LECTURE NOTES 5 FOR 247A

TERENCE TAO

1. Weighted inequalities

In the previous notes we have taken a linear or sublinear operator T on a measure space X and established L^p mapping properties, such as

$$\int_Y |Tf(y)|^p \, d\mu_Y(y) \lesssim \int_X |f(x)|^p \, d\mu_X(x).$$

These estimates reveal information about how the height and width of Tf is controlled by the height and width of f. However, these estimates do not indicate *where* the support of Tf is located, due to the rearrangement-invariant nature of the L^p norms. One way to establish this type of control is to start establishing *weighted* L^p estimates, such as

$$\int_Y |Tf(y)|^p w_2(y) \, d\mu(y) \lesssim \int_X |f(x)|^p w_1(x) \, d\mu_X(x). \tag{1}$$

for various pairs of *weight functions* $w_1 : X \to \mathbf{R}^+$ and $w_2 : Y \to \mathbf{R}^+$. Roughly speaking, these weighted estimates are a variant of the usual L^p estimates which assert that if f avoids the regions where w_1 is large, then Tf avoids the regions where w_2 is large. If $X = Y$ and one can establish the above type of estimate for many pairs w_1, w_2, with w_1 "looking similar to" w_2, then these estimates begin to confirm an assertion that T does not significantly change the support of f.

One can view the estimate (1) as an assertion that w_1 is at least as large as some sort of "nonlinear adjoint" of T applied to w_2. For instance:

Problem 1.1. Let $T = T_K$ be an integral operator with bounded non-negative kernel $K : X \times Y \to \mathbf{R}^+$, and let w_1, w_2 be bounded weights. Show that the inequality

$$\int_Y |Tf(y)| w_2(y) \, d\mu_Y(y) \lesssim \int_X |f(x)| w_1(x) \, d\mu_X(x)$$

holds for all simple functions f with finite measure support if and only if one has the pointwise estimate $w_1 \geq T^* w_2$.

We know that boundedness properties of an adjoint can be used to establish boundedness properties of the original operator. There is a counterpart for weighted estimates:

Proposition 1.2 (Equivalence of weight bounds and vector-valued L^p estimates)**.** *Let $1 \leq p < q < \infty$, let $A > 0$, let $T : L^p(X) \to L^p(Y)$ be an arbitrary operator* *(possibly nonlinear!), and let r be such that $r' = q/p$. Then the following are equivalent:*

(i) *For any sequence $(f_n)_{n \in \mathbf{Z}}$ in $L^p(X)$, we have*

$$\|(\sum_n |Tf_n|^p)^{1/p}\|_{L^q(Y)} \leq A\|(\sum_n |f_n|^p)^{1/p}\|_{L^q(X)}.$$

In particular (setting all but one of the f_n equal to zero),

$$\|Tf\|_{L^q(Y)} \leq A\|f\|_{L^q(X)}.$$

(ii) *For every non-negative $w_2 \in L^r(Y)$ there exists a non-negative $w_1 \in L^r(X)$ with $\|w_1\|_{L^r(X)} \leq \|w_2\|_{L^r(Y)}$ such that*

$$\int_Y |Tf(y)|^p w_2(y)\ d\mu_Y(y) \leq A^p \int_X |f(x)|^p w_1(x)\ d\mu_X(x) \tag{2}$$

for all $f \in L^p(X)$.

Proof By dividing T by A we may normalise $A = 1$. We first show that (ii) implies (i). From the converse Hölder inequality we have

$$\|(\sum_n |Tf_n|^p)^{1/p}\|^p_{L^q(Y)} := \sup\{\int_Y \sum_n |Tf_n(y)|^p w_2(y)\ d\mu_Y(y) : \|w_2\|_{L^r(Y)} = 1, w_2 \geq 0\}.$$

By monotone convergence we may interchange the sum and integral. From (2) and Hölder inequality we have

$$\sum_n \int_Y |Tf_n(y)|^p w_2(y)\ d\mu_Y(y) \leq \sum_n \int_Y |f_n(y)|^p w_1(y)\ d\mu_Y(y) \leq \|(\sum_n |f_n|^p)^{1/p}\|^p_{L^q(X)} \|w_1\|_{L^r(X)}$$

where $\|w_1\|_{L^r(X)} \leq \|w_2\|_{L^r(Y)} = 1$. The claim follows.

Now we establish that (i) implies (ii). Fix w_2; we may normalise $\|w_2\|_{L^r(Y)} = 1$. We need to find a w_1 obeying (2). From the above arguments we know that a necessary condition for w_1 is that

$$\sum_n \int_Y |Tf_n(y)|^p w_2(y)\ d\mu_Y(y) \leq \int_Y |F(y)| w_1(y)\ d\mu_Y(y)$$

for all f_n and F such that $\sum_n |f_n|^p \leq |F|$. Motivated by this, given any $F \in L^{r'}(Y)$, we consider the "capacity"

$$\lambda(F) := \sup\{\sum_n \int_Y |Tf_n(y)|^p w_2(y)\ d\mu_Y(y) : \sum_n |f_n|^p \leq |F|\},$$

where $(f_n)_{n \in \mathbf{Z}}$ ranges over all sequences $L^p(X)$ with $\sum_n |f_n|^p$ bounded pointwise by F. It will suffice to locate $w_1 \in L^r(X)$ with $\|w_1\|_{L^r(X)} \leq 1$ such that

$$\lambda(F) \leq \int_X |F| w_1\ d\mu_X$$

for all $F \in L^{r'}(X)$, since we may then substitute $F = |f|^p$ and observe that $\lambda(|f|^p) \geq \int_Y |f|^p w_2\ d\mu_Y$.

Observe from Hölder that

$$\sum_n \int_Y |Tf_n(y)|^p w_2(y)\ d\mu_Y(y) \leq \|(\sum_n |Tf_n|^p)^{1/p}\|_{L^q(Y)}^p \|w_2\|_{L^r(Y)}$$
$$\leq \|(\sum_n |f_n|^p)^{1/p}\|_{L^q(X)}^p$$
$$\leq \|F\|_{L^{r'}(X)}$$

and so we have the boundedness

$$\lambda(F) \leq \|F\|_{L^{r'}(X)}.$$

We also easily verify the superlinearity properties

$$\lambda(cF) = |c|\lambda(F); \quad \lambda(F+G) \geq \lambda(F) + \lambda(G)$$

for all non-negative $F, G \in L^{r'}(X)$. Observe that the sets

$$A := \{F \in L^{r'}(X) : \|F\|_{L^{r'}(X)} < 1\}$$

and

$$B := \{F \in L^{r'}(X) : F > 0; \lambda(F) > 1\}$$

are algebraically open (see appendix), convex, and disjoint. Thus by the geometric Hahn-Banach theorem[1] (see appendix), we can find a linear functional $\nu : L^{r'}(X) \to \mathbf{C}$ such that $\nu < 1$ on A and $\nu > 1$ on B. The former implies that ν is a continuous linear functional with norm at most 1 and thus $\nu(F) = \int_X F w_1\ d\mu_X$ for some $w_1 \in L^r(X)$ with $\|w_1\|_{L^r(X)} \leq 1$. The latter implies that $\nu(F) \geq \lambda(F)$ for all non-negative F, and in particular (setting $F := |f|^p$) that

$$\int_Y |Tf|^p w_2\ d\mu_Y \leq \lambda(|f|^p) \leq \nu(|f|^p) = \int_X |f|^p w_1\ d\mu_X$$

as desired. ■

Remark 1.3. By combining this with Q15 from last week's notes, we obtain a remarkable fact: if $T : L^2(X) \to L^2(Y)$ is a linear operator, then there is a unique continuous linear extension from $L^q(X)$ to $L^q(Y)$ for $q > 2$ if and only if for every $w_2 \in L^r(Y)$ there exists $w_1 \in L^r(X)$ with $\|w_1\|_{L^r(X)} \lesssim_T \|w_2\|_{L^r(Y)}$ for which one has the weighted estimate (1), where $r := (q/2)'$. Thus in the case of linear operators there is an equivalence between higher unweighted L^p estimates and weighted L^2 estimates.

We thus see that weighted estimates can be quite powerful; they not only imply unweighted estimates, but also vector-valued versions of those estimates. Indeed, more is true: if we have a sequence of operators T_n which obey weighted estimates with the *same* set of weights, then we automatically obtain vector-valued estimates

[1]This is a typical use of the Hahn-Banach theorem: philosophically, this theorem asserts that in linear or convex programming, any possibility which is not obviously excluded by the constraints is in fact permissible. Here, it is the vector-valued inequality in (i) which shows that all the "obvious" obstructions to the estimate (ii) do not exist, and Hahn-Banach tells one that there are no non-obvious obstructions.

which involve all the T_n at once. For instance, if p, q, r are as in the above proposition, and for each $w_2 \in L^r(Y)$ there was a $w_1 \in L^r(X)$ (*independent* of n) with $\|w_1\|_{L^r(X)} \lesssim \|w_2\|_{L^r(Y)}$ such that

$$\int_Y |T_n f(y)|^p w_2(y)\ d\mu_Y(y) \leq \int_X |f(x)|^p w_1(x)\ d\mu_X(x)$$

for all n, then we can conclude the vector-valued estimate

$$\|(\sum_n |T_n f_n|^p)^{1/p}\|_{L^q(Y)} \leq A\|(\sum_n |f_n|^p)^{1/p}\|_{L^q(X)}.$$

It is thus of interest to obtain weighted estimates, preferably those in which the final weight w_2 is given explicitly in terms of w_1. A particularly pleasant situation occurs when $X = Y$ and the weights w_1, w_2 are in fact *equal*, although having estimates with asymmetric weights can still be very useful.

One particularly common (and useful) class of weights are *power weights* such as $w(x) = \langle x\rangle^\alpha$ or $w(x) = |x|^\alpha$ for Euclidean spaces $\mathbf{R}^d$, especially in PDE applications where such weights tend to appear quite naturally (for instance, via the introduction of Sobolev spaces). However, we will work here with more abstract weights. Here it turns out that one can often mimic the proof of an unweighted estimate to deduce the weighted counterpart.

1.4. Weighted estimates for the Hardy-Littlewood maximal function.

To illustrate weighted estimates, let us establish a basic estimate for the maximal function.

Proposition 1.5 (Weighted Hardy-Littlewood maximal inequality)**.** *Let $w : \mathbf{R}^d \to \mathbf{R}^+$. Then we have the weak-type estimate*

$$\int_{\mathbf{R}^d} 1_{|Mf(x)|>\lambda}\ w(x)\ dx \lesssim_d \frac{1}{\lambda}\int_{\mathbf{R}^d} |f(x)| Mw(x)\ dx$$

for all locally integrable f and all $\lambda > 0$, and the L^p estimate

$$\int_{\mathbf{R}^d} |Mf(x)|^p\ w(x)\ dx \lesssim_{d,p} \int_{\mathbf{R}^d} |f(x)|^p Mw(x)\ dx$$

for all locally integrable f and $1 < p < \infty$.

Proof By monotone convergence we may assume that both w and f are bounded with compact support. The Stein-Weiss interpolation theorem (and the boundedness of M on L^∞ no matter what the weights are) ensure that it suffices to prove the weak-type $(1, 1)$ estimate. For this we mimic the proof of the unweighted estimate. We normalise $\lambda = 1$, and let K be any compact set inside $\{|Mf(x)| > 1\}$; it suffices to show that

$$\int_K w(x)\ dx \lesssim_d \int_{\mathbf{R}^d} |f(x)| Mw(x)\ dx.$$

The usual Vitali covering lemma argument lets us locate a finite number of disjoint open balls $B_1, \ldots, B_N$ such that $\bigcup_{n=1}^N 3B_n$ covers K, and such that $⨍_{B_n} |f(x)| \gtrsim_d 1$

for each ball B_n. Since

$$\int_K w(x)\ dx \leq \sum_{n=1}^N \int_{3B_n} w(x)\ dx$$

it thus suffices to show that

$$\int_{3B_n} w(x)\ dx \lesssim_d \int_{B_n} |f(x)| Mw(x)\ dx$$

for each ball B_n. But we observe that

$$Mw(x) \gtrsim_d \frac{1}{|B_n|} \int_{3B_n} w(y)\ dy$$

for all $x \in B_n$, and the claim follows. ∎

As a corollary we obtain the following useful vector-valued generalisation of the Hardy-Littlewood maximal inequality.

Theorem 1.6 (Fefferman-Stein vector-valued maximal inequality)**.** *If $1 < p, q < \infty$ and $f_n : \mathbf{R}^d \to \mathbf{C}$ are any sequence of locally integrable functions then*

$$\|(\sum_n |Mf_n|^p)^{1/p}\|_{L^q(\mathbf{R}^d)} \lesssim_{d,p,q} \|(\sum_n |f_n|^p)^{1/p}\|_{L^q(\mathbf{R}^d)}$$

and the weak-type estimate

$$|\{(\sum_n |Mf_n|^p)^{1/p} \geq \lambda\}| \lesssim_{d,p} \frac{1}{\lambda}\|(\sum_n |f_n|^p)^{1/p}\|_{L^1(\mathbf{R}^d)}$$

for all locally integrable f_n.

Proof By the usual monotone convergence arguments we may restrict only finitely many of the f_n to be non-zero, and to have all f_n bounded and finite measure support. We can then restrict the supremum in the maximal function M to only be over finitely many radii.

When $p = q$ the first claim follows from interchanging the norms and using the usual Hardy-Littlewood maximal inequality. For $q > p$ the claim follows by combining Proposition 1.5, Proposition 1.2, and the fact that M is bounded on $L^r(\mathbf{R}^d)$. For $q < p$, it suffices by vector-valued real interpolation to prove the weak-type $(1, 1)$ estimate. We may take each f_n to be non-negative. By linearisation we can replace $Mf_n(x)$ by $⨍_{B(x,r_n(x))} f_n$ for some measurable radius function $r_n(x)$ (note that we must permit r_n to depend on n), which is one reason why this weak-type estimate does not immediately follow from the Hardy-Littlewood inequality). Actually it is more convenient to replace this "rough" average by a smoother average

$$\frac{1}{|B(x, r_n(x))|} \int_{\mathbf{R}^d} \psi_{n,x}(y) f_n(y)\ dy$$

where $\psi_{n,x}$ is a bump function adapted to $B(x, 2r_n(x))$ which equals one on $B(x, r_n(x))$. This reduces matters to establishing a vector-valued weak-type $(1, 1)$ operator for the linear operator

$$T : (f_n)_{n\in\mathbf{Z}}(x) \mapsto (\int_{\mathbf{R}^d} \frac{1}{|B(x, r_n(x))|} \psi_{n,x}(y) f_n(y))_{n\in\mathbf{Z}}.$$

This is an integral operator whose kernel $K(x,y)$ is a diagonal matrix whose n^{th} diagonal entry is $\frac{1}{|B(x,r_n(x))|}\psi_{n,x}(y)$. We observe that this kernel is a *one-sided* vector-valued singular kernel; it obeys the required regularity in the y variable but not the x. However, this is still enough to convert the L^p boundedness of T (which as observed before, follows from switching norms and using the usual Hardy-Littlewood maximal inequality) and then using the standard Calderón-Zygmund argument. ■

The theory of weighted estimates (and the A_p weight class) has been further developed intensively, but we will not cover it here; see Stein's "Harmonic analysis" for a thorough exposition of the topic.

2. Pseudodifferential operators

In the previous set of notes we established constructed a reasonably large class of CZOs, namely the Hörmander-Mikhlin multipliers. These multipliers suffice for the analysis of *translation-invariant* settings, such as studying operators arising from constant-coefficient differential operators such as the Laplacian. However they are not general enough to handle non-translation-invariant situations. For this, we must turn to the *pseudodifferential operators*, which generalise the variable-coefficient differential operators $\sum_\alpha c_\alpha(x)\partial_x^\alpha$ when the coefficients $c_\alpha(x)$ are smooth.

To motivate these operators, we begin with the Fourier inversion formula

$$f(x) = \int_{\mathbf{R}^d} e^{2\pi i x\cdot\xi}\hat f(\xi)\ d\xi$$

(for f Schwartz, say) and apply a monomial differential operator $\partial_x^\alpha = \partial_{x_1}^{\alpha_1}\dots\partial_{x_d}^{\alpha_d}$ for some multiindex $\alpha = (\alpha_1,\dots,\alpha_d)$ to obtain

$$\partial_x^\alpha f(x) = \int_{\mathbf{R}^d} (2\pi i x)^\alpha e^{2\pi i x\cdot\xi}\hat f(\xi)\ d\xi$$

where $x^\alpha := x_1^{\alpha_1}\dots x_d^{\alpha_d}$. Thus we see that for any (smooth) coefficients $c_\alpha(x)$, with only finitely many c_α non-zero, we have

$$\sum_\alpha c_\alpha(x)\partial_x^\alpha f(x) = \int_{\mathbf{R}^d} a(x,\xi)e^{2\pi i x\cdot\xi}\hat f(\xi)\ d\xi$$

where the *symbol* $a(x,\xi)$ is given by

$$a(x,\xi) := \sum_\alpha c_\alpha(x)(2\pi i\xi)^\alpha.$$

Inspired by this, we define for any smooth symbol $a : \mathbf{R}^d\times\mathbf{R}^d \to \mathbf{C}$ of at most polynomial growth, the operator $a(X,D)$ (sometimes also called $\mathrm{Op}(a)$) on Schwartz functions $f\in\mathcal{S}(\mathbf{R}^d)$ by the formula

$$a(X,D)f(x) := \int_{\mathbf{R}^d} a(x,\xi)e^{2\pi i x\cdot\xi}\hat f(\xi)\ d\xi.$$

This operator $a(X,D)$ is sometimes called the *Kohn-Nirenberg quantization* of the symbol[2] $a(x,\xi)$. By expanding out the Fourier transform and interchanging integrals, we (formally) have

$$a(X,D)f(x) = \int_{\mathbf{R}^d} [\int_{\mathbf{R}^d} a(x,\xi) e^{2\pi i (x-y)\cdot \xi}\ d\xi] f(y)\ dy; \tag{3}$$

this formula is rigorous for instance if a is not only smooth but also compactly supported. As such we see that there is a slight asymmetry between x and y in this formula; because of this one sometimes uses the *Weyl quantization*

$$a^{\mathrm{w}}(X,D)f(x) := \int_{\mathbf{R}^d} [\int_{\mathbf{R}^d} a(\frac{x+y}{2},\xi) e^{2\pi i (x-y)\cdot \xi}\ d\xi] f(y)\ dy$$

instead, though we will not use this quantization here.

Remark 2.1. Formally, if we apply $a(X,D)$ to a plane wave $f(x) := e^{2\pi i x\cdot\xi}$, we see that

$$a(X,D)f(x) = a(x,\xi)f(x).$$

Thus we can (in principle) reconstruct the symbol a from the operator $a(X,D)$ by testing it against plane waves. This computation also shows that, *formally*, every linear operator is of the form $a(X,D)$ for some a; but in practice most operators will give horrible symbols a which obey no useful estimates.

The operators $a(X,D)$ generalise both spatial multipliers $a(X) : f(x) \mapsto a(x)f(x)$ and Fourier multipliers $a(D)$, at least when the symbol is smooth. The map from symbol a to operator $a(X,D)$ is linear. However, in contrast to the homomorphism laws

$$a(X)b(X) = (ab)(X); \quad a(D)b(D) = (ab)(D)$$

for smooth multipliers a, b of space or frequency only, it is not quite true that the quantization operation preserves products:

$$a(X,D)b(X,D) \neq (ab)(X,D).$$

Indeed, this would imply in particular that any two operators $a(X,D)$ and $b(X,D)$ commute, which is untrue (indeed, X and D themselves do not commute with each other). Roughly speaking, the Kohn-Nirenberg quantization always arranges the X factors to the "left" of the D factors, whereas the Weyl quantization spreads them around evenly. One can obtain partial substitutes of the homomorphism law; for instance, it is still true that $a(X,D)b(X,D) = (ab)(X,D)$ if $a(x,\xi)$ does not depend on ξ, or $b(x,\xi)$ does not depend on x. Later on we will see that for suitably smooth a and b, that $a(X,D)b(X,D)$ is equal to $(ab)(X,D)$ modulo "lower order" terms. These types of observations form the foundation for the *pseudodifferential symbol calculus*, which is a very useful tool for analysing variable-coefficient linear differential operators (especially those with smooth elliptic symbols), although we will only touch on the foundations of that vast subject here.

[2]In the theory of semiclassical analysis, it is often customary to introduce a small parameter $\hbar$, called *Planck's constant*, and replace $a(x,\xi)$ in the above equation by $a(x,\xi/\hbar)$, but we will not do so here.

There are two main aspects to the theory of pseudodifferential operators: the analytic aspects, in which the boundedness properties of operators $a(X,D)$ are analysed, and the (mainly) algebraic aspects, in which one learns how to compose, commute, or invert operators $a(X,D)$ (usually modulo lower order terms). We shall begin with the analytic aspects.

2.2. **Symbol classes.** Of course, to get any sort of usable bound on the operators $a(X,D)$, we shall need some quantitative control on the symbol $a(x,\xi)$. This requires placing a in one of the *symbol classes*. There are quite a variety of symbol classes to use, which determine the growth and regularity of the symbol a in both the spatial variable x and the frequency variable ξ. For simplicitly, we shall stick here with the *standard*[3] symbol class $S^k = S^k_{0,1}$, which suffices for the local theory of elliptic variable coefficient operators with smooth coefficients[4].

Definition 2.3 (Standard pseudodifferential operators)**.** A smooth symbol $a : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{C}$ is a *(standard) symbol of order* k for some $k \in \mathbf{R}$ if we have the estimates

$$|\partial_x^\alpha \partial_\xi^\beta a(x,\xi)| \lesssim_{\alpha,\beta,d,k} \langle \xi \rangle^{k-|\beta|} \tag{4}$$

for all multi-indices α, β and $x, \xi \in \mathbf{R}^d$ (in particular, $a = O(\langle \xi \rangle^k)$ has at most polynomial growth), and we write $a \in S^k = S^k_{0,1}$ in this case. We refer to the operator $a(X,D)$ as a *(standard) pseudo-differential operator* (or ΨDO for short) of order k.

Remark 2.4*.* Again, this is a quasi-definition rather than a precise definition, due to the presence of implicit constants; in any conclusion involving these symbols, the implied constants in the conclusion will almost certainly depend on the constants in this definition. In practice, one does not need (4) for *all* multi-indices α, β, but only for finitely many (e.g. for all α, β with $|\alpha|, |\beta| \leq 100d$), but for notational reasons we shall assume infinite regularity instead of finite. The indices $0, 1$ in $S^k_{0,1}$ refer to the fact that each derivative in the x variable reduces the growth in x by 0 orders, whereas each derivative in the ξ variable reduces the growth in ξ by 1 order.

Examples 2.5*.* If $L := \sum_{|\alpha| \leq k} c_\alpha(x) \partial_x^\alpha$ is a variable-coefficient differential operator with the coefficients c_α smooth, and obeying the bounds $\partial_x^\beta c_\alpha(x) = O_\beta(1)$ for all multi-indices β, then L is a pseudodifferential operator of order k. For any fixed $s \in \mathbf{C}$, a fractional differentiation operator $\langle D \rangle^s$ is a pseudodifferential operator of order $\mathrm{Re}(s)$. A Littlewood-Paley operator $\psi_j(D)$ is a pseudodifferential operator of order 0 for any $j \geq 0$ (but not as $j \to -\infty$ - why?), and if $j = O(1)$ then $\psi_j(D)$ is a pseudodifferential operator of any fixed order.

Remark 2.6*.* By copious application of Lemma 5.2 one can show that pseudodifferential operators of any order map Schwartz functions to Schwartz functions.

[3]These are also called the *classical* symbol classes, but this is confusing, given that the distinction between classical and quantum mechanics in this subject.

[4]Each symbol class represents a decomposition of phase space into regions respecting the Heisenberg uncertainty principle, with the symbol required to be smoothly adapted to each such region and have a certain upper bound on amplitude. In this case, the relevant regions of phase space are the regions where the spatial variable x is localised to a ball of radius $O(1)$, and the frequency variable ξ is localised to a dyadic annulus $|\xi| \sim 2^k$.

Remark 2.7 (Classical multiplication of symbols). Numerous applications of the Leibnitz rule reveal that if a, b are symbols of order k, l respectively, then ab is a symbol of order $k + l$. Also, observe that if a is a symbol of order k, it is automatically a symbol of order k' for any $k' \geq k$. Also, $\partial_x^\alpha \partial_\xi^\beta a$ will be a symbol of order $k - |\beta|$.

The basic estimate here is

Theorem 2.8 (Calderón-Vaillancourt theorem). *Every pseudodifferential operator of order* 0 *is a CZO. In particular, it is bounded on* $L^p(\mathbf{R}^d)$ *with* $1 < p < \infty$ *with norm* $O_{p,d}(1)$.

Remark 2.9. The theorem of Calderón and Vaillancourt is in fact more general, allowing other types of symbol classes than that presented here, but we will not attempt to describe the strongest version of the theorem in these notes.

Proof We shall allow all implied constants to depend on d, and henceforth omit this dependence explicitly from our notation.

Let $a(x, \xi)$ be a symbol of order 0. We have to check two different things: firstly, that $a(X, D)$ is bounded on $L^2(\mathbf{R}^d)$ with operator norm $O(1)$, and secondly that $a(X, D)$ has a singular kernel.

Observe that if we smoothly truncate $a(x, \xi)$ to be compactly supported in ξ, by multiplying by some cutoff function $\chi(|\xi|/R)$ for some large $R \gg 1$, that a remains a symbol of order 0 (this is a special case of Remark 2.7). Because of this, and because of various limiting arguments (restricting all functions to the Schwartz class, of course), we will be able to restrict to the case where a is compactly supported in ξ, as long as our bounds do not depend on the size of this support.

Let us check the singular kernel property first. If f, g are test functions with disjoint supports, we observe from Fubini's theorem (and the compact support of a in ξ) that

$$\int_{\mathbf{R}^d} a(X, D)f(x)\overline{g(x)}\ dx = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} K(x, y)\overline{g(x)}f(y)\ dxdy$$

where

$$K(x, y) := \int_{\mathbf{R}^d} a(x, \xi)e^{2\pi i(x-y)\cdot\xi}\ d\xi.$$

Now notice that for any fixed x, $a(x, \xi)$ is a symbol of order 0, so by Lemma 5.3 we have

$$|K(x, y)| \lesssim |x - y|^{-d} \text{ and } |\nabla_y K(x, y)| \lesssim_d |x - y|^{-d-1}.$$

However we have to do a little bit more work to get the x derivative, due to the dependence of a on x. Fortunately, a is so smooth in x that this is not an issue. Observe that

$$(\nabla_x + \nabla_y)K(x, y) = \int_{\mathbf{R}^d} (\nabla_x a)(x, \xi)e^{2\pi i(x-y)\cdot\xi}\ d\xi.$$

Now observe from the symbol estimates that for fixed x, $\nabla_x a$ is still a symbol of order 0, and so we have

$$|(\nabla_x + \nabla_y) K(x,y)| \lesssim_N |x-y|^{-d} \langle x-y \rangle^{-N}$$

for any $N \geq 0$, and the claim follows.

Now we check the $L^2(\mathbf{R}^d)$ boundedness property, which is harder. We first take advantage of some additional spatial decay to localise in space. An inspection of Lemma 5.3 shows that we have the rapid decay

$$|K(x,y)| \lesssim |x-y|^{-100d}$$

(say). To use this, we cover $\mathbf{R}^d$ by finitely overlapping balls B of unit radius, and write

$$\|a(X,D)f\|_{L^2(\mathbf{R}^d)} \lesssim (\sum_B \|1_B a(X,D) f\|_{L^2(\mathbf{R}^d)}^2)^{1/2} \lesssim (\sum_B \|1_B a(X,D)(f 1_{2B})\|_{L^2(\mathbf{R}^d)}^2)^{1/2} + (\sum_B \|1_B a(X,D)(f 1_{2B^c}$$

The kernel bound gives the pointwise estimate

$$1_B a(X,D)(f 1_{2B^c}) \lesssim f * \langle x \rangle^{-100d}$$

and so the second term is controlled by $\|f * \langle x \rangle^{-100d}\|_{L^2(\mathbf{R}^d)}$, which is acceptable by Young's inequality. Thus it will suffice to show that

$$(\sum_B \|1_B a(X,D)(f 1_{2B})\|_{L^2(\mathbf{R}^d)}^2)^{1/2} \lesssim \|f\|_{L^2(\mathbf{R}^d)}.$$

But since

$$(\sum_B \|f 1_{2B}\|_{L^2(\mathbf{R}^d)}^2)^{1/2} \sim \|f\|_{L^2(\mathbf{R}^d)}$$

we see that it will suffice to show that each $1_B a(X,D)$ is bounded on $L^2(\mathbf{R}^d)$ uniformly in B. We may translate B to be $B(0,1)$; by applying a smooth cutoff we may then assume that the symbol $a(x,\xi)$ is supported on the region $\{|x| \leq 2\}$. The B localisation has served its purpose of localising a, and we now discard it.

By duality, it suffices to show that

$$|\langle a(X,D)f, g \rangle| \lesssim \|f\|_{L^2(\mathbf{R}^d)} \|g\|_{L^2(\mathbf{R}^d)}$$

for any Schwartz f. Here we shall use the Littlewood-Paley decomposition $1 = \sum_j \psi_j(D)$. The intuition is that $a(X,D)$, as it resembles a Fourier multiplier, "almost commutes" with these projections $\psi_j(D)$, and in particular $\psi_j(D) a(X,D) \psi_k(D)$ is expected to be very small when j and k are very different. This type of Fourier localisation phenomenon can be exploited by Littlewood-Paley decomposition. Specifically, we decompose

$$a(X,D) = \sum_{j,k \geq 0} \tilde\psi_j(D) \psi_j(D) a(X,D) \psi_k(D) \tilde\psi_k(D)$$

where $\tilde\psi_j$ is a slightly wider variant of ψ_j such that $\tilde\psi_j \psi_j = \psi_j$, where we redefine ψ_0 and $\tilde\psi_0$ to be one on the ball $|\xi| \lesssim 1$ rather than the annulus $|\xi| \sim 1$. Thus by the triangle inequality and Cauchy-Schwarz

$$|\langle a(X,D)f, g \rangle| \lesssim \sum_{j,k \geq 0} \|\psi_j(D) a(X,D) \psi_k(D)\|_{L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)} \|\tilde\psi_k(D) f\|_{L^2(\mathbf{R}^d)} \|\tilde\psi_j(D)^* g\|_{L^2(\mathbf{R}^d)}.$$

But from the Littlewood-Paley inequality (or Plancherel) we have

$$(\sum_k \|\tilde\psi_k(D) f\|_{L^2(\mathbf{R}^d)}^2)^{1/2} \sim \|f\|_{L^2(\mathbf{R}^d)}$$

and

$$(\sum_j \|\tilde\psi_j(D)^* g\|_{L^2(\mathbf{R}^d)}^2)^{1/2} \sim \|g\|_{L^2(\mathbf{R}^d)}$$

and so from Schur's test it will suffice to show that

$$\|\psi_j(D) a(X,D) \psi_k(D)\|_{L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)} \lesssim 2^{-\varepsilon|j-k|}$$

for some $\varepsilon > 0$ and all j, k (in fact the argument below allows us to take any $\varepsilon > 0$).

To establish this, let us first look at the operator norm of $a(X,D)\psi_k(D)$. This is an integral operator with kernel

$$K_k(x,y) := \int_{\mathbf{R}^d} a(x,\xi)\psi_k(\xi) e^{2\pi i (x-y)\cdot\xi}\ d\xi.$$

For any fixed x, $a(x,\xi)\psi_k(\xi)$ is a bump function adapted to the annulus $\{|\xi| \sim 2^k\}$. Lemma 5.2 then gives

$$|K_k(x,y)| \lesssim 2^{dk} \langle 2^k (x-y)\rangle^{-d-1}$$

(say). But then Schur's test can be applied to conclude

$$\|a(X,D)\psi_k(D)\|_{L^2(\mathbf{R}^d)} \lesssim 1.$$

This gives the claim when $j - k = O(1)$. It remains to establish the cases when $j > k + 10$ (say) and when $j < k - 10$. The cases are not completely symmetric because our choice of quantization was not symmetric, but it turns out that both cases can be treated in essentially the same manner. Let us first look at the case $j > k + 10$. A little computation shows that $a(X,D)\psi_k(D)$ is an integral operator with kernel

$$K_{j,k}(x,y) = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} \psi_j(\eta) a(z,\xi)\psi_k(\xi) e^{2\pi i (z-y)\cdot\xi} e^{2\pi i (x-z)\cdot\eta}\ d\xi dz d\eta.$$

One can view this kernel in two ways. Firstly, it is $\psi_j(D_x)$ applied to $K_k(x,y)$ (keeping y fixed). Using the prior bounds on $K_k(x,y)$ and the convolution kernel bounds on $\psi_j(D_x)$ we establish that

$$|K_{j,k}(x,y)| \lesssim 2^{dk} \langle 2^k (x-y)\rangle^{-d-1}.$$

But this is not enough by itself (it doesn't get the $2^{-\varepsilon|j-k|}$ decay). We can get a different bound by using one of the basic heuristics of stationary phase, which is to locate the variable in which the phase is oscillating in order to fully exploit integration by parts. In this case, the correct variable to analyse is z. We take absolute values in the ξ, η integrations to obtain

$$|K_{j,k}(x,y)| \leq \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} \left|\int_{\mathbf{R}^d} e^{2\pi i z\cdot(\xi-\eta)} a(z,\xi)\ dz\right| |\psi_j(\eta)||\psi_k(\xi)|\ d\xi d\eta.$$

The support of ψ_j and ψ_k forces $|\xi - \eta| \sim 2^j$, and then since for fixed ξ, $a(z,\xi)$ is a bump function adapted to $B(0,2)$, we see from Lemma 5.2 that

$$\int_{\mathbf{R}^d} e^{2\pi i z\cdot(\xi-\eta)} a(z,\xi)\ dz = O(2^{-100dj})$$

(say). We conclude that
$$|K_{j,k}(x,y)| \lesssim 2^{-98dj}.$$
Taking a suitable combination of this with the previous kernel bound and then applying Schur's test one obtains the desired result (in fact one gets a much better decay by this type of argument, namely $O_N(2^{-Nj})$ for any $N > 0$.

The case $j < k - 10$ is very similar and is left to the reader. This proves the theorem. ∎

Remark 2.10. It is possible to deduce a weak version of the Hörmander-Mikhlin theorem (in which the j^{th} derivatives of the symbol $m(\xi)$ are assumed to be $O_j(|\xi|^{-j})$ for all $j \geq 0$, not just $0 \leq j \leq d+2$) as a consequence of the Calderón-Vaillancourt theorem and a rescaling argument. Firstly, observe that if $m(\xi)$ was a symbol of order 0, then $m(D)$ is already a pseudodifferential operator of order 0 and the claim is immediate. If instead $m(\xi)$ is merely a homogeneous symbol of order 0, we first truncate m smoothly in an ε-neighbourhood of the origin for some small $\varepsilon > 0$ to create a truncated symbol $m_\varepsilon(\xi)$. Then one checks that $m_\varepsilon(\varepsilon\xi)$ is a symbol of order 0 and hence $m_\varepsilon(\varepsilon D)$ is a CZO. Since the class of CZOs is scale-invariant we conclude that $m_\varepsilon(D)$ is a CZO uniformly in ε. The claim then follows by a limiting argument taking $\varepsilon \to 0$.

2.11. **The pseudodifferential calculus.** Now we start performing algebraic manipulations on these pseudodifferential operators, collectively referred to as the *pseudodifferential calculus.* We shall first need a technical lemma.

Lemma 2.12 (Oscillatory integral estimate)**.** *Let $a(x,y,\xi,\eta)$ be a compactly supported function obeying the estimates*
$$|\partial_x^\alpha \partial_y^\beta \partial_\xi^\gamma \partial_\eta^\delta a(x,y,\xi,\eta)| \lesssim_{\alpha,\beta,\gamma,\delta,k,k',d} \langle\xi\rangle^{k-|\gamma|}\langle\eta\rangle^{k'-|\delta|}$$
for some $k, k' \in \mathbf{R}$, all $x,y,\xi,\eta \in \mathbf{R}^d$, and all multiindices $\alpha,\beta,\gamma,\delta$. Then the function
$$c(x,\xi) := \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} a(x,y,\xi,\eta)e^{2\pi i(x-y)\cdot(\xi-\eta)}\ dyd\eta$$
is a symbol of order $k+k'$ (with implied constants depending on k,k',d).

Proof We begin with some reductions. The main task will be to establish the zeroth order bound
$$|c(x,\xi)| \lesssim_{k,k',d} \langle\xi\rangle^{k+k'}.$$
Once one has this, the higher derivatives in x can be dealt with by moving the derivative inside the integral sign, noting that the x derivatives of $e^{2\pi i(x-y)\cdot(\xi-\eta)}$ are negative the y derivative, and then integrating by parts to move all x,y derivatives onto the symbol a where they can be harmlessly absorbed. A similar argument lets one deal with higher derivatives in ξ.

By dividing a by $\langle\xi\rangle^k$ we may take $k=0$. By translating a by x we may take $x=0$. Our task is now to show that
$$|\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} b(y,\eta)e^{2\pi iy\cdot(\eta-\xi)}\ dyd\eta| \lesssim_{k',d} \langle\xi\rangle^{k'} \quad (5)$$

where $b(y,\eta) := a(0,y,\xi,\eta)$; note that

$$|\partial_y^\beta \partial_\eta^\delta b(y,\eta)| \lesssim_{\beta,\delta,k',d} \langle \eta \rangle^{k'-|\delta|}$$

for all $y, \eta \in \mathbf{R}^d$ and all multiindices β, δ. Henceforth we omit the dependence of constants on k', d.

Let us first deal with the case when b is supported on the region $|y| \lesssim 1$. For fixed η, we may apply Lemma 5.2 to estimate the y integral by $O_N(\langle \eta \rangle^{k'} \langle \xi - \eta \rangle^{-N})$ for any N, which by the triangle inequality is equal to $O_N(\langle \xi \rangle^{k'} \langle \xi - \eta \rangle^{-N})$. Integrating in η we obtain the claim.

By a smooth partition of unity it remains to deal with the case when b is supported on the region $|y| \gg 1$. Here what we do is take a large integer N, write

$$e^{2\pi i y \cdot (\eta - \xi)} = (\frac{y}{2\pi i |y|^2} \cdot \nabla_\eta)^N e^{2\pi i y \cdot (\eta - \xi)}$$

and integrate by parts N times to express the left-hand side of (5) as

$$|\int_{\mathbf{R}^d} \int_{\mathbf{R}^d} [(\nabla_\eta \cdot \frac{y}{2\pi |y|^2})^N b(y,\eta)] e^{2\pi i y \cdot (\eta - \xi)}\ dy d\eta|.$$

For fixed η, one can estimate the y integral using Lemma 5.2 (and a smooth partition of unity) as $O_N(\langle \eta \rangle^{k'} \langle \xi - \eta \rangle^{-N})$, and by repeating the previous argument we obtain the desired bound. ■

We now use this lemma to establish that the (formal) adjoint of a pseudodifferential operator is also a pseudodifferential operator.

Lemma 2.13 (Adjoints of ΨDOs)**.** *Let $a(X,D)$ be a pseudodifferential operator of order k for some $k \in \mathbf{R}$. Then $a(X,D)$ has an adjoint $a^*(X,D)$ which is also a pseudodifferential operator of order k. In fact we have*

$$a^*(x,\xi) = \overline{a(x,\xi)} \mod S^{k-1}$$

i.e. a^* differs from $\overline{a}$ by a symbol of order $k-1$.

Proof We assume that a is compactly supported in frequency; this assumption can be removed by limiting arguments which we leave as an exercise to the reader (the point being that our estimates are uniform in the size of this support). Then the adjoint of $a(X,D)$ is given by

$$a(X,D)^* f(x) = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \overline{a(y,\xi)} e^{2\pi i (x-y)\cdot \xi} f(y)\ d\xi dy.$$

On the other hand, observe that

$$\overline{a}(X,D) f(x) = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \overline{a(x,\xi)} e^{2\pi i (x-y)\cdot \xi} f(y)\ d\xi dy.$$

Writing $\overline{a(y,\xi)} = \overline{a(x,\xi)} + \int_0^1 (x-y) \cdot a_x((1-t)x + ty, \xi)$, we conclude that

$$a(X,D)^* f(x) = \overline{a}(X,D) + \int_0^1 \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} (x-y) \overline{a_x((1-t)x+ty,\xi)} e^{2\pi i (x-y)\cdot \xi} f(y)\ d\xi dy.$$

So it will suffice to show that the operator

$$T_t f(x) := \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} (x-y) \cdot \overline{a_x((1-t)x + ty, \xi)} e^{2\pi i (x-y)\cdot \xi} f(y)\ d\xi dy$$

is a pseudodifferential operator of order $k-1$, uniformly for $0 \leq t \leq 1$. We can integrate by parts to eliminate the $x-y$, writing

$$T_t f(x) := \frac{-1}{2\pi i} \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} b((1-t)x + ty, \xi) e^{2\pi i (x-y)\cdot \xi} f(y)\ d\xi dy$$

where $b(x,\xi) := \nabla_\xi \cdot \overline{a_x(x,\xi)}$. Observe that b is a symbol of order $k-1$. Some playing around with the Fourier inversion formula reveals that we can write $T_t = c_t(X,D)$ where

$$c_t(x,\eta) := \frac{-1}{2\pi i} \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} b((1-t)x + ty, \xi) e^{2\pi i (x-y)\cdot (\xi - \eta)}\ d\xi dy.$$

But from Lemma 2.12 we have that c_t is a symbol of order $k-1$, and the claim follows. ∎

Remark 2.14. A corollary of this lemma is that if a ΨDO of order k has real symbol, then it is self-adjoint modulo a ΨDO of order $k-1$, and conversely if a ΨDO of order k is self-adjoint, then it has a real symbol modulo a symbol of order $k-1$. If one switches to the Weyl calculus instead of the Kohn-Nirenberg calculus then one can drop the lower order terms here, and assert simply that a ΨDO is self-adjoint if and only if it has real symbol.

Next, we study compositions of pseudo-differential operators.

Lemma 2.15 (Composition of ΨDOs)**.** *Let $a(X,D)$ and $b(X,D)$ be pseudodifferential operators of order k, k' respectively. Then $a(X,D)b(X,D)$ is a pseudodifferential operator of order $k+k'$, and $a(X,D)b(X,D) - ab(X,D)$ is a pseudodifferential operator of order $k+k'-1$.*

Proof We assume a, b are compactly supported; one can remove these hypotheses by a limiting argument which we omit here. For Schwartz functions f, we expand

$$\begin{aligned} a(X,D)b(X,D)f(x) &:= \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} a(x,\xi) e^{2\pi i (x-y)\cdot\xi} b(X,D) f(y)\ dy d\xi \\ &= \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} a(x,\xi) b(y,\eta) e^{2\pi i (x-y)\cdot \xi} e^{2\pi i y \cdot \eta} \hat f(\eta)\ d\eta dy d\xi \\ &= c(X,D) f(x) \end{aligned}$$

where

$$c(x,\eta) := \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} a(x,\xi) b(y,\eta) e^{2\pi i (x-y)\cdot(\xi-\eta)}\ dy d\xi.$$

From Lemma 2.12 we obtain the first claim. To obtain the second claim, observe from the Fourier inversion formula that

$$ab(x,\eta) := \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} a(x,\xi) b(x,\eta) e^{2\pi i (x-y)\cdot(\xi-\eta)}\ dy d\xi$$

so it suffices to show that

$$\int_{\mathbf{R}^d} \int_{\mathbf{R}^d} a(x,\xi)(b(y,\eta) - b(x,\eta)) e^{2\pi i (x-y)\cdot(\xi-\eta)}\ dy d\xi$$

is a symbol of order $k+k'-1$. As in the previous lemma, we exploit the fundamental theorem of calculus to write

$$b(y,\eta) - b(x,\eta) = \int_0^1 (y-x)\cdot b_x((1-t)x+ty,\eta)\ dt$$

and then use integration by parts to deal with the $y-x$, eventually reducing one to showing that

$$\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} \nabla_\xi a(x,\xi)\cdot b_x((1-t)x+ty,\eta)e^{2\pi i(x-y)\cdot(\xi-\eta)}\ dyd\xi$$

is a symbol of order $k+k'-1$. But this follows from Lemma 2.12. ∎

It is instructive to establish the above lemma by hand in the case where $a(X,D)$ and $b(X,D)$ are variable-coefficient differential operators of order k and k' respectively. The defect between $a(X,D)b(X,D)$ and $ab(X,D)$ can be interpreted physically as the difference between quantum and classical mechanics.

The above lemma implies in particular that the commutator

$$[a(X,D),b(X,D)] := a(X,D)b(X,D) - b(X,D)a(X,D)$$

of two pseudodifferential operators of order k and k' will be an operator of order $k+k'-1$. In fact we can pin down this commutator more precisely. Define the *Poisson bracket* $\{a,b\}$ of a,b to be the quantity

$$\{a,b\} := \nabla_\xi a\cdot\nabla_x b - \nabla_x a\cdot\nabla_\xi b.$$

Thus for instance $\{x,\xi\} = -1$. One can easily verify that the Poisson bracket of two symbols of order k,k' respectively will be a symbol of order $k+k'-1$.

Lemma 2.16 (Commutator of ΨDOs)**.** *Let $a(X,D)$ and $b(X,D)$ be pseudodifferential operators of order k, k' respectively. Then $[a(X,D),b(X,D)] - \frac{1}{2\pi i}\{a,b\}(X,D)$ is a pseudodifferential operator of order $k+k'-2$.*

Proof From the proof of the previous lemma, we see that $[a(X,D),b(X,D)]$ is a pseudodifferential operator with symbol

$$\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} [a(x,\xi)b(y,\eta) - a(y,\eta)b(x,\xi)]e^{2\pi i(x-y)\cdot(\xi-\eta)}\ dyd\xi.$$

We Taylor expand

$$b(y,\eta) = b(x,\xi)+(y-x)\cdot\nabla_x b(x,\xi)+(\eta-\xi)\cdot\nabla_\eta b(x,\xi)+\int_0^1\int_0^1 [(y-x)\cdot\nabla_x][(\eta-\xi)\cdot\nabla_\eta]b((1-t)x+ty,(1-s)\xi+s\eta)\ dsdt$$

and similarly for $a(y,\eta)$. Expanding this we obtain a variety of terms. Any factor of $(y-x)$ can be converted into $\frac{-1}{2\pi i}$ a ξ derivative on $e^{2\pi i(x-y)\cdot(\xi-\eta)}$, which can then by integration by parts be converted to $\frac{1}{2\pi i}$ times a ξ derivative somewhere else. Similarly, any factor of $(\eta-\xi)$ can be converted to $\frac{-1}{2\pi i}$ times a y derivative somewhere else. Because of this, any term which contains two or more factors of $(y-x)\cdot\nabla_x$ or $(\eta-\xi)\cdot\nabla_\eta$ can be integrated to an expression which has two or more derivatives in ξ or η on the symbols a,b, and Lemma 2.12 will show that such terms give symbols of order $k+k'-2$. Thus we only need to consider terms with exactly one factor of $(y-x)\cdot\nabla_x$ or $(\eta-\xi)\cdot\nabla_\eta$ (the terms with zero factors cancel each other out). In other words, we can replace the expression $a(x,\xi)b(y,\eta) - a(y,\eta)b(x,\xi)$ with

$$a(x,\xi)(y-x)\cdot\nabla_x b(x,\xi)+a(x,\xi)(\eta-\xi)\cdot\nabla_\eta b(x,\xi)-b(x,\xi)(y-x)\cdot\nabla_x a(x,\xi)+b(x,\xi)(\eta-\xi)\cdot\nabla_\eta a(x,\xi).$$

Converting the $(y-x)$ and $(\eta-\xi)$ terms to derivatives by integration by parts, as outlined above, we can re-express this as

$$\frac{1}{2\pi i} a_\xi(x,\xi)\cdot b_x(x,\xi) - \frac{1}{2\pi i} a_x(x,\xi)\cdot b_\xi(x,\xi) = \frac{1}{2\pi i}\{a,b\}(x,\xi)$$

and the claim follows. ■

We thus see that the algebraic structure of pseudodifferential operators is intimately connected to the Poisson geometry structure of the phase plane $\{(x,\xi) : x,\xi \in \mathbf{R}^d\}$, which in turn is induced by the symplectic geometry of that plane. There is a rich theory of how symplectic geometry (which represents Hamiltonian or "classical" mechanics) interacts with pseudodifferential operators (which represents quantum mechanics); this is the subject of semiclassical analysis and geometric quantisation. However, these topics would take us too far afield.

3. Sobolev spaces

With all the various inequalities at our disposal (the Hörmander-Mikhlin multiplier theorem, the Calderón-Vaillancourt theorem, and the Hardy-Littlewood-Sobolev inequality) we can now quickly develop some of the classical theory of homogeneous and inhomogeneous Sobolev spaces on Euclidean spaces $\mathbf{R}^d$; of course we will not attempt an exhaustive survey of this vast subject here[5]. We begin with the inhomogeneous spaces which contain fewer technicalities.

Definition 3.1 (Inhomogeneous Sobolev norms)**.** Let $1 < p < \infty$ and $s \in \mathbf{R}$. If $f \in \mathcal{S}(\mathbf{R}^d)$ is a Schwartz function, we define the inhomogeneous Sobolev norm $\|f\|_{W^{s,p}(\mathbf{R}^d)}$ by the formula

$$\|f\|_{W^{s,p}(\mathbf{R}^d)} := \|\langle\nabla\rangle^s f\|_{L^p(\mathbf{R}^d)}.$$

We let $W^{s,p}(\mathbf{R}^d)$ be the closure of the Schwartz functions under this norm.

Remark 3.2. When $p = 2$, the spaces $W^{s,p}(\mathbf{R}^d)$ are often denoted $H^s(\mathbf{R}^d)$ (but not to be confused with the Hardy spaces $\mathcal{H}^p(\mathbf{R}^d)$). From the Fourier inversion formula we see that

$$\|f\|_{H^s(\mathbf{R}^d)} \sim_{s,d} \|\langle\xi\rangle^s \hat{f}\|_{L^2(\mathbf{R}^d)}.$$

Clearly $W^{0,p}(\mathbf{R}^d)$ is isometric to $L^p(\mathbf{R}^d)$, so Sobolev spaces include Lebesgue spaces as special cases.

One can easily verify that Sobolev spaces are Banach spaces, and that the dual of $W^{s,p}(\mathbf{R}^d)$ can be identified with $W^{-s,p'}(\mathbf{R}^d)$. We have the following basic embeddings.

[5] In particular, we omit one important analytical aspect of these spaces, namely their local stability under smooth diffeomorphisms, which allows one to transplant these spaces onto manifolds and thus have application to various problems in differential geometry.

Proposition 3.3 (Sobolev embeddings)**.** *Let $1 < p < \infty$ and $s \in \mathbf{R}$ and $f \in W^{s,p}(\mathbf{R}^d)$.*

(i) *(Monotonicity in s) If $s' < s$ then $f \in W^{s',p}(\mathbf{R}^d)$, and*

$$\|f\|_{W^{s',p}(\mathbf{R}^d)} \lesssim_{p,s,s',d} \|f\|_{W^{s,p}(\mathbf{R}^d)}. \tag{6}$$

(ii) *(Behaviour with respect to pseudodifferential operators) If $a(X,D)$ is a pseudodifferential operator of order k for some $k \in \mathbf{R}$, then a extends continuously from $W^{s,p}$ to $W^{s-k,p}$:*

$$\|a(X,D)f\|_{W^{s-k,p}(\mathbf{R}^d)} \lesssim_{p,s,d,k} \|f\|_{W^{s,p}(\mathbf{R}^d)}.$$

(iii) *(Behaviour with respect to Hörmander-Mikhlin multipliers) If $m(D)$ is a Hörmander-Mikhlin multiplier, then $m(D)$ extends to a bounded linear operator on $W^{s,p}(\mathbf{R}^d)$, with*

$$\|m(D)f\|_{W^{s,p}(\mathbf{R}^d)} \lesssim_{p,s,d} \|f\|_{W^{s,p}(\mathbf{R}^d)}.$$

(iv) *(Characterisation using derivatives) If $f, \nabla^k f \in W^{s,p}(\mathbf{R}^d)$ for*[6] *some integer $k \geq 0$, then $f \in W^{s+k,p}(\mathbf{R}^d)$ and*

$$\|f\|_{W^{s+k,p}(\mathbf{R}^d)} \sim_{p,s,d,k} \|f\|_{W^{s,p}(\mathbf{R}^d)} + \|\nabla^k f\|_{W^{s,p}(\mathbf{R}^d)} \sim_{p,s,d,k} \sum_{j=0}^k \|\nabla^j f\|_{W^{s,p}(\mathbf{R}^d)}.$$

In particular,

$$\|f\|_{W^{k,p}(\mathbf{R}^d)} \sim_{p,d,k} \|f\|_{L^p(\mathbf{R}^d)} + \|\nabla^k f\|_{L^p(\mathbf{R}^d)} \sim_{p,d,k} \sum_{j=0}^k \|\nabla^j f\|_{W^{s,p}(\mathbf{R}^d)}.$$

(v) *(Sobolev embedding theorem) If $p < q < \infty$ is such that $\frac{d}{q} \geq \frac{d}{p} - s$, then $f \in L^q(\mathbf{R}^d)$ and*

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

If instead $s > d/p$, then f is bounded and continuous with

$$\|f\|_{L^\infty(\mathbf{R}^d)} \lesssim_{p,s,q,d} \|f\|_{W^{s,p}(\mathbf{R}^d)}.$$

Proof To prove (i), observe that $\langle\nabla\rangle^{s'-s}$ is certainly a Hörmander-Mikhlin multiplier (or a pseudodifferential operator of order $s'-s$ and hence of order 0), and hence bounded on $L^p(\mathbf{R}^d)$. One can then quickly verify (6) for Schwartz functions and then extend by density.

To prove (ii), observe from the pseudodifferential calculus that $\langle\nabla\rangle^{s-k} a(X,D) \langle\nabla\rangle^{-s}$ is a pseudodifferential operator of order 0, and hence bounded on $L^p(\mathbf{R}^d)$. The claim then follows as in (i).

To prove (iii), use the Hörmander multiplier theorem coupled with the observation that $m(D)$ commutes with $\langle\nabla\rangle^s$ (working in the Schwartz category to begin with).

[6] Note from (ii) that $\nabla^k f$ is already well-defined as an element of $W^{s-k,p}(\mathbf{R}^d)$ at least.

To prove (iv), we can use Fourier multiplier calculus to write $\langle\nabla\rangle^k = m_0(D) + m_1(D)\cdot\nabla^k$ where m_0, m_1 are Hörmander-Mikhlin multipliers, and hence

$$\langle\nabla\rangle^{s+k} f = m_0(D)\langle\nabla\rangle^s f + m_1(D)\cdot\langle\nabla\rangle^s\nabla^k f$$

Taking L^p norms (first for Schwartz functions, and then one can take weak limits) we establish that

$$\|f\|_{W^{s+k,p}(\mathbf{R}^d)} \lesssim_{p,s,d,k} \|f\|_{W^{s,p}(\mathbf{R}^d)} + \|\nabla^k f\|_{W^{s,p}(\mathbf{R}^d)}.$$

The remaining inequalities then follow from (ii).

To prove (v), we may lower s to assume that $0 < s < d$. We need to show that $\langle\nabla\rangle^{-s}$ maps $L^p(\mathbf{R}^d)$ to $L^q(\mathbf{R}^d)$ in the first case, or $L^p(\mathbf{R}^d)$ to $L^\infty(\mathbf{R}^d)$ in the second case. But by Lemma 5.3, the convolution kernel $K(x)$ of this Fourier multiplier is $O_{s,d,N}(\min(|x|^{-d+s}, |x|^{-N}))$ for any $N > 0$. In particular it is $O_{s,d,q}(|x|^{-d/p'-d/q})$, at which point the first claim follows from the Hardy-Littlewood-Sobolev inequality; if $s > d/p$ then the kernel lies in $L^{p'}$, at which point we get the second claim (the continuity follows by starting with Schwartz functions and taking limits). ■

Next, we establish a Littlewood-Paley characterisation of Sobolev spaces.

Theorem 3.4 (Littlewood-Paley characterisation)**.** *Let $1 < p < \infty$ and $s \in \mathbf{R}$. Let $\phi_0 : \mathbf{R}^d \to \mathbf{R}$ be a bump function adapted to the ball $B(0,2)$, and for each $j \geq 1$ let $\psi_j : \mathbf{R}^d \to \mathbf{R}$ be a bump function adapted to the annulus $B(0,2^{j+1})\backslash B(0,2^{j-1})$, such that one has the pointwise estimate $|\phi_0|^2 + \sum_{j=1}^\infty |\psi_j|^2 \sim 1$. Then for any $f \in W^{s,p}(\mathbf{R}^d)$ we have*

$$\|f\|_{W^{s,p}(\mathbf{R}^d)} \sim_{s,p,d} \|\phi_0(D)f\|_{L^p(\mathbf{R}^d)} + \|(\sum_{j=1}^\infty 2^{2js}|\psi_j(D)f|^2)^{1/2}\|_{L^p(\mathbf{R}^d)}.$$

Proof Let us just prove the upper bound. From Proposition 3.3(ii) we already know that $\phi_0(D)$ maps $W^{s,p}$ to L^p, so it suffices to show that

$$\|(\sum_{j=1}^\infty 2^{2js}|\psi_j(D)f|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \lesssim_{s,p,d} \|f\|_{W^{s,p}(\mathbf{R}^d)}.$$

It suffices to do this for Schwartz functions. Substituting $g := \langle\nabla\rangle^s f$, our task is to show

$$\|(\sum_{j=1}^\infty |2^{js}\psi_j(D)\langle\nabla\rangle^{-s} g|^2)^{1/2}\|_{L^p(\mathbf{R}^d)} \lesssim_{s,p,d} \|f\|_{L^p(\mathbf{R}^d)}.$$

But observe that $2^{js}\psi_j(D)\langle\nabla\rangle^{-s}$ is a Fourier multiplier whose symbol is adapted to the annulus $\{|\xi| \sim 2^{js}\}$, so the claim follows from the upper Littlewood-Paley inequality.

The lower bound can be proven in a similar manner using the full Littlewood-Paley inequality, as well as a further frequency decomposition into low frequencies $|\xi| \lesssim 1$ and high frequencies $|\xi| \gtrsim 1$, and we leave this as an exercise to the reader. ■

Sobolev spaces can be interpolated with each other fairly easily using the complex interpolation method. We present here a sample result to give the flavour.

Proposition 3.5. *Let $T : \mathcal{S} \to \mathcal{S}$ be a linear operator on Schwartz functions which is bounded on both $W^{s_0,p_0}(\mathbf{R}^d)$ and $W^{s_1,p_1}(\mathbf{R}^d)$ for some $1 < p_0, p_1 < \infty$ and $s_0, s_1 \in \mathbf{R}$. Then T is also bounded on $W^{s_\theta,p_\theta}(\mathbf{R}^d)$ for $0 \leq \theta \leq 1$, where $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$ and $s_\theta := (1-\theta)s_0 + \theta s_1$.*

Proof By hypothesis, the operator $\langle\nabla\rangle^{s_j} T \langle\nabla\rangle^{-s_j}$ is bounded on $L^{p_j}(\mathbf{R}^d)$ for $j = 0, 1$. By the Hörmander-Mikhlin multiplier theorem, we also see that $\langle\nabla\rangle^{s_j+it} T \langle\nabla\rangle^{-s_j-it}$ is bounded on $L^{p_j}(\mathbf{R}^d)$ for $t \in \mathbf{R}$ whose operator norm grows at most polynomially in t. The claim then follows from the Stein interpolation theorem. ∎

We now briefly discuss homogeneous Sobolev spaces $\dot{W}^{s,p}(\mathbf{R}^d)$, defined using $|\nabla|^s$ instead of $\langle\nabla\rangle^s$; these have the advantage of behaving well under dilations Dil^q_λ, although they have some other drawbacks to compensate for this. Here one can run into some technicalities for the very low regularities ($s \leq -d$), because the operator $|\nabla|^s$ is not well defined distributionally in that case. This problem can be avoided by restricting the Schwartz functions to avoid the frequency origin:

Definition 3.6 (Homogeneous Sobolev norms)**.** Let $1 < p < \infty$ and $s \in \mathbf{R}$. If $f \in \mathcal{S}(\mathbf{R}^d)$ is a Schwartz function whose Fourier transform vanishes near the origin, we define the homogeneous Sobolev norm $\|f\|_{\dot{W}^{s,p}(\mathbf{R}^d)}$ by the formula

$$\|f\|_{\dot{W}^{s,p}(\mathbf{R}^d)} := \||\nabla|^s f\|_{L^p(\mathbf{R}^d)}.$$

We let $\dot{W}^{s,p}(\mathbf{R}^d)$ be the closure of all such functions under this norm.

Because $\dot{W}^{s,p}$ does not always contain all Schwartz functions, one has to sometimes take a little care with arguments. Nevertheless, one can still establish a reasonable theory for these spaces. For instance, ∇^k is an isomorphism between $\dot{W}^{s,p}$ and $\dot{W}^{s-k,p}$, and we have the homogeneous Sobolev embedding $\dot{W}^{s,p}(\mathbf{R}^d) \subset L^q(\mathbf{R}^d)$ whenever $1 < p < q < \infty$ and $d/q = d/p - s$. For positive s we also have the relationship $W^{s,p}(\mathbf{R}^d) = \dot{W}^{s,p}(\mathbf{R}^d) \cap L^p(\mathbf{R}^d)$, with

$$\|f\|_{W^{s,p}(\mathbf{R}^d)} \sim_{s,p,d} \|f\|_{L^p(\mathbf{R}^d)} + \|f\|_{\dot{W}^{s,p}(\mathbf{R}^d)};$$

however we caution that this relationship breaks down for negative s. We leave these facts (which are simple modifications of the arguments given above) to the reader.

4. Appendix: the geometric Hahn-Banach theorem

Here we give a geometric formulation of the Hahn-Banach theorem, sometimes also called the *Dieudonné hyperplane separation theorem.*

Definition 4.1 (Algebraic openness)**.** A set $A \subset V$ in a (real) vector space is *algebraically open* if the set $\{t \in \mathbf{R} : v + tw \in A\}$ is open for all $v, w \in V$ (i.e. the intersection of A with any line is an open subset of that line).

Algebraic openness is a very mild property; indeed, the topological vector space structure given by weak openness is finer than any other topological vector space structure. For instance, in a normed vector space, any open set will be algebraically open.

Theorem 4.2 (Geometric Hahn-Banach theorem)**.** *Let A, B be disjoint convex subsets of a real vector space V, with A algebraically open. Then there exists a linear functional $\lambda : V \to \mathbf{R}$ and $c \in \mathbf{R}$ such that $\lambda < c$ on A, and $\lambda \geq c$ on B.*

Remark 4.3. In finite dimensions, it is not difficult to drop the algebraic openness hypothesis on A as long as one now replaces the condition $\lambda < c$ by $\lambda \leq c$. However in infinite dimensions one cannot do this. Consider for instance the space $V = \bigcup_{n=0}^\infty \mathbf{R}^n$ of sequences $(x_n)_{n=1}^\infty$ with only finitely many x_n non-zero, let A consist of those sequences whose last non-zero element is strictly positive, and $B = -A$ consist of those sequences whose last non-zero element is strictly negative. Then there is no hyperplane separating A from B.

Remark 4.4. If B is algebraically open, and A, B are non-empty, then λ is non-zero, and it is not hard to see that the condition $\lambda \geq c$ can be upgraded to $\lambda > c$. If A contains the origin, then c must be positive, and can then be rescaled to be 1.

Proof We first observe that it suffices to verify the homogeneous case, when A, B are convex cones and $c = 0$. Indeed, to then establish the general case, one applies the homogeneous case to the convex cones $A', B' \in \mathbf{R} \times V$ defined by

$$A' := \{(t, tx) : t > 0, x \in A\}; \quad B' := \{(t, tx) : t > 0, x \in B\};$$

we leave the details to the reader.

Consider all the pairs (A, B) of disjoint convex cones, with A algebraically open. We can order these pairs by set inclusion, so that $(A, B) \leq (A', B')$ whenever $A \subseteq A'$ and $B \subseteq B'$, and observe that every chain has an upper bound. By Zorn's lemma, we thus see that to prove the claim it suffices to do so under the additional assumption that (A, B) is *maximal*. (This is the one and only place where we use (crucially) the axiom of choice.)

We can of course assume that neither A nor B is empty. We now claim that B is the complement of A. For if not, then there exists $v \in V$ which does not lie in either A or B. By the maximality of (A, B), the convex cone generated by $B \cup \{v\}$ must intersect A at some point, say w. By dilating w if necessary we may assume that w lies on a line segment between v and some point b in B. By using the convexity and disjointness of A and B one can then deduce that for any $a \in A$, the ray $\{a + t(w - b) : t > 0\}$ is disjoint from B. Thus one can enlarge A to the convex cone generated by A and $w - b$, which is still algebraically open and now strictly larger than A (because it contains v), a contradiction. Thus B is the complement of A.

Let us call a line in V *monochromatic* if it is entirely contained in A or entirely contained in B. Note that if a line is not monochromatic, then (because A and B are convex and partition the line, and A is algebraically open), the line splits into an open ray contained in A, and a closed ray contained in B. From this we can conclude that if a line is monochromatic, then all parallel lines must also be monochromatic, because otherwise we look at the ray in the parallel line which contains A and use convexity of both A and B to show that this ray is adjacent to a halfplane contained in B, contradicting algebraic openness. Now let W be the space of all vectors w for which there exists a monochromatic line in the direction w (including 0). Then W is easily seen to be a vector space; since A, B are non-empty, W is a proper subspace of V. On the other hand, if w and w' are not in W, some playing around with the property that A and B are convex sets partitioning V shows that the plane spanned by w and w' contains a monochromatic line, and hence some non-trivial linear combination of w and w' lies in W. Thus V/W is precisely one-dimensional. Since every line with direction in w is monochromatic, A and B also have well-defined quotients A/W and B/W on this one-dimensional subspace, which remain convex (with A/W still algebraically open). But then it is clear that A/W and B/W are an open and closed ray from the origin in V/W respectively. It is then a routine matter to construct a linear functional $\lambda : V \to \mathbf{R}$ (with null space W) such that $A = \{\lambda < 0\}$ and $B = \{\lambda \geq 0\}$, and the claim follows. ■

To illustrate the power of this theorem, let us give a famous consequence:

Corollary 4.5 (Hahn-Banach theorem)**.** *Let V be a normed vector space, and let W be any subspace of V. Then any bounded linear functional $\lambda : W \to \mathbf{R}$ has an extension $\tilde{\lambda} : V \to \mathbf{R}$ with the same norm.*

Proof We may normalise $\|\lambda\|_{W^*} = 1$. Then the algebraically open convex sets $\{v \in V : \|v\|_V < 1\}$ and $\{w \in W : \lambda(w) > 1\}$ are disjoint, with the former containing 0, and so by the geometric Hahn-Banach theorem we can find $\tilde{\lambda}$ such that $\tilde{\lambda} < 1$ on the first set and $\tilde{\lambda} > 1$ on the second. The former fact establishes that $\tilde{\lambda}$ is bounded on V with norm at most 1, and the latter implies that the null space of $\tilde{\lambda}$ contains the null space of λ. Restricting to W we then quickly conclude that $\tilde{\lambda}$ when restricted to W equals λ, and the claim follows. ■

In the converse direction, one can deduce the geometric Hahn-Banach theorem from its more familiar formulations, but this requires some work. The first key observation is that to prove the geometric Hahn-Banach theorem it suffices to do so when B is the origin $\{0\}$, since the general case will eventuall follow from the trick of replacing the pair (A, B) by $(A - B, \{0\})$, where $A - B := \{a - b : a \in A, b \in B\}$ is the Minkowski difference of A and B. The remainder of the argument proceeds either by mimicking the usual proof of the Hahn-Banach theorem or by building a norm somehow out of $A - B$.

5. Appendix: Schwartz functions

To facilitate computations we devise some notion regarding bump functions. (Retrospectively, having this appendix in last week's notes would have helped out with the proof of the Hörmander-Mikhlin theorem.)

We already have defined a notion of what it means for a function ϕ to be a bump function adapted to a region such as a ball or an annulus; now we modify the notion to also cover Schwartz functions adapted to similar regions.

Definition 5.1. Let $B(x_0, r)$ be a ball and $H > 0$. We say that a function $\psi : \mathbf{R}^d \to \mathbf{C}$ is a *Schwartz function of height H adapted to the ball $B(x_0, r)$* if we have a representation

$$\psi(x) = H\phi(\frac{x - x_0}{r})$$

for some ϕ which is boundedly Schwartz in the sense that

$$|\nabla_x^k \phi(x)| \lesssim_{k,d,N} \langle x\rangle^{-N-k}$$

for all $k, N \geq 0$.

Observe that one may equivalently define a Schwartz function of height H adapted to $B(x_0, r)$ to be a function ψ which obeys the bounds

$$|\nabla_x^k \psi(x)| \lesssim_{k,d,N} Hr^{-k}\langle\frac{x - x_0}{r}\rangle^{-N-k}$$

for all $k, N \geq 0$. Note that any bump function adapted to $B(x_0, r)$ (or any region similar to $B(x_0, r)$, such as the annulus $\{x : |x-x_0| \sim r\}$ is also a Schwartz function of height 1 adapted to $B(x_0, r)$.

For Schwartz functions adapted to balls centred at the origin, one can compute their Fourier transform easily:

Lemma 5.2. *Let ϕ be a Schwartz function of height H adapted to the ball $B(0, r)$. Then $\hat{\phi}$ and $\check{\phi}$ are a Schwartz functions of height Hr^d adapted to the ball $B(0, 1/r)$.*

Proof We can normalise $H = 1$; by rescaling we can normalise $r = 1$. The claim then follows from the usual proof of the fact that the Fourier transform (or inverse Fourier transform) of a Schwartz function is Schwartz. ■

This allows us to take the Fourier transform of symbols:

Lemma 5.3. *Let $m : \mathbf{R}^d \to \mathbf{R}$ be a Schwartz function which is also a symbol of order k for some $k > -d$, thus*

$$|\partial_\xi^\alpha m(\xi)| \lesssim_{\alpha,k,d} \langle\xi\rangle^{k-|\alpha|}$$

for all $\xi \in \mathbf{R}^d$ and all multiindices α. Then $\check{m}$ obeys the dual symbol estimates

$$|\partial_x^\alpha \check{m}(x)| \lesssim_{\alpha,k,d,N} |x|^{-d-k-|\alpha|}\langle x\rangle^{-N} \tag{7}$$

for all $x \in \mathbf{R}^d$, all multiindices α, and all $N \geq 0$.

Remark 5.4. The qualitative hypothesis that m to be Schwartz is needed to ensure that the inverse Fourier transform of m makes sense classically. It can be removed, but at the cost of interpreting $\check{m}$ as a distribution rather than as a function (and excluding the origin $x = 0$, which may now be singular).

Proof Using smooth dyadic cutoffs we may split

$$m = m_0 + \sum_{j=1}^\infty 2^{jk} m_j$$

where m_0 is a bump function adapted to the ball $\{|\xi| \leq 1\}$ and each m_j for $j \geq 1$ is a bump function adapted to the annulus $\{|\xi| \sim 2^j\}$. In particular, for each $j \geq 0$, m_j is a Schwartz function of height 1 adapted to $B(0, 2^j)$. Then

$$\check m = \check m_0 + \sum_{j=1}^\infty 2^{jk} \check m_j.$$

By Lemma 5.2, $\check m_0$ is a Schwartz function of height 1 adapted to $B(0,1)$, while $\check m_j$ is a Schwartz function of height 2^{dj} adapted to $B(0, 2^{-j})$. On summing we obtain the claim. ■

6. Exercises

- Q1. (Hardy-Littlewood maximal inequality for A_p weights) Let $w : \mathbf{R}^d \to \mathbf{R}^+$ be strictly positive, and let $1 < p < \infty$. Show that the following three statements are equivalent up to changes in the implied constant:
- (i) For every ball B, we have $(⨍_B w)(⨍_B w^{-p'/p})^{-p'/p} \lesssim_{p,d} 1$.
- (ii) For every ball B, we have $(⨍_B w)(⨍_B w^{-p'/p})^{-p'/p} \sim_{p,d} 1$.
- (iii) For every locally integrable function f and $\lambda > 0$, we have

$$\int_{\mathbf{R}^d} 1_{Mf(x) > \lambda} w(x)\ dx \lesssim_{p,d} \frac{\int_{\mathbf{R}^d} |f(x)|^p w(x)\ dx}{\lambda^p}.$$

(Note: Weights w with the above properties are known as A_p *weights*. Somewhat counter-intuitively, the weighted weak-type (p,p) estimate in (iii) is in fact equivalent to its strong-type counterpart, due to a certain "openness" property of A_p weights, but this will not be shown here.)

- Q2. (Calderón-Zygmund theory for power weights) Let $T : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$ be a CZO. Show that for any $1 < p < \infty$ and $-d < \alpha < \frac{dp'}{p}$ that

$$\int_{\mathbf{R}^d} |Tf(x)|^p \langle x \rangle^\alpha \lesssim_{d,p,\alpha} \int_{\mathbf{R}^d} |f(x)|^p \langle x \rangle^\alpha$$

and

$$\int_{\mathbf{R}^d} |Tf(x)|^p |x|^\alpha \lesssim_{d,p,\alpha} \int_{\mathbf{R}^d} |f(x)|^p |x|^\alpha$$

for all bounded, compactly supported f. (Note that the former can be obtained from the latter by a scaling argument.) Hint: use duality to reduce matters to establishing a bilinear weighted estimate for $\int_{\mathbf{R}^d} Tf(x)\overline{g(x)}\ dx$, and then decompose f and g into dyadic pieces. For the "diagonal" interactions use the fact that T is bounded on L^p, while for the off-diagonal interactions use something like Q8 from Week 2 notes.

(Note: the above theory works when $\langle x \rangle^\alpha$ or $|x|^\alpha$ is replaced with a general A_p weight, but this is harder to prove.)

- Q3. Establish a partial converse to Lemma 5.3: if m is Schwartz and $\check m$ obeys (7) with $k < 0$, then m is a symbol of order k.

- Q4. Show that if a is a symbol of order k, then $a^w(X,D) = a(X,D) + b(X,D)$ and $a(X,D) = a^w(X,D) + c^w(X,D)$, where b and c are symbols of order $k-1$. Thus the Kohn-Nirenberg and Weyl quantisations are equivalent modulo lower order operators.
- Q5. (Endpoint Sobolev embedding) Let $1 < p < \infty$ and $s = d/p$. Show that if $f \in \dot W^{s,p}(\mathbf{R}^d)$, then $f \in \mathrm{BMO}(\mathbf{R}^d)$ with

$$\|f\|_{\mathrm{BMO}(\mathbf{R}^d)} \lesssim_{d,p} \|f\|_{\dot W^{s,p}(\mathbf{R}^d)}.$$

In fact, slightly more is true: show that f has *vanishing mean oscillation* in the sense that for any $x \in \mathbf{R}^d$ we have

$$⨍_{B(x,r)} |f - ⨍_{B(x,r)} f| \to 0 \hbox{ as } r \to 0 \hbox{ or } r \to \infty.$$

- Q6. (Hölder-Sobolev embedding) Let Let $1 < p < \infty$, $0 < \delta < 1$, and $s = d/p + \delta$. Show that if $f \in \dot W^{s,p}(\mathbf{R}^d)$, then we have the Hölder estimate

$$|f(x) - f(y)| \lesssim_{d,p} \|f\|_{\dot W^{s,p}(\mathbf{R}^d)} |x-y|^\delta$$

for all $x, y \in \mathbf{R}^d$.

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

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