2)$ which equals 1 on $B(0,1)$. One can verify that in the limit $R \to \infty$ this does not affect either the BMO norm of g or the expression (7), and that this function has mean zero on $B(0,1)$. Thus without loss of generality we can assume that $⨍_{B(0,1)} g = 0$. From the John-Nirenberg inequality and the normalisation $\|g\|_{\mathrm{BMO}(\mathbf{R}^d)} = 1$ we then have

$$\int_{B(0,1)} |g|^2 \lesssim 1$$

but also more generally we have for all $t \geq 1$

$$⨍_{B(0,t)} g = O(t^d)$$

(actually, John-Nirenberg allows us to improve the right-hand side to $O_d(1+\log t)$, but we will not need this) and so

$$\int_{B(0,t)} |g|^2 \lesssim t^{3d}.$$

We split g into $g1_{B(0,4)}$ and $g(1-1_{B(0,4)})$. From the above estimate and the rapid decrease of the kernel of ψ_j we readily see that

$$|\psi_j(D)(g(1-1_{B(0,4)}))(x)| \lesssim 2^{-100j}$$

for all $j \geq 0$ and $x \in B(0,3)$. So this "global" portion is negligible and we only need to control the local part:

$$\int_{B(0,3)} \sum_{j\geq 0} |\psi_j(D)(g1_{B(0,4)})|^2 \lesssim_\varepsilon 1.$$

But the left-hand side is bounded by

$$\|(\sum_j |\psi_j(D)(g1_{B(0,4)})|^2)^{1/2}\|^2_{L^2(\mathbf{R}^d)} \lesssim \|g1_{B(0,4)}\|^2_{L^2(\mathbf{R}^d)} \lesssim 1$$

and the claim follows. ■

Corollary 4.2. *The same claim holds for more general low-high Coifman-Meyer paraproducts.*

Proof We again normalise $\|g\|_{\mathrm{BMO}(\mathbf{R}^d)} = \|f\|_{L^2(\mathbf{R}^d)} = 1$, and allow all implicit constants to depend on d. By arguing as in the preceding section, we can decompose

$$\pi_{lh}(f,g) = \sum_j T_{m_j}(\psi_{<j-C}(D)f, \psi_j(D)g)$$

where m_j is adapted to the region where $\{|\xi_1+\xi_2| \sim |\xi_2| \sim 2^j\}$. By Plancherel as before we have

$$\|\pi_{lh}(f,g)\|^2_{L^2(\mathbf{R}^d)} \lesssim \sum_j \|T_{m_j}(\psi_{<j-C}(D)f, \psi_j(D)g)\|^2_{L^2(\mathbf{R}^d)}$$

so it suffices to show that

$$(\sum_j \|T_{m_j}(\psi_{<j-C}(D)f, \psi_j(D)g)\|^2_{L^2(\mathbf{R}^d)})^{1/2} \lesssim 1.$$

On the other hand, by a Fourier decomposition we have

$$\|T_{m_j}(\psi_{<j-C}(D)f, \psi_j(D)g)\|_{L^2(\mathbf{R}^d)}$$
$$\lesssim \sum_{n_1,n_2\in\mathbf{Z}^d} (1+|n_1|+|n_2|)^{-100d} \|\psi_{<j-C}(D)f(x-n_1/C2^j)\psi_j(D)g(x-n_2/C2^j)\|_{L^2(\mathbf{R}^d)}$$

and so by the triangle inequality it suffices to show that

$$\sum_{n_1,n_2\in\mathbf{Z}^d} (1+|n_1|+|n_2|)^{-100d} (\sum_j \|\psi_{<j-C}(D)f(x-n_1/C2^j)\psi_j(D)g(x-n_2/C2^j)\|^2_{L^2(\mathbf{R}^d)})^{1/2} \lesssim 1.$$

But by repeating the proof of the previous proposition we can show that

$$(\sum_j \|\psi_{<j-C}(D)f(x-n_1/C2^j)\psi_j(D)g(x-n_2/C2^j)\|^2_{L^2(\mathbf{R}^d)})^{1/2} \lesssim (1+|n_1|+|n_2|)^{20d}.$$

The basic point is that the n_1 and n_2 shifts can be viewed as a phase modulation on the symbol of $\psi_{<j-C}$ and ψ_j. An inspection of the previous proof will reveal that one only needed to bound the first $10d$ derivatives of these symbols in order to close the argument, which is what leads to the loss of $(1+|n_1|+|n_2|)^{20d}$. The claim follows. ∎

We now see that if $g \in \mathrm{BMO}$, then the linear operator $f \mapsto \pi_{lh}(f,g)$ is bounded on L^2. In fact more is true:

Proposition 4.3. *Let $\|g\|_{\mathrm{BMO}(\mathbf{R}^d)} = O(1)$, and let π_{lh} be a low-high Coifman-Meyer paraproduct. Then the operator $f \mapsto \pi_{lh}(f,g)$ is a CZO. Similarly for any high-high Coifman-Meyer paraproduct π_{hh}, the operator $f \mapsto \pi_{hh}(f,g)$ is a CZO.*

Proof The second claim follows from the first by taking adjoints (note the adjoint of a CZO is a CZO). Since we have already established L^2 boundedness, the only remaining task is to establish singular integral bounds.

We will do this for the model example

$$\pi_{lh}(f,g) = \sum_j (\psi_{<j-C}(D)f)\psi_j(D)g;$$

the more general cases can be obtained by using the Fourier decompositions already employed previously. To avoid technicalities let us implicitly restrict j to a finite range, e.g. $-J \le j \le J$; our bounds will be independent of J and so standard limiting arguments will allow us to remove the finite range restriction. The kernel $K(x,y)$ of this operator is given by the formula

$$K(x,y) = \sum_j \hat{\psi}_{<j-C}(y-x)\psi_j(D)g(x).$$

Since $\psi_{<j-C}$ is a bump function adapted to the ball of radius $O(2^j)$, we have

$$|\nabla^k \hat{\psi}_{<j-C}(x-y)| \lesssim 2^{jk}2^{dj}\langle 2^j(x-y)\rangle^{-100d}$$

for $k=0,1$; since g is bounded in BMO one also easily verifies that

$$|\nabla^k \psi_j(D)g(x)| \lesssim 2^{jk}$$

for $k=0,1$. Thus

$$|\nabla^k_{x,y} K(x,y)| \lesssim \sum_j 2^{jk}2^{dj}\langle 2^j(x-y)\rangle^{-100d} \lesssim |x-y|^{d-k}$$

for $k=0,1$, and the claim follows. ∎

Corollary 4.4. *Corollary 3.4 also holds in the boundary cases when exactly one of p,q,r' is equal to ∞.*

Proof By duality we may assume that q is ∞. By Lemma 2.7 we may assume that T_m is either π_{hh}, π_{lh}, or π_{hl}. In the first two cases we can use the above Corollary and Calderón-Zygmund theory; in the last case we use Lemma 3.3 with the roles of f and g reversed. ∎

We remark that one can also show that Corollary 3.4 holds in the regime when $1 < p, q < \infty$ and r is allowed to be less than 1 (as in Lemma 3.2), but this is a little trickier to show (though in much the same spirit as the earlier results) and will not be done here.

5. Residual paraproducts

As mentioned earlier, Coifman-Meyer multipliers have the disadvantage that they are not closed under composition with linear Fourier multipliers such as Hörmander-Mikhlin multipliers. The following class of multipliers performs somewhat better in this regard[3].

Definition 5.1 (Residual paraproducts)**.** A *residual symbol* is a function $m : \Gamma \to \mathbf{C}$ on the space $\Gamma := \{(\xi_1, \xi_2, \xi_3) \in \mathbf{R}^d \times \mathbf{R}^d \times \mathbf{R}^d : \xi_1 + \xi_2 + \xi_3 = 0\}$ obeying the estimates

$$|\nabla_1^{j_1} \nabla_2^{j_2} \nabla_3^{j_3} m(\xi_1, \xi_2, \xi_3)| \lesssim_{j_1,j_2,j_3,d,\varepsilon} [\frac{\min(|\xi_1|, |\xi_2|, |\xi_3|)}{\max(|\xi_1|, |\xi_2|, |\xi_3|)}]^\varepsilon \prod_{abc=123,231,312} \min(|\xi_b|, |\xi_c|)^{-j_a} \tag{8}$$

for all[4] $j_1, j_2, j_3 \geq 0$ and all non-zero $\xi_1, \xi_2, \xi_3 \in \mathbf{R}^d$ and some $\varepsilon > 0$, where ∇_1 is the gradient in the directions[5] $\xi_1 = const$, etc. The corresponding multiplier T_m (where we abuse notation and write $m(\xi_1, \xi_2) := m(\xi_1, \xi_2, -\xi_1 - \xi_2)$) will be called a *residual paraproduct* and will be denoted π^r. We can classify these residual paraproducts further as high-high, low-high, and high-low residual paraproducts π^r_{hh}, π^r_{lh}, π^r_{hl} respectively.

Examples 5.2. The symbol

$$[\frac{\min(|\xi_1|, |\xi_2|, |\xi_3|)}{\max(|\xi_1|, |\xi_2|, |\xi_3|)}]^\varepsilon |\xi_1|^{it_1} |\xi_2|^{it_2} |\xi_3|^{it_3}$$

is a residual symbol for any bounded real t_1, t_2, t_3. The product of a residual symbol with either a residual symbol or a Coifman-Meyer symbol is another residual symbol. The operator

$$\pi^r(f, g) := \sum_{j_1,j_2,j_3} c_{j_1,j_2,j_3} \psi_{j_3}(D)[(\psi_{j_1}(D) f)(\psi_{j_2}(D) g)],$$

where each ψ_j is a bump function adapted to the annulus $\{|\xi| \sim 2^j\}$, is a residual operator as long as the constants c_{j_1,j_2,j_3} obey the decay condition

$$c_{j_1,j_2,j_3} = O(2^{-\varepsilon(\max(j_1,j_2,j_3) - \min(j_1,j_2,j_3))}).$$

Remark 5.3. To understand the estimates (8), let us work for instance in the "low-high" case when $|\xi_2| \sim |\xi_3|$ (and so $|\xi_1| \lesssim |\xi_2|$). Then (8) asserts that m is bounded by $(|\xi_1|/|\xi_2|)^\varepsilon$ (so it decays a little away from the diagonal $|\xi_1| \sim |\xi_2|$), and that one can move each of ξ_1, ξ_2, ξ_3 by a small multiple of $|\xi_1|, |\xi_2|, |\xi_3|$ respectively (staying in Γ, of course) without encountering any significant fluctuations or irregularity in the symbol.

[3] This notation is not standard in the literature.

[4] Actually, in practice only a finite number of j_1, j_2 depending on d are needed.

[5] If one dropped the ξ_3 variable and viewed m as a function purely of ξ_1 and ξ_2, then (up to irrelevant constants) $\nabla_1 = \nabla_{\xi_2}$, $\nabla_2 = \nabla_{\xi_1}$, and $\nabla_3 = \nabla_{\xi_1} + \nabla_{\xi_2}$.

The residual and Coifman-Meyer paraproducts are incomparable; residual paraproducts oscillate too much to be Coifman-Meyer, and Coifman-Meyer paraproducts do not have the decay factor $[\frac{\min(|\xi_1|,|\xi_2|,|\xi_3|)}{\max(|\xi_1|,|\xi_2|,|\xi_3|)}]^\varepsilon$ which is crucial to residual paraproducts. Unfortunately, it turns out that eliminating this factor can cause the multipliers to cease obeying good estimates (see Proposition 5.5 below). However, with this decay we have good estimates:

Proposition 5.4 (Residual multiplier theorem). *Let π^r be a residual multiplier. Then*

$$\|\pi^r(f,g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,r,d} \|f\|_{L^p(\mathbf{R}^d)}\|g\|_{L^q(\mathbf{R}^d)}$$

whenever $1 \leq p,q,r \leq \infty$ and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, with $(p,q) \neq (\infty,\infty),(\infty,1),(1,\infty)$.

Proof We suppress all dependence of constants on p,q,r,d. By applying Lemma 2.7 (which applies just as easily to residual multipliers as to Coifman-Meyer multipliers) we may assume that π^r is a high-high, low-high, or high-low paraproduct. By duality we may reduce to the low-high case. We use Littlewood-Paley decomposition to split

$$\pi^r(f,g) = \sum_{j_1,j_2,j_3} \psi_{j_3}(D) T_{m_{j_1,j_2,j_3}}(\psi_{j_1}(D)f, \psi_{j_2}(D)g)$$

where

$$m_{j_1,j_2,j_3}(\xi_1,\xi_2,\xi_3) = m(\xi_1,\xi_2,\xi_3)\prod_{i=1}^2 \psi_{j_i}(\xi_i).$$

Since we are in the low-high case we may take $j_2 = j_3 + O(1)$ and $j_1 \leq j_2 + O(1)$. Let us write $j_1 = j_2 - k$, then we see that $m_{j_1,j_2,j_3} = 2^{-\varepsilon k} m'_{j_1,j_2,j_3}$ where m' is a bump function adapted to the region $\{(\xi_1,\xi_2,-\xi_1-\xi_2) : |\xi_1| \sim 2^{j_1}; |\xi_2| \sim 2^{j_2}\}$. By the triangle inequality it thus suffices to show that

$$\| \sum_{j_1,j_2,j_3 : j_1 = j_2 - k, j_3 = j_2 + l} \psi_{j_3}(D) T_{m'_{j_1,j_2,j_3}}(\psi_{j_1}(D)f, \psi_{j_2}(D)g)\|_{L^r(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}\|g\|_{L^q(\mathbf{R}^d)}$$

for all $k \geq -O(1)$ and $l = O(1)$.

Fix k,l. By the Littlewood-Paley inequality the left-hand side is

$$\lesssim \|(\sum_{j_2} |T_{m'_{j_2-k,j_2,j_2+l}}(\psi_{j_2-k}(D)f, \psi_{j_2}(D)g)|^2)^{1/2}\|_{L^r(\mathbf{R}^d)}.$$

By the usual Fourier decomposition we can estimate

$$\begin{aligned}&|T_{m'_{j_2-k,j_2,j_2+l}}(\psi_{j_2-k}(D)f, \psi_{j_2}(D)g)|(x) \\ &\quad\lesssim \sum_{n_1,n_2} (1+|n_1|+|n_2|)^{-100d} |\psi_{j_2-k}(D)f(x - n_1/C2^{j_2-k})||\psi_{j_2}(D)g(x-n_2/C2^{j_2})|\end{aligned}$$

and then by using (6) (or Lemma 3.1) we obtain

$$|T_{m'_{j_2-k,j_2,j_2+l}}(\psi_{j_2-k}(D)f, \psi_{j_2}(D)g)|(x) \lesssim \sum_{n_1,n_2} (1+|n_1|+|n_2|)^{-100d}(1+|n_1|)^d(1+|n_2|)^d (M\psi_{j_2-k}(D)f)(M\psi_{j_2}(D)g)$$

The n_1,n_2 sum is $O(1)$ and can be discarded. One then uses Cauchy-Schwarz, Hölder, Fefferman-Stein, and Littlewood-Paley as before. ■

Finally, we give an example which shows that multiplier estimates can sometimes fail.

Proposition 5.5 (Grafakos-Kalton example)**.** *Let $1 < p, q, r < \infty$ be such that $1/p + 1/q = 1/r$. The estimate*

$$\| \sum_{j_1,j_2,j_3} c_{j_1,j_2,j_3} \psi_{j_3}(D)(\psi_{j_1}(D)f, \psi_{j_2}(D)g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,r,d} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}$$

does not *hold uniformly for all choices of uniformly bounded constants c_{j_1,j_2,j_3}, and for bump functions ψ_j uniformly adapted to annuli $\{|\xi| \sim 2^j\}$.*

Proof To simplify the notation slightly we work in one dimension $d = 1$. The idea is to assume the estimate is false, and then (by inspecting various frequency limits of the estimate) derive increasingly ridiculous (and eventually patently false) estimates as a result.

If the estimate failed, then

$$\| \sum_{j_2 > 0} (\psi_{j_2}(D)g)(\sum_{j_1 < 0} c_{j_1,j_2} \psi_{j_1}(D)f)\|_{L^r(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})} \|g\|_{L^q(\mathbf{R})}$$

whenever c_{j_1,j_2} are bounded, where we shall suppress all dependence of implicit constants on p, q, r, d. Thus if we let $T_1, \ldots, T_N$ be an arbitrary collection of Hörmander-Mikhlin multipliers, each of the form

$$T_n = \sum_{j<0} c_{j,n} \psi_j(D)$$

then we have

$$\| \sum_{n=1}^N (T_n f) \psi_{j_n}(D) g\|_{L^r(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})} \|g\|_{L^q(\mathbf{R})}$$

for arbitrary distinct positive $j_1, \ldots, j_N$. By the Littlewood-Paley inequality this would imply

$$\|(\sum_{n=1}^N |T_n f|^2 |\psi_{j_n}(D) g|^2)^{1/2}\|_{L^r(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})} \|g\|_{L^q(\mathbf{R})}.$$

Now we specialise to a function g of the form

$$g = \sum_{n=1}^N e^{2\pi i 2^{j_n} x} g_n(x)$$

where g_n are fixed Schwartz functions with compactly supported Fourier transform. If the j_n are sufficiently large and separated from each other (depending on the g_n), we have

$$|\psi_{j_n}(D) g| = |g_n|$$

and (by the Littlewood-Paley inequality)

$$\|g\|_{L^q(\mathbf{R})} \sim \|(\sum_{n=1}^N |g_n|^2)^{1/2}\|_{L^q(\mathbf{R})}.$$

We thus conclude that

$$\|(\sum_{n=1}^N |T_n f|^2 |g_n|^2)^{1/2}\|_{L^r(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})} \|(\sum_{n=1}^N |g_n|^2)^{1/2}\|_{L^q(\mathbf{R})}.$$

By density this estimate is in fact true for arbitrary functions g_n. If we set $g_n(x) := g(x) 1_{n=n(x)}$, where $n(x)$ is the index n which maximises $|T_n f(x)|$ for each x, we have

$$(\sum_{n=1}^N |T_n f|^2 |g_n|^2)^{1/2} = |g| \sup_{1 \leq n \leq N} |T_n f|$$

and thus

$$\||g| \sup_{1 \leq n \leq N} |T_n f|\|_{L^r(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})} \|g\|_{L^q(\mathbf{R})}.$$

This is true for all g, so by the converse Hölder inequality we obtain a "grand universal maximal inequality"

$$\| \sup_{1 \leq n \leq N} |T_n f|\|_{L^p(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})}.$$

Since N is arbitrary, we can use monotone convergence and conclude

$$\| \sup_{|c_j| \leq 1} |\sum_{j<0} c_j \psi_j(D) f|\|_{L^p(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})}.$$

But

$$\sup_{|c_j| \leq 1} |\sum_{j<0} c_j \psi_j(D) f| = \sum_{j<0} |\psi_j(D) f|$$

and so we have shown

$$\| \sum_{j<0} |\psi_j(D) f|\|_{L^p(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})}.$$

By rescaling we may replace $j < 0$ by $j < J$ for any J; letting $J \to \infty$ and using monotone convergence we conclude

$$\| \sum_{j} |\psi_j(D) f|\|_{L^p(\mathbf{R})} \lesssim \|f\|_{L^p(\mathbf{R})}.$$

Setting $f(x) = \sum_{n=1}^N e^{2\pi i 2^{j_n} x} f_n(x)$ for some band-limited f_n as before, and letting the j_n get widely spaced and go to infinity, we conclude

$$\| \sum_{n=1}^N |f_n|\|_{L^p(\mathbf{R})} \lesssim \|(\sum_{n=1}^N |f_n|^2)^{1/2}\|_{L^p(\mathbf{R})}$$

for all such f_n, and hence (by limiting arguments) for arbitrary f_n. But this is clearly false, as can be seen for instance by setting $f_n := 1_{[0,1]}$. ∎

6. The paradifferential calculus

Now we put all these paraproduct estimates to work. Let us recall some definitions of linear symbols.

Definition 6.1. Let $s \geq 0$. A *homogeneous symbol of order s* is a symbol $m : \mathbf{R}^d \to \mathbf{C}$ which obeys the estimates

$$|\nabla^j m(\xi)| \lesssim_{j,s,d} |\xi|^{s-j}$$

for all $j \geq 0$ and $\xi \neq 0$, whereas an *inhomogeneous symbol of order s* is a symbol $m : \mathbf{R}^d \to \mathbf{C}$ which obeys the estimates

$$|\nabla^j m(\xi)| \lesssim_{j,s,d} \langle \xi \rangle^{s-j}.$$

The corresponding Fourier multipliers $m(D)$ are referred to as *homogeneous* and *inhomogeneous Fourier multipliers of order s* respectively.

These multipliers almost commute with paraproducts in certain ways:

Lemma 6.2 (Kato-Ponce type commutator identities)**.** *Let $s \in \mathbf{R}$, and let D^s be an inhomogeneous symbol of order s.*

- *If π_{lh} is a low-high paraproduct, then*
$$D^s \pi_{lh}(f, g) = \pi_{lh}(f, D^s g) + \tilde{\pi}_{lh}(\nabla f, \langle \nabla \rangle^{s-1} g) = \pi'_{lh}(f, \langle \nabla \rangle^s g)$$
for some other (vector-valued) low-high paraproduct $\tilde{\pi}_{lh}$ and low-high paraproduct π'_{lh}.
- *If π_{hl} is a high-low paraproduct, then*
$$D^s \pi_{hl}(f, g) = \pi_{hl}(D^s f, g) + \tilde{\pi}_{hl}(\langle \nabla \rangle^{s-1} g, \nabla f) = \pi'_{hl}(D^s, g)$$
for some high-low paraproducts $\tilde{\pi}_{lh}$, π'_{lh}.
- *If π_{hh} is a high-high paraproduct, then*
$$\pi_{hh}(D^s f, g) = \pi_{hh}(f, D^s g) + \tilde{\pi}_{hh}(\nabla f, \langle \nabla \rangle^{s-1} g) = \pi'_{hh}(f, \langle \nabla \rangle^s g)$$
for some high-high paraproducts $\tilde{\pi}_{hh}$, π'_{hh}.

The implied constants in the symbol bounds for $\tilde{\pi}_{lh}$, etc. may depend on s. Similar claims hold for homogeneous symbols $\dot{D}^s$ of order s, but with $\langle \nabla \rangle$ replaced by $|\nabla|$.

Proof We just prove the low-high inhomogeneous case, as the other cases are similar. Writing the symbol of D^s as $m^s(\xi)$ and the symbol of π_{lh} as $m_{lh}(\xi_1, \xi_2)$, we observe that the bilinear operator $D^s \pi_{lh}(f, g) - \pi_{lh}(f, D^s g)$ has symbol

$$m_{lh}(\xi_1, \xi_2)[m^s(\xi_1 + \xi_2) - m^s(\xi_2)].$$

This is supported in the region $|\xi_1 + \xi_2| \sim |\xi_2|$. Using Littlewood-Paley multipliers, we can subdivide further into the regions $|\xi_1| \leq \frac{1}{2}|\xi_2|$ and $|\xi_1| \geq \frac{1}{4}|\xi_2|$. Suppose we are in the former region. Then by the fundamental theorem of calculus we have

$$m_{lh}(\xi_1, \xi_2)[m^s(\xi_1 + \xi_2) - m^s(\xi_2)] = \int_0^1 \xi_1 \cdot m_{lh}(\xi_1, \xi_2) \nabla m^s(t\xi_1 + \xi_2)\ dt.$$

One can easily verify that $m_{lh}(\xi_1, \xi_2) \nabla m^s(t\xi_1 + \xi_2) \langle 2\pi \xi_2 \rangle^{1-s}$ is a low-high Coifman-Meyer paraproduct uniformly for $t \in [0, 1]$, and the claim follows (using Minkowski's inequality to deal with the integration in t).

Now suppose instead we are in the region $|\xi_1| \geq \frac{1}{4}|\xi_2|$, which when combined with the low-high nature of π_{lh} implies that $\xi_1, \xi_2, \xi_1 + \xi_2$ are all comparable. In this case we cannot use the fundamental theorem of calculus as before because the line segment $\{t\xi_1 + \xi_2 : 0 \leq t \leq 1\}$ can pass close to the origin. Instead, we write

$$m_{lh}(\xi_1, \xi_2)[m^s(\xi_1 + \xi_2) - m^s(\xi_2)] = \xi_1 \cdot \frac{\xi_1}{|\xi_1|^2}(m_{lh}(\xi_1, \xi_2)[m^s(\xi_1 + \xi_2) - m^s(\xi_2)]).$$

One easily verifies that $\frac{\xi_1}{|\xi_1|^2}(m_{lh}(\xi_1, \xi_2)[m^s(\xi_1+\xi_2)-m^s(\xi_2)])\langle 2\pi\xi_2\rangle^{1-s}$ is a low-high Coifman-Meyer multiplier (without attempting to exploit any cancellation between the m^s terms), and the claim follows.

To obtain the cruder representation of $\pi'_{lh}(f, \langle\nabla\rangle^s g)$, one simply notes that $m^s(\xi_1 + \xi_2)m_{lh}(\xi_1, \xi_2)\langle\xi_2\rangle^{-s}$ is a Coifman-Meyer low-high symbol. ■

One can also move positive-order operators from the low frequency to the high, leaving a residual error:

Lemma 6.3 (Moving derivatives)**.** *Let $s = s_1 + s_2 > 0$ for some $s_1, s_2 \geq 0$, and let D^s be an inhomogeneous symbol of order s.*

- *If π_{lh} is a low-high paraproduct, then*
$$\pi_{lh}(D^s f, g) = \langle\nabla\rangle^{s_1}\pi^r_{lh}(f, \langle\nabla\rangle^{s_2} g)$$
for some residual low-high paraproduct π^r_{lh}.
- *If π_{hl} is a high-low paraproduct, then*
$$\pi_{hl}(f, D^s g) = \langle\nabla\rangle^{s_1}\pi^r_{hl}(\langle\nabla\rangle^{s_2} f, g)$$
for some residual high-low paraproduct π^r_{hl}.
- *If π_{hh} is a high-high paraproduct, then*
$$D^s\pi_{hh}(f, g) = \pi^r_{hh}(\langle\nabla\rangle^{s_1} f, \langle\nabla\rangle^{s_2} g)$$
for some residual high-high paraproduct π^r_{hh}.

The implied constants in the symbol bounds for $\tilde{\pi}_{lh}$, etc. may depend on s. Similar claims hold for homogeneous symbols $\dot{D}^s$ of order s, but with $\langle\nabla\rangle$ replaced by $|\nabla|$.

Proof We again just prove the low-high case. If we let $m^s(\xi)$ and $m_{lh}(\xi_1, \xi_2)$ denote the symbols of D^s and π_{lh} as before, then we see that π^r_{lh} will be a multiplier with symbol

$$\langle 2\pi(\xi_1 + \xi_2)\rangle^{-s_1} m^s(\xi_1)\langle 2\pi\xi_2\rangle^{-s_2} m_{lh}(\xi_1, \xi_2).$$

But one easily verifies that this is a residual low-high symbol (with $\varepsilon = s$). ■

There are analogues of the above two lemmas for residual paraproducts, but we shall leave them to the reader as we do not need them here.

Let us now put all of these above estimates to work. Let D^s be an inhomogeneous Fourier multiplier of order s. Applying D^s to the decomposition of the product in Lemma 2.7 we have

$$D^s(fg) = D^s\pi_{hh}(f, g) + D^s\pi_{lh}(f, g) + D^s\pi_{hl}(f, g)$$

and similarly

$$(D^s f)g = \pi_{hh}(D^s f, g) + \pi_{lh}(D^s f, g) + \pi_{hl}(D^s f, g)$$

and

$$f(D^s g) = \pi_{hh}(f, D^s g) + \pi_{lh}(f, D^s g) + \pi_{hl}(f, D^s g).$$

Subtracting and using Lemma 6.2, we have the *fractional Leibnitz rule*

$$D^s(fg) = (D^s f)g + f(D^s g) + T(f, g)$$

where T is a bilinear operator of the form

$$\begin{aligned} T(f, g) = {} & D^s \pi_{hh}(f, g) - \pi_{hh}(D^s f, g) - \pi_{hh}(f, D^s g) \\ & + \tilde{\pi}_{lh}(\nabla f, \langle \nabla \rangle^{s-1} g) - \pi_{lh}(D^s f, g) + \tilde{\pi}_{hl}(\langle \nabla \rangle^{s-1} f, \nabla g) \\ & - \pi_{hl}(f, D^s g). \end{aligned}$$

This operator can be simplified a fair bit by using the above identities. For instance one can express T in the form

$$T(f, g) = \pi(\langle \nabla \rangle^\theta f, l \langle \nabla \rangle^{s-\theta} g) + \pi'(\langle \nabla \rangle^{s-\theta} f, l \langle \nabla \rangle^\theta g)$$

for any $0 < \theta < \min(s, 1)$, where π, π' are linear combinations of Coifman-Meyer and residual paraproducts; we leave this as an exercise to the reader. This should be compared with the usual Leibnitz rule

$$\nabla^k(fg) = (\nabla^k f)g + f(\nabla^k g) + O(|\nabla f||\nabla^{k-1} g| + \ldots + |\nabla^{k-1} f||\nabla g|)$$

for integer $k \geq 1$.

The fractional Leibnitz rule leads to a number of product estimates in Sobolev spaces. A complete list of such estimates would be impossible here, so let us just give a sample:

Proposition 6.4. *Let $0 < s < 1$ and let $1 < p, q, r < \infty$ be such that $1/p + 1/q = 1/r$, and let D^s be a Fourier multiplier of order s. Let $s_1, s_2 > 0$ be such that $s_1 + s_2 = s$. Then*

$$\|D^s(fg) - (D^s f)g - f(D^s g)\|_{L^r(\mathbf{R}^d)} \lesssim_{p,q,r,s,s_1,s_2,d} \|f\|_{W^{s_1,p}(\mathbf{R}^d)} \|g\|_{W^{s_2,q}(\mathbf{R}^d)}$$

and

$$\|fg\|_{W^{r,s}(\mathbf{R}^d)} \lesssim_{p,q,s,d} \|f\|_{W^{s,p}(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)} + \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{W^{s,q}(\mathbf{R}^d)}$$

for all Schwartz f, g.

Proof For the first claim, we see from the previous discussion that

$$D^s(fg) - (D^s f)g - f(D^s g) = \pi(\langle \nabla \rangle^{s_1} f, \langle \nabla \rangle^{s_2} g)$$

for some linear combination π of Coifman-Meyer and residual paraproducts, and so by applying Corollary 3.4 and Proposition 5.4

$$\|D^s(fg) - (D^s f)g - f(D^s g)\|_{L^r(\mathbf{R}^d)} \lesssim \|\langle \nabla \rangle^{s_1} f\|_{L^p(\mathbf{R}^d)} \|\langle \nabla \rangle^{s_2} g\|_{L^q(\mathbf{R}^d)}$$

hence the claim.

For the second claim, we need to estimate $\|\langle\nabla\rangle^s(fg)\|_{L^r(\mathbf{R}^d)}$. One could use the full Leibnitz formula as before, but for this particular estimate we can instead use the simpler paraproduct decomposition

$$\begin{aligned}\langle\nabla\rangle^s(fg) &= \langle\nabla\rangle^s\pi_{hh}(f,g) + \langle\nabla\rangle^s\pi_{lh}(f,g) + \langle\nabla\rangle^s\pi_{hl}(f,g)\\ &= \pi^r_{hh}(\langle\nabla\rangle^s f,g) + \pi'_{lh}(f,\langle\nabla\rangle^s g) + \pi'_{hl}(\langle\nabla\rangle^s f,g)\end{aligned}$$

thanks to Lemma 2.7, Lemma 6.2, and Lemma 6.3. The claim then easily follows by using Corollary 4.4 and Proposition 5.4. ∎

It is also instructive to establish such estimates by direct Littlewood-Paley decomposition, avoiding the Coifman-Meyer paradifferential calculus.

There is of course an extension of this bilinear calculus to trilinear operators, etc. but the theory becomes notationally messy, and in practice one can usually obtain whatever trilinear or multilinear estimates necessary by concatenating bilinear estimates together, or by working things out by hand using Littlewood-Paley multipliers.

Finally, we remark that in the theory of nonlinear dispersive equations (such as nonlinear Schrödinger and wave equations) there has been significant interest in establishing bilinear or trilinear multiplier estimates when the symbol does not obey Coifman-Meyer or residual type bounds, but instead has singularities concentrated near larger dimensional sets such as paraboloids or cones. These estimates are most effective at the L^2 level, in which case they go by the name of $X^{s,b}$ *estimates*. But these are beyond the scope of this course.

7. Fractional chain rule

We now turn from bilinear estimates to nonlinear ones, in particular understanding the relationship between the size of a function $u : \mathbf{R}^d \to \mathbf{C}$ and a composition $F(u) : \mathbf{R}^d \to \mathbf{C}$, where $F : \mathbf{C} \to \mathbf{C}$ is a known function[6]. For various technical reasons we assume $F(0) = 0$. Model examples include the *Lipschitz case*, when $|F(z) - F(w)| \lesssim |z - w|$ for all $z, w \in \mathbf{C}$, and the *power nonlinearity* case, in which $|F(z)| \lesssim_p |z|^p$ and (more generally) $|F(z) - F(w)| \lesssim_p |z - w|(|z|^{p-1} + |w|^{p-1})$ for some $p \geq 1$. (Thus the Lipschitz case corresponds to the $p = 1$ power nonlinearity case.) A very typical example of a power nonlinearity is the function $F(z) = |z|^{p-1}z$. If p is an odd integer, then $F(u)$ is a multilinear combination of u and $\overline{u}$ and can thus (at least in principle) be treated by the theory of the previous section, but we now allow F to be "non-algebraic" and "non-analytic" in the sense that $F(u)$ cannot be expressed in terms of multilinear combinations (or convergent power series) of u and $\overline{u}$.

[6]This is of course not the only nonlinear function one wishes to consider. Another common problem which arises in PDE is to estimate $F(u) - F(v)$ in terms of u, v, and $u - v$. But one can often reduce these more general problems to this basic one, for instance by writing $F(u) - F(v) = \int_0^1 (u-v)\cdot F'((1-t)u+tv)\, dt$, or else expressing $F(u)-F(v) = (u-v)G(u, v)$ where $G : \mathbf{C}^2 \to \mathbf{C}$ is the function $G(z, w) := \frac{F(z)-F(w)}{z-w}$ and applying a vector-valued version of the previous analysis.

In Lebesgue spaces it is clear what the relationship between u and $F(u)$ is. For instance in the Lipschitz case we have $F(u) = O(|u|)$, and hence we have $\|F(u)\|_{L^q(\mathbf{R}^d)} \lesssim \|u\|_{L^q(\mathbf{R}^d)}$ for all $0 < q \leq \infty$. In the power nonlinearity case (with exponent p) we have $F(u) = O_p(|u|^p)$ and hence $\|F(u)\|_{L^{q/p}(\mathbf{R}^d)} \lesssim_p \|u\|^p_{L^q(\mathbf{R}^d)}$ for all $0 < q \leq \infty$.

In Sobolev spaces with exactly one derivative of regularity we can also get the right estimates quickly from the chain rule. For instance, in the Lipschitz case (and for Schwartz u and F, for simplicity) we have

$$\nabla(F(u)) = \nabla u \cdot F'(u) = O(|\nabla u|) \tag{9}$$

and so

$$\|\nabla F(u)\|_{L^q(\mathbf{R}^d)} \lesssim \|\nabla u\|_{L^q(\mathbf{R}^d)}$$

for all $0 < q < \infty$. In particular this implies that

$$\|F(u)\|_{W^{1,q}(\mathbf{R}^d)} \lesssim_{q,d} \|u\|_{W^{1,q}(\mathbf{R}^d)}$$

for $1 < q < \infty$. A similar computation in the power nonlinearity case (which we leave to the reader) gives

$$\|F(u)\|_{W^{1,q/p}(\mathbf{R}^d)} \lesssim_{q,d} \|u\|_{W^{1,q}(\mathbf{R}^d)}$$

for all $p < q < \infty$. Indeed, by use of Sobolev embedding one can even improve the q/p exponent somewhat; we leave this again to the reader.

One might then hope (perhaps by some sort of interpolation) to obtain some intermediate result for Sobolev spaces with regularity s between 0 and 1. (For $s > 1$ we do not expect any estimates unless more regularity is also placed on F; for instance, in order to get two degrees of regularity on $F(u)$ it is reasonable to first demand two degrees of regularity on F.) For instance based on the above it is reasonable to conjecture that in the Lipschitz case we have

$$\|F(u)\|_{W^{s,q}(\mathbf{R}^d)} \lesssim_{q,d,s} \|u\|_{W^{s,q}(\mathbf{R}^d)}$$

for all $0 \leq s \leq 1$. Unfortunately, it is difficult to apply standard interpolation theorems here, because the operator $u \mapsto F(u)$ is not linear, multilinear, or sublinear. Nevertheless, one can still try to follow the *spirit* of the (real) interpolation method, by trying to decompose u into "low regularity" (or "high frequency") pieces with which one uses the $s = 0$ theory, and a "high regularity" (or "low regularity") piece which can be controlled by the $s = 1$ theory. One has to deal with the non-linearity of F, of course, but the plan would be to use formulae such as Taylor's theorem with remainder, e.g.

$$F(v) = F(u) + \int_0^1 (v - u) \cdot F'((1 - t)u + tv)\ dt.$$

(Here, $F'(z)$ is viewed as a real-linear map from $\mathbf{C}$ to $\mathbf{C}$.) There is another, closely related, strategy, which is to try to generalise the chain rule (9) (similarly to how we generalised the Leibnitz rule in previous sections). For integer k, repeated iterations of the chain rule give

$$\nabla^k F(u) = (\nabla^k u) \cdot F'(u) + \dots$$

where the remaining terms require fewer than k derivatives on u (and plenty on F). We can extrapolate from this and guess a *fractional chain rule*

$$D^s F(u) \approx (D^s u) \cdot F'(u).$$

This type of rule turns out to be a little tricky to formalise properly, but it serves as good intuition to guide the results which follow.

A good compromise between the above strategies is not to work on expressions such as $D^s F(u)$ directly, but instead to work with Littlewood-Paley components $\psi_j(D)F(u)$, hoping to reconstruct $D^s F(u)$ later. The analogue of the fractional chain rule here is the heuristic

$$\psi_j(D)F(u) \approx \psi_j(D)u \cdot F'(\psi_{<j}(D)u). \tag{10}$$

An informal motivation of this heuristic is as follows. Splitting $u = \psi_{<j}(D)u + \psi_{\geq j}(D)u$ and using Newton's approximation we roughly have

$$F(u) \approx F(\psi_{<j}(D)u) + \psi_{\geq j}(D)u \cdot F'(\psi_{<j}(D)u).$$

Now we apply $\psi_j(D)$ to both sides. The function $\psi_{<j}(D)u$ is almost annihilated by $\psi_j(D)$, and we expect the same for multilinear combinations of this function; thus the first term $F(\psi_{<j}(D)u)$ should disappear and the second term, being a kind of high-low paraproduct between u and $F'(u)$, should then have the $\psi_j(D)$ term fall solely on the $\psi_{\geq j}(D)u$ factor, giving (10).

One can eventually make the above heuristics precise, but they require more regularity on F than currently assumed (for instance, one needs something like a Hölder continuity estimate on F'). Here, we will present here a cruder estimate which does not demand as high regularity on F and suffices for many purposes.

Lemma 7.1. *Let u be Schwartz and let $1 = \sum_j \psi_j(D)$ be a Littlewood-Paley decomposition. If F is a Lipschitz nonlinearity with $F(0) = 0$ (in order to ensure F decays at infinity), then we have the pointwise estimate*

$$|\psi_j(D)F(u)| \lesssim_d \sum_k \min(2^k, 1) M(\psi_{j+k}(D)u)$$

where M is the Hardy-Littlewood maximal inequality. More generally, if F is a power-type nonlinearity with exponent $p \geq 1$, then

$$|\psi_j(D)F(u)| \lesssim_{p,d} \sum_k \min(2^k, 1)[M(|u|^{p-1})M(\psi_{j+k}(D)u) + M(|u|^{p-1}\psi_{j+k}(D)u)].$$

Proof We just prove the first estimate and leave the second as an exercise. By translation invariance it suffices to estimate this at the origin. By rescaling (noting that the rescaled version $2^j F(2^{-j}z)$ of F is still Lipschitz) we may take $j = 0$, thus we need to show

$$|\psi_0(D)(F(u))(0)| \lesssim_d \sum_k \min(2^k, 1) M(\psi_k(D)u)(0).$$

We express $\psi_0(D)$ in Fourier space to obtain

$$\psi_0(D)(F(u))(0) = \int_{\mathbf{R}^d} \check{\psi}_0(-y)F(u(y))\, dy.$$

We use the Lipschitz condition to write

$$F(u(y)) = F(\psi_{\leq 0}(D)u(y)) + O(|\psi_{>0}(D)u(y)|).$$

Since ψ_0 vanishes at the origin, $\check{\psi}_0$ has mean zero, so we can rewrite this as

$$\psi_0(D)(F(u))(0) = \int_{\mathbf{R}^d} \check{\psi}_0(-y)(F(\psi_{\leq 0}(D)u(y)) - F(\psi_{\leq 0}(D)u(0)) + O(|\psi_{>0}(D)u(y)|))\, dy.$$

Now we take advantage of the rapid decrease of $\check{\psi}_0$ and the triangle inequality to obtain

$$|\psi_0(D)(F(u))(0)| \lesssim_d \int_{\mathbf{R}^d} \langle y\rangle^{-100d}(|F(\psi_{\leq 0}(D)u(y)) - F(\psi_{\leq 0}(D)u(0))| + |\psi_{>0}(D)u(y)|).$$

At this point we use the Lipschitz hypothesis to bound

$$|F(\psi_{\leq 0}(D)u(y)) - F(\psi_{\leq 0}(D)u(0))| \lesssim |\psi_{\leq 0}(D)u(y) - \psi_{\leq 0}(D)u(0)|.$$

Applying a Littlewood-Paley decomposition we then have

$$|\psi_0(D)(F(u))(0)| \lesssim_d \int_{\mathbf{R}^d} \langle y\rangle^{-100d} \sum_{k\leq 0} |\psi_k(D)u(y) - \psi_k(D)u(0)| + \sum_{k>0} |\psi_k(D)u(y)|.$$

The second term can be easily bounded by $\sum_{k>0} M\psi_k(D)u(0)$, which is acceptable. For the $k < 0$ contributions, we use the fundamental theorem of calculus to express

$$\psi_k(D)u(y) - \psi_k(D)u(0) = \int_0^1 y \cdot \nabla\psi_k(D)u(ty)\; dt.$$

By Lemma 3.1 we have

$$\nabla\psi_k(D)u(ty) = O(\langle 2^k y\rangle^d M\psi_k(D)u(0)).$$

We thus obtain a contribution of $\sum_{k\leq 0} 2^k M\psi_k(D)u(0)$, which is acceptable. ■

This estimate is already enough to control $F(u)$ adequately in many situations. We give just one example:

Corollary 7.2. *Let u be Schwartz. If F is a Lipschitz nonlinearity then*

$$\|F(u)\|_{W^{s,q}(\mathbf{R}^d)} \lesssim_{s,q,d} \|u\|_{W^{s,q}(\mathbf{R}^d)}$$

for all $1 < q < \infty$ and $0 < s < 1$. If instead F is a power-type nonlinearity with exponent p, then

$$\|F(u)\|_{W^{s,q}(\mathbf{R}^d)} \lesssim_{s,q,d} \|u\|_{L^r(\mathbf{R}^d)}^{p-1} \|u\|_{W^{s,t}(\mathbf{R}^d)}$$

whenever $1 < q, r, t < \infty$ and $0 < s < 1$ are such that $1/q = (p-1)/r + 1/t$.

Proof Again we just handle the Lipschitz case and leave the power case as an exercise. Since $F(u) = O(|u|)$, we have

$$\|F(u)\|_{L^q(\mathbf{R}^d)} \lesssim \|u\|_{L^q(\mathbf{R}^d)} \lesssim_{s,q,d} \|u\|_{W^{s,q}(\mathbf{R}^d)}$$

so it suffices to show that

$$\|F(u)\|_{\dot W^{s,q}(\mathbf{R}^d)} \lesssim_{s,q,d} \|u\|_{\dot W^{s,q}(\mathbf{R}^d)}.$$

From the Littlewood-Paley characterisation of Sobolev spaces, we have

$$\|F(u)\|_{\dot W^{s,q}(\mathbf{R}^d)} \sim_{s,q,d} \|(\sum_j 2^{2sj} |\psi_j(D)F(u)|^2)^{1/2}\|_{L^q(\mathbf{R}^d)}.$$

Applying Lemma 7.1 and the triangle inequality, we conclude

$$\|F(u)\|_{\dot W^{s,q}(\mathbf{R}^d)} \sim_{s,q,d} \sum_k \min(2^k,1) \|(\sum_j 2^{2sj} |M(\psi_{j+k}(D)u)|^2)^{1/2}\|_{L^q(\mathbf{R}^d)}.$$

Applying Fefferman-Stein we then have

$$\|F(u)\|_{\dot W^{s,q}(\mathbf{R}^d)} \sim_{s,q,d} \sum_k \min(2^k,1) \|(\sum_j 2^{2sj} |\psi_{j+k}(D)u|^2)^{1/2}\|_{L^q(\mathbf{R}^d)}.$$

Shifting j by k and using the Littlewood-Paley inequality we conclude

$$\|F(u)\|_{\dot W^{s,q}(\mathbf{R}^d)} \sim_{s,q,d} \sum_k \min(2^k,1) 2^{-sk} \|u\|_{\dot W^{s,q}(\mathbf{R}^d)}$$

and the claim follows. ∎

8. Exercises

- Q1. Show that Corollary 3.4 fails when $p = q = r = \infty$. (Hint: Consider a high-high paraproduct $\sum_j \psi_j(D) f \psi_j(D) g$ in one dimension applied to $f = g$ equal to the signum function $\operatorname{sgn}(x)$ (or some smoothed out version thereof) and evaluate this at $x = 0$.) By duality obtain similar results when $(p,q) = (1,\infty)$ or $(p,q) = (\infty, 1)$.
- Q2. (Carleson embedding theorem) Let μ be a positive Radon measure on $\mathbf{R}^+ \times \mathbf{R}^d$. Show that the following are equivalent up to changes in the implied constant:
  - (i) We have
  $$\mu([0,r] \times B(x,r)) \lesssim_d |B(x,r)|$$
  for all balls $B(x,r)$.
  - (ii) For any $1 < p < \infty$, we have
  $$\int_{\mathbf{R}^+ \times \mathbf{R}^d} [\fint_{B(x,r)} |f|]^p \, d\mu(r,x) \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}^p.$$

  (Hints: the implication of (i) from (ii) is easy; in fact one only needs (ii) for a single p, such as $p = 2$, in order to deduce (i). For the converse implication, dyadically decompose $\fint_{B(x,r)} |f|$ as in the proof of (4.1).) Measures μ which obey either (i) or (ii) are known as *Carleson measures*.
- Q3. (Moser's inequality) Let $s > 0$ and $1 < p < \infty$. Show that
  $$\|fg\|_{W^{s,p}(\mathbf{R}^d)} \lesssim_{s,p,d} \|f\|_{W^{s,p}(\mathbf{R}^d)} \|g\|_{L^\infty(\mathbf{R}^d)} + \|f\|_{L^\infty(\mathbf{R}^d)} \|g\|_{W^{s,p}(\mathbf{R}^d)}$$
  for all Schwartz f, g. Conclude in in particular that if $s > d/p$, then $W^{s,p}(\mathbf{R}^d)$ is closed under multiplication.
- Q4. (Div-curl lemma) Let $1 < p < \infty$, and let $f : \mathbf{R}^d \to \mathbf{R}^d$ and $g : \mathbf{R}^d \to \mathbf{R}^d$ be vector-valued Schwartz functions such that $\nabla \cdot f = 0$ and $\nabla \wedge g = 0$ (thus f is divergence-free and g is curl-free). Show that for any Hörmander-Mikhlin multiplier $a(D)$, that
  $$\|a(D)(fg)\|_{L^1(\mathbf{R}^d)} \lesssim_{d,p} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^{p'}(\mathbf{R}^d)}.$$
  This is despite $a(D)$ not necessarily being bounded on L^1. (Readers familiar with Hardy spaces will thus be able to conclude that fg lies in $\mathcal{H}^1(\mathbf{R}^d)$.)
- Q5. Complete the proof of Lemma 7.1 for nonlinearities of power type.

- Q6. Complete the proof of Corollary 7.2 for nonlinearities of power type.
- Q7. For $1 \leq p \leq \infty$ and $0 < \alpha < 1$, define the *Hölder space* $\Lambda^\alpha_p(\mathbf{R}^d)$ to be the functions whose norm
$$\|u\|_{\Lambda^\alpha_p(\mathbf{R}^d)} := \|u\|_{L^p(\mathbf{R}^d)} + \sup_{0<|h|\leq 1} \|\mathrm{Trans}_h u - u\|_{L^p(\mathbf{R}^d)}/|h|^\alpha$$
is finite, where $\mathrm{Trans}_h u(x) := u(x-h)$ is the shift of u by h. Thus for instance Λ^α_∞ is the class of Hölder continuous functions of order α.
- (i) Show that
$$\|u\|_{\Lambda^\alpha_p(\mathbf{R}^d)} \sim_{p,\alpha,d} \|u\|_{L^p(\mathbf{R}^d)} + \sup_{j\geq 0} 2^{-j\alpha}\|\psi_j(D)u\|_{L^p(\mathbf{R}^d)}$$
where we use the usual Littlewood-Paley decomposition. Conclude that for $1 < p < \infty$ we have
$$\|u\|_{W^{\alpha-\varepsilon,p}(\mathbf{R}^d)} \lesssim_{p,\alpha,d,\varepsilon} \|u\|_{\Lambda^\alpha_p(\mathbf{R}^d)} \lesssim_{p,\alpha,d,\varepsilon} \|u\|_{W^{\alpha+\varepsilon,p}(\mathbf{R}^d)}.$$
This fact allows us to use Hölder spaces as a "cheap" substitute for Sobolev spaces, provided we are willing to lose an epsilon.
- (ii) Using only elementary estimates (such as Hölder's inequality - no Fourier analysis!) show that
$$\|F(u)\|_{\Lambda^\alpha_q(\mathbf{R}^d)} \lesssim \|u\|_{\Lambda^\alpha_q(\mathbf{R}^d)}$$
for any Lipschitz nonlinearity F and $1 \leq q \leq \infty$, and more generally for a power nonlinearity of order p that
$$\|F(u)\|_{\Lambda^\alpha_q(\mathbf{R}^d)} \lesssim \|u\|^{p-1}_{L^r(\mathbf{R}^d)}\|u\|_{\Lambda^\alpha_t(\mathbf{R}^d)}$$
whenever $1 \leq q, r, t \leq \infty$ is such that $1/q = (p-1)/r + 1/t$.
- Q8. (Localisation of Sobolev spaces) Let $s \geq 0$, $1 \leq p \leq \infty$, and let ϕ be a Schwartz function of height 1 adapted to a ball B. If the radius of B is at least 1, show that
$$\|\phi f\|_{W^{s,p}(\mathbf{R}^d)} \lesssim_{s,p,d} \|f\|_{W^{s,p}(\mathbf{R}^d)}$$
for all $f \in \mathcal{S}(\mathbf{R}^d)$. If we also have $s < d/p$, show that the hypothesis that the radius of B is at least 1 can be omitted.

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

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