LECTURE NOTES 2 FOR 247A

TERENCE TAO

1. Complex interpolation

It is instructive to compare the real interpolation method of the preceding section with the *complex interpolation method*. Now we need to take T to be linear rather than sublinear; we now make the mild assumption that the form

$$\int_X Tfg d\mu \qquad (1)$$

makes sense for all simple functions f, g of finite measure support.

Theorem 1.1 (Riesz-Thorin complex interpolation)**.** *Let T be a linear operator such that the form (1) is well-defined. Let $0 < p_0, p_1 \leq \infty$, $1 \leq q_0, q_1 \leq \infty$ and $A_0, A_1 > 0$ be such that we have the strong-type bounds*

$$\|Tf\|_{L^{q_i}(Y)} \leq A_i \|f\|_{L^{p_i}(X)}$$

for all simple functions f of finite measure support, and $i = 0, 1$. Then we have

$$\|Tf\|_{L^{q_\theta}(Y)} \leq A_\theta \|f\|_{L^{p_\theta}(X)}$$

for all simple functions f of finite measure support and $0 \leq \theta \leq 1$, where p_θ, q_θ, A_θ is as in equation (23) of last week's notes.

Proof We may assume that $(p_0, q_0) \neq (p_1, q_1)$ otherwise the claim is trivial. We may then normalise $A_0 = A_1 = 1$ as in the previous week's notes. By duality and homogeneity it suffices to show that

$$|\int_X Tf_\theta g_\theta \ d\mu| \leq 1 \text{ whenever } \|f_\theta\|_{L^{p_\theta}(X)} = \|g_\theta\|_{L^{q'_\theta}(Y)} = 1$$

for all simple functions f_θ, g_θ of finite measure support and all $0 \leq \theta \leq 1$. Note that the claim is already true by hypothesis when $\theta = 0, 1$.

The idea is to use the three lines lemma from last week's notes. The problem is that the inequality is not complex analytic in θ as stated. However we can fix this as follows. Observe that if f_θ is a simple function with $L^{p_\theta}(X)$ norm 1, then we can factorise

$$f_\theta = F_0^{1-\theta} F_1^\theta a$$

where F_0, F_1 are non-negative simple functions with $L^{p_0}(X)$ and $L^{p_1}(X)$ norms respectively equal to 1, and a is a simple function of magnitude at most 1. Indeed we can set $a = \operatorname{sgn}(f)$ and $F_i = |f_\theta|^{p_\theta/p_i}$. (Some minor changes need to be made for the limiting case when one or both of the p_i are equal to infinity; we leave this to the reader.) Similarly we can factorise

$$g_\theta = G_0^{1-\theta} G_1^\theta b$$

where G_0, G_1 are non-negative simple functions with $L^{q_0'}(Y)$ and $L^{q_1'}(Y)$ norms respectively equal to 1, and b is a simple function of magnitude at most 1.

Now consider the complex-valued function

$$H(z) := \int_X T(F_0^{1-z} F_1^z a) G_0^{1-z} G_1^z b \; d\mu.$$

Because T is linear and all functions are simple, it is not difficult to see that H is an entire function of z of at most exponential growth. It is bounded by 1 on both sides of the strip $0 \leq \mathrm{Re}(z) \leq 1$, hence bounded also within the strip. Setting $z = \theta$ we obtain the result. ■

Comparing the Riesz-Thorin theorem with the Marcinkiewicz theorem, we observe some key differences. On the plus side, we do not lose an unspecified constant in the estimates, and we do not have the restriction $q \geq p$. On the other hand, the hypotheses require strong-type control on T, not restricted weak-type control.

The Riesz-Thorin theorem can also be rephrased as follows. Let $\|T\|_{L^p(X) \to L^q(Y)}$ be the operator norm of T from $L^p(X)$ to $L^q(Y)$; then this quantity is log-convex as a function of $(1/p, 1/q)$. Because of this, the above theorem is sometimes known as the *Riesz-Thorin convexity theorem.*

An important observation of Elias Stein is that the above proof can be generalised to the case where the operator T itself varies analytically:

Theorem 1.2 (Stein interpolation)**.** *Let T_z be a family of linear operators on the strip $0 \leq \mathrm{Re}(z) \leq 1$ such that the for each $f \in D_X$, $g \in D_Y$, the form*

$$\int_Y (T_z f) g \; d\mu_X$$

is absolutely convergent and a complex analytic function of z on the strip which grows slower than double-exponentially in z. Let $0 < p_0, p_1 \leq \infty$, $1 \leq q_0, q_1 \leq \infty$ and $A_0, A_1 > 0$ be such that we have the strong-type bounds

$$\|T_z f\|_{L^{q_i}(Y)} \leq A_i \|f\|_{L^{p_i}(X)}$$

for all simple functions f of finite measure support, $i = 0, 1$, and $\mathrm{Re}(z) = i$. Then we have

$$\|T_\theta f\|_{L^{q_\theta}(Y)} \leq A_\theta \|f\|_{L^{p_\theta}(X)}$$

for all simple functions f of finite measure support and $0 \leq \theta \leq 1$.

Note that the Riesz-Thorin theorem corresponds to the case where $T_z = T$ is independent of z.

Proof We repeat the previous argument; the only observation to make is that the function

$$H(z) := \int_X T_z(F_0^{1-z}F_1^z a)G_0^{1-z}G_1^z b \, d\mu$$

continues to be complex-analytic; this is easiest seen by decomposing all the simple functions into indicator functions. ∎

One particularly common special case of this theorem is

Corollary 1.3 (Stein-Weiss interpolation theorem)**.** *Let T be a linear operator such that* (1) *is well-defined, and let $w_0, w_1 : X \to \mathbf{R}^+$ and $v_0, v_1 : Y \to \mathbf{R}^+$ be non-negative weights which are integrable on every finite measure set. Let $0 < p_0, p_1 \leq \infty$, $1 \leq q_0, q_1 \leq \infty$ and $A_0, A_1 > 0$ be such that we have the strong-type bounds*

$$\|Tf\|_{L^{q_j}(Y, v_j \ d\mu_Y)} \leq A_j \|f\|_{L^{p_j}(X, w_j \ d\mu_X)} \quad (2)$$

for all simple functions f of finite measure support, and $j = 0, 1$. Then we have

$$\|Tf\|_{L^{q_\theta}(Y, v_\theta \ d\mu_Y)} \leq A_\theta \|f\|_{L^{p_\theta}(X, w_\theta \ d\mu_X)}$$

for all simple functions f of finite measure support and $0 \leq \theta \leq 1$, where $v_\theta := v_0^{1-\theta} v_1^\theta$ and $w_\theta := w_0^{1-\theta} w_1^\theta$.

Proof We begin with a number of qualitative reductions. A monotone convergence argument (replacing the weights v_j, w_j by various truncated versions) lets one reduce to the case where v_j are bounded from above, and w_j are bounded from below. Another monotone convergence argument (this time replacing Y by various truncated versions) lets us reduce to the case where v_θ is bounded from below, which implies that the v_j are also bounded from below. One can also similarly restrict Y to have finite measure, and another truncation argument (to the support of f) also reduces X to have finite measure. Then another limiting argument (restricting f to the region where w_θ is bounded by some N, and then letting N go to infinity, using the hypotheses (2) and Fatou's lemma to get enough convergence) lets one reduce to the case where w_θ is bounded from above, which implies that w_i are also bounded from above. Now that all the v_j and w_j are bounded both above and below one can approximate by simple functions and use a limiting argument to reduce to the case where v_j, w_j are simple and never vanishing.

We can now legimately define the analytic family of operators

$$T_z := v_0^{(1-z)/q_0} v_1^{z/q_1} T w_0^{-(1-z)/p_0} w_1^{-z/p_1}$$

for any complex z, which is well defined on simple functions. The hypotheses (2) then ensure that $\|T_z\|_{L^{p_j}(X,\mu_X) \to L^{q_j}(Y,\mu_Y)} \leq A_j$ whenever $j = 0, 1$ and $\mathrm{Re}(z) = j$, the point being that multiplication by imaginary exponentials such as v_0^{it} are contractions on every L^p space; one also easily checks that the other hypotheses of the Stein interpolation theorem are satisfied. We conclude that

$$\|T_\theta\|_{L^{p_\theta}(X,\mu_X) \to L^{q_\theta}(Y,\mu_Y)} \leq A_\theta$$

from which the claim follows. ∎

The real and complex interpolation methods provide a surprisingly powerful way to prove estimates involving L^p norms: in order to prove a range of such estimates, it suffices to do so for the extreme cases (possibly weakening strong type to weak type or restricted weak type in the process). We shall shortly give several disparate examples of this interpolation approach.

2. Duality

Both real and complex interpolation give a means to deduce new bounds on an operator T from old ones. Another important technique is that of *duality*, which basically replaces the input function f with an output function T^*g, and the output function Tf with an input function g. Generally speaking, the situation is as follows. One has a domain X, and some class of "test functions" D_X on X (such as simple functions, or perhaps the Schwartz class), and one also has a domain Y with a class D_Y. We shall require D_X and D_Y to be vector spaces, though we will not require any topological structure on these spaces (in particular, they do not need to be complete with respect to any particular topology). We can then define the dual space D_Y^* of linear functionals $g \mapsto \langle g, h\rangle_Y$ on D_Y; we do not require these functionals to be continuous in any topology, and so D_Y^* is in fact a very large space. In particular, it typically contains very general objects such as locally integrable functions, measures, or distributions, where we (formally at least) identify a function (or distribution, etc.) $h(y)$ with the linear functional

$$g \mapsto \langle g, h\rangle_Y := \int_Y g(y)\overline{h(y)}\ d\mu_Y(y).$$

We also adopt the convention $\langle h, g\rangle_Y := \overline{\langle g, h\rangle_Y}$. One can of course have similar conventions for D_X and D_X^*.

We assume that T is a linear operator from D_X to D_Y^*, thus given any $f \in D_X$ and $g \in D_Y$ the sesquilinear form

$$(f, g) \mapsto \langle Tf, g\rangle_Y$$

makes sense. We then automatically have an adjoint $T^* : D_Y \to D_X^*$ defined by

$$\langle f, T^*g\rangle_X := \langle Tf, g\rangle_Y.$$

At present, T^*g only takes values in the abstract dual space D_X^*. However in many cases T^*g can be identified more concretely with a function, measure, or at least a distribution. For instance, if X, Y are topological measure spaces, D_X and D_Y are the compactly supported continuous functions on X and Y, and T is given by an integral operator

$$Tf(y) := \int_X K(x, y)f(x)\ d\mu_X(x)$$

for some locally integrable kernel $K : X \times Y \to \mathbf{C}$, then the sesquilinear form

$$\langle Tf, g\rangle_Y = \int_Y \int_X K(x, y)f(x)\overline{g(y)}\ d\mu_X(x)d\mu_Y(y)$$

makes sense for all $f \in D_X$, $g \in D_Y$, and by the Fubini-Tonelli theorem we see that the adjoint T^* is given by

$$T^*g(x) := \int_Y \overline{K(x,y)} g(y)\ d\mu_Y(y).$$

The point of introducing the adjoint T^* is that many bounds on the original operator T are equivalent to those for T^*. We illustrate this point with the L^p spaces, although from the proof one sees that the same would hold true for any spaces whose norm can be characterised by a duality relationship.

Theorem 2.1 (Duality). *Let $1 \leq p, q \leq \infty$ and $A > 0$. Let D_X and D_Y be linear spaces of functions on X and Y which are dense in $L^p(X)$ and $L^{q'}(Y)$ respectively. Let $T : D_X \to D_Y^*$ be a linear transformation, and let $T^* : D_Y \to D_X^*$ be its adjoint. Suppose we make the (qualitative) assumptions that T maps D_X to $L^q(Y)$ and T^* maps D_Y to $L^{p'}(X)$. Then the following (quantitative) assertions are equivalent.*

(i) *For all $f \in D_X$ we have $\|Tf\|_{L^q(Y)} \leq A\|f\|_{L^p(X)}$.*
(ii) *For all $g \in D_Y$ we have $\|T^*g\|_{L^{p'}(X)} \leq A\|g\|_{L^{q'}(Y)}$.*
(iii) *For all $f \in D_X$ and $g \in D_Y$ we have $|\langle Tf, g\rangle_Y| \leq A\|f\|_{L^p(X)}\|g\|_{L^{q'}(Y)}$.*
(iv) *For all $f \in D_X$ and $g \in D_Y$ we have $|\langle f, T^*g\rangle_X| \leq A\|f\|_{L^p(X)}\|g\|_{L^{q'}(Y)}$.*

Proof The equivalence of (i) with (iii) follows from the dual characterisation of the $L^q(Y)$ norm from last week's notes (and the density of D_X in $L^{q'}(Y)$). Similarly for the equivalence of (ii) with (iv). The equivalence of (iii) with (iv) follows from the definition of T^*. ■

Remark 2.2. This theorem differs very slightly from the usual duality theorem, which asserts that if T is continuous from a normed vector space V to another W, then its adjoint T^* is well defined from W^* to V^* and has the same operator norm. That duality theorem is fine for most applications, but runs into a slight difficulty when V is an L^∞ type space, as the dual can then get excessively large. (Having the domain space W^* too large is of course not a problem.) In particular when dealing with non-reflexive spaces such as L^1 or L^∞, one loses the equivalence of the two statements. By passing to dense classes and making the *a priori* qualitative hypotheses that T maps D_X to $L^q(Y)$ and T^* maps D_Y^* to $L^{p'}(X)$ (which are often easily verified in practice, since we can usually[1] take the dense classes D_X and D_Y to be very nice functions such as the Schwartz class).

3. Conditional expectation

Let us now start using our interpolation theorems. The first application is to the useful operation of *conditional expectation*. This operation was first developed in probability theory but has since proven to be useful in ergodic theory and harmonic analysis; furthermore, several further operations in harmonic analysis (such

[1] One caveat comes when L^∞ spaces are involved, because classes which decay at infinity, such as the Schwartz class, are usually not dense in L^∞.

as Littlewood-Paley projections) behave sufficiently like conditional expectation that it is worth studying conditional expectation simply to gain intuition on these other operators.

Let $(X, \mathcal{B}_{\max}, \mu)$ be a measure space (which need not necessarily be a probability space, though this is certainly an important special case), and let $\mathcal{B}$ be[2] a σ-finite sub-σ-algebra of $\mathcal{B}_{\max}$. Then $L^2(\mathcal{B}) = L^2(X, \mathcal{B}, \mu)$ is a closed subspace of the Hilbert space $L^2(\mathcal{B}_{\max}) = L^2(X, \mathcal{B}_{\max}, \mu)$, and thus has an orthogonal projection, which we shall denote by the map $f \mapsto \mathbf{E}(f|\mathcal{B})$. Thus $\mathbf{E}(f|\mathcal{B})$ is $\mathcal{B}$-measurable, and is equal to f if and only if f is also $\mathcal{B}$-measurable.

Problem 3.1. Let $X = \mathbf{R}$ with Lebesgue measure $d\mu = dx$, and let $\mathcal{B}_{\max}$ be the Borel or Lebesgue σ-algebra. Let $\mathcal{B}$ be the sub-σ-algebra generated by the intervals $[n, n+1)$ for $n \in \mathbf{Z}$. Show that for $f \in L^2(\mathcal{B}_{\max})$, $\mathbf{E}(f|\mathcal{B})$ is given by the formula

$$\mathbf{E}(f|\mathcal{B})(x) := \int_{\lfloor x \rfloor}^{\lfloor x \rfloor + 1} f(y)\ dy$$

for all $x \in \mathbf{R}$, where $\lfloor x \rfloor$ is the greatest integer less than x. This model example is worth keeping in mind for the rest of this discussion.

We have the orthogonal projection identities

$$\int_X f\overline{\mathbf{E}(g|\mathcal{B})}\ d\mu = \int_X \mathbf{E}(f|\mathcal{B})\overline{g}\ d\mu = \int_X \mathbf{E}(f|\mathcal{B})\overline{\mathbf{E}(g|\mathcal{B})}\ d\mu$$

for all $f, g \in L^2(\mathcal{B}_{\max})$. In particular we have

$$\int_X \mathbf{E}(f|\mathcal{B})g\ d\mu = \int_X fg\ d\mu$$

whenever $f \in L^2(\mathcal{B}_{\max})$ and $g \in L^2(\mathcal{B})$. By the self-duality of the Hilbert space $L^2(\mathcal{B})$, this is a full description of the function $\mathbf{E}(f|\mathcal{B}) \in L^2(\mathcal{B})$. From this description we quickly infer that

- If f is real-valued, then so is $\mathbf{E}(f|\mathcal{B})$;
- If f is real and non-negative, then so is $\mathbf{E}(f|\mathcal{B})$;
- If $h \in L^\infty(\mathcal{B})$, then
$$\mathbf{E}(fh|\mathcal{B}) = h\mathbf{E}(f|\mathcal{B}). \tag{3}$$
  In other words, conditional expectation is linear when $L^2(\mathcal{B}_{\max})$ is viewed as a module over $L^\infty(\mathcal{B})$, not just over $\mathbf{C}$.

By linearity, we thus see that $\overline{\mathbf{E}(f|\mathcal{B})} = \mathbf{E}(\overline{f}|\mathcal{B})$ and $\mathbf{E}(\mathrm{Re} f|\mathcal{B}) = \mathrm{Re}\mathbf{E}(f|\mathcal{B})$. Also, if f, g are real-valued and $f \leq g$ pointwise, then $\mathbf{E}(f|\mathcal{B}) \leq \mathbf{E}(g|\mathcal{B})$ pointwise. For real-valued functions this already gives the *triangle inequality* $|\mathbf{E}(f|\mathcal{B})| \leq \mathbf{E}(|f||\mathcal{B})$. For complex-valued functions, a naive splitting into real and imaginary parts will lose a factor of two: $|\mathbf{E}(f|\mathcal{B})| \leq 2\mathbf{E}(|f||\mathcal{B})$. This factor can eventually be recovered by a tensor power trick (try it!) but this is overkill. Instead, we exploit phase

[2] A small subtlety here: the hypothesis that $\mathcal{B}$ is σ-finite does not automatically follow from the hypothesis that $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{B}_{\max}$. Of course this is not an issue when X has finite measure, which is for instance the case in probability theory.

invariance to deduce the complex triangle inequality from the real one. Observe that the complex triangle inequality $|\mathbf{E}(f|\mathcal{B})| \leq \mathbf{E}(|f||\mathcal{B})$ is invariant if we multiply f by a $\mathcal{B}$-measurable phase h (thus $|h| = 1$). Using this (and the $\mathcal{B}$-measurability of $\mathbf{E}(f|\mathcal{B})$, and the module property) we can reduce to the case where $\mathbf{E}(f|\mathcal{B})$ is real and non-negative. But in that case

$$|\mathbf{E}(f|\mathcal{B})| = \mathrm{Re}\mathbf{E}(f|\mathcal{B}) \leq \mathbf{E}(\mathrm{Re} f|\mathcal{B}) \leq \mathbf{E}(|f||\mathcal{B})$$

as desired.

By definition of orthogonal projection we know that the map $f \mapsto \mathbf{E}(f|\mathcal{B})$ is a contraction on $L^2(\mathcal{B}_{\max})$:

$$\|\mathbf{E}(f|\mathcal{B})\|_{L^2(\mathcal{B}_{\max})} \leq \|f\|_{L^2(\mathcal{B}_{\max})}.$$

We can then ask whether the same is true for other L^p:

Proposition 3.2 (Conditional expectation contracts L^p)**.** *If $1 \leq p \leq \infty$ and $f \in L^2(\mathcal{B}_{\max}) \cap L^p(\mathcal{B}_{\max})$, then*

$$\|\mathbf{E}(f|\mathcal{B})\|_{L^p(\mathcal{B}_{\max})} \leq \|f\|_{L^p(\mathcal{B}_{\max})}.$$

Proof By complex interpolation[3] it suffices to check the endpoint cases $p = 1, \infty$. We start with the $p = \infty$ case. For any set E in $\mathcal{B}$ of finite measure and $f \in L^2(\mathcal{B}_{\max}) \cap L^\infty(\mathcal{B}_{\max})$, we have the pointwise bound

$$|\mathbf{E}(f|\mathcal{B})|1_E = |\mathbf{E}(f1_E|\mathcal{B})| \leq \mathbf{E}(|f|1_E|\mathcal{B}) \leq \|f\|_{L^\infty(\mathcal{B}_{\max})}\mathbf{E}(1_E|\mathcal{B}) = \|f\|_{L^\infty(\mathcal{B}_{\max})}1_E$$

and thus (since X is the countable union of sets of finite measure in $\mathcal{B}$) the desired bound follows.

Now take $p = 1$. If $f \in L^1(\mathcal{B}_{\max}) \cap L^2(\mathcal{B}_{\max})$, and $E \in \mathcal{B}$ has finite measure then

$$\begin{aligned}\int_X |\mathbf{E}(f|\mathcal{B})|1_E \, d\mu &\leq \int_X \mathbf{E}(|f||\mathcal{B})1_E \, d\mu \\ &= \int_X |f|1_E \, d\mu \\ &\leq \|f\|_{L^1(\mathcal{B}_{\max})}\end{aligned}$$

and hence by monotone convergence (and the σ-finiteness)

$$\|\mathbf{E}(f|\mathcal{B})\|_{L^1(\mathcal{B})} \leq \|f\|_{L^1(\mathcal{B}_{\max})}.$$

(One can also deduce the $p = 1$ case from the $p = \infty$ case using the self-adjointness of conditional expectation and a duality argument; we omit the details.) ■

For $1 \leq p < \infty$ we know that $L^2(\mathcal{B}_{\max}) \cap L^p(\mathcal{B}_{\max})$ (which contains all simple functions of finite measure support) is dense in L^p, and thus there is a unique continuous extension of conditional expectation to $L^p(\mathcal{B}_{\max})$, which is also a contraction. The same argument works when $p = \infty$ and X has finite measure. When X has infinite measure, the Hahn-Banach theorem gives an infinity of possible extensions to

[3]Real interpolation would also work, but then one needs the tensor product trick to eliminate the constant. This in turn requires some knowledge of product measures, and how conditional expectation reacts to tensor products.

$L^\infty(\mathcal{B}_{\max})$ which are contractions, however there is a unique such extension which still retains the module property (3); we leave the verification of this to the reader.

Remark 3.3. When $p = 2$, $\mathbf{E}(f|\mathcal{B})$ is the closest $\mathcal{B}$-measurable function to f in L^p norm; however we caution that this claim is not true for other values of p.

Let us give an application of this contractibility. Suppose we have an increasing family

$$\mathcal{B}_1 \subset \mathcal{B}_2 \subset \ldots \subset \mathcal{B}_n \subset$$

of σ-finite sub-σ-algebras of $\mathcal{B}_{\max}$. We can then form their limit $\bigvee_{n=1}^\infty \mathcal{B}_n$, defined as the σ-algebra generated by $\bigcup_{n=1}^\infty \mathcal{B}_n$.

Example 3.4. *If $X = \mathbf{R}$, and $\mathcal{B}_n$ is the σ-algebra generated by the dyadic intervals $[k/2^n, (k+1)/2^n)$ for $k \in \mathbf{Z}$, then $\bigvee_{n=1}^\infty \mathcal{B}_n$ is the Borel σ-algebra (why?).*

Proposition 3.5 (Norm convergence of nested expectations)**.** *If $1 \leq p < \infty$ and $f \in L^p(\mathcal{B}_{\max})$, then $\mathbf{E}(f|\mathcal{B}_n)$ converge in L^p norm to $\mathbf{E}(f|\bigvee_{n=1}^\infty \mathcal{B}_n)$.*

Proof From the nesting property we have

$$\mathbf{E}(f|\mathcal{B}_n) = \mathbf{E}(\mathbf{E}(f|\bigvee_{m=1}^\infty \mathcal{B}_m)|\mathcal{B}_n)$$

and so we see that (by replacing f with $\mathbf{E}(f|\bigvee_{m=1}^\infty \mathcal{B}_m)$ if necessary) to prove the proposition it suffices to do so in the case where f is already $\bigvee_{n=1}^\infty \mathcal{B}_n$-measurable.

Observe that the space of sets in $\mathcal{B}_{\max}$ which can be approximated in measure to arbitrary accuracy by a set in $\bigcup_{n=1}^\infty \mathcal{B}_n$ is a σ-algebra and thus contains $\bigvee_{n=1}^\infty \mathcal{B}_n$. In particular, every set in $\bigvee_{n=1}^\infty \mathcal{B}_n$ can be so approximated. This implies that simple functions in $L^p(\bigvee_{n=1}^\infty \mathcal{B}_n)$ can be approximated in L^p norm by simple functions in $L^p(\mathcal{B}_n)$ for some n; since simple functions are themselves dense in L^p, this implies that $\bigcup_{n=1}^\infty L^p(\mathcal{B}_n)$ is dense in $L^p(\bigvee_{n=1}^\infty \mathcal{B}_n)$.

Now for $f \in \bigcup_{n=1}^\infty L^p(\mathcal{B}_n)$ it is clear that $\mathbf{E}(f|\mathcal{B}_n)$ converges to f in L^p norm as $n \to \infty$. We also know that the conditional expectation operators $f \mapsto \mathbf{E}(f|\mathcal{B}_n)$ are contractions on L^p, and in particular are uniformly bounded on L^p. The claim now follows from the previously mentioned density and a standard limiting argument. ■

Remarks 3.6. The claim is false at $p = \infty$; consider for instance the indicator function of $[0, 1/3)$ in Example 3.4. We shall also address the issue of pointwise almost everywhere convergence in next week's notes. The trick of using uniform bounds on operators, together with norm convergence on a dense class of functions, to deduce norm convergence on all functions in the space, is a common[4] one and will be seen several times in these notes.

[4]Indeed, the uniform boundedness principle implies, in some sense, that this is the *only* way to establish a norm convergence result.

4. Norm interchange

Consider a function $f(x, y)$ of two variables. More precisely, consider two measure spaces $X = (X, \mathcal{B}_X, \mu_X)$ and $Y = (Y, \mathcal{B}_Y, \mu_Y)$, and consider a function f on the product space $X \times Y = (X \times Y, \mathcal{B}_X \times \mathcal{B}_Y, \mu_X \times \mu_Y)$. We can consider the mixed norms $L^p_x L^q_y(X \times Y)$ of f for $1 \leq p, q \leq \infty$ by

$$\|f\|_{L^p_x L^q_y(X\times Y)} := \|\|f(x, \cdot)\|_{L^q_y(Y)}\|_{L^p_x(X)}$$

thus for instance when p, q are finite we have

$$\|f\|_{L^p_x L^q_y(X\times Y)} = (\int_X (\int_Y |f(x, y)|^q \; d\mu_Y(y))^{p/q} \; d\mu_X(x))^{1/p}.$$

We can similarly define the interchanged norm $L^q_y L^p_x(X \times Y)$. By iterating the triangle inequality we can see that these are indeed norms. The relationship between the norms is as follows.

Proposition 4.1 (Interchange of norms)**.** *Let the notation and assumptions be as above.*

(i) *When $p = q$, then $\|f\|_{L^p_x L^q_y} = \|f\|_{L^q_y L^p_x} = \|f\|_{L^p_{x,y}(X\times Y)}$.*
(ii) *When $p \geq q$, then $\|f\|_{L^p_x L^q_y} \leq \|f\|_{L^q_y L^p_x}$.*
(iii) *When $p \leq q$, then $\|f\|_{L^p_x L^q_y} \geq \|f\|_{L^q_y L^p_x}$.*

Proof Claim (i) follows from the Fubini-Tonelli theorem (with the case $p = q = \infty$ treated separately). For claim (ii), observe that the claim is already true for $q = p$, and is also true for $q = 1$ by Minkowski. For the intermediate values of p, we can modify the proof of the Riesz-Thorin interpolation theorem (adapted now to mixed norms) to conclude the argument; we leave the details to the reader as an exercise. Claim (iii) of course follows from claim (ii) by reversing p and q. ■

A more direct proof of (ii) is as follows. We can take p, q finite since the cases $p = \infty$ and $q = \infty$ are easy. Raising things to the q^{th} power, it suffices to prove

$$\||f|^q\|_{L^{p/q}_x L^1_y} \leq \||f|^q\|_{L^1_y L^{p/q}_x},$$

but this follows from Minkowski's inequality (since $L^{p/q}_x$ is a Banach space). This proof is shorter, but perhaps a little harder to remember as it relies on an ad hoc trick; in contrast, the interpolation proof, while more complicated, is at least relies on a standard method and is thus easier to remember.

Problem 4.2. Let $1 \leq q < p < \infty$, and suppose that $\|f\|_{L^p_x L^q_y} = \|f\|_{L^q_y L^p_x} < \infty$. Show that $|f|$ is a tensor product, i.e. there exist functions $f_x \in L^p_x$ and $f_y \in L^q_y$ such that $|f(x, y)| = f_x(x) f_y(y)$. Thus we only expect interchanging norms to be a good idea when we believe the worst case example of f in our problem to be roughly tensor product-like in nature.

Let us note two special (and easy) cases of norm interchange. Firstly, that maximal functions control individual functions:

$$\sup_n \|f_n\|_p \leq \|\sup_n |f_n|\|_p.$$

Secondly, we have the triangle inequality

$$\| \sum_n f_n \|_p \leq \sum_n \|f_n\|_p.$$

The moral here is that maximal functions are harder to upper-bound, but are conversely more powerful for dominating other expressions; conversely, it is easier to bound norms of sums than sums of norms, though the latter are good for dominating other things.

Problem 4.3 (Duality for mixed norm spaces). Show that if $1 \leq p, q \leq \infty$ and $f \in L^p_x L^q_y$, then

$$\|f\|_{L^p_x L^q_y} = \sup\{ | \int_X \int_Y f(x,y) \overline{g(x,y)} \, d\mu_X(x) d\mu_Y(y) | : \|g\|_{L^{p'}_x L^{q'}_y} \leq 1 \}.$$

Note that Proposition 4.1 is in some sense "self-dual" with respect to with this duality relationship.

5. Schur's test

We now consider a general class of operators, the *integral operators* $T = T_K$, defined (formally, at least) by

$$Tf(y) = T_K f(y) = \int_X K(x,y) f(x) \, d\mu_X(x) \tag{4}$$

where $K : X \times Y \to \mathbf{C}$ is some measurable function, called the *integral kernel* or simply the *kernel*[5] of the operator T. Of course, one needs some sort of integrability condition[6] on K in order for this operator to make sense, even for simple functions f. For instance, if K is bounded, then Tf makes sense (as a bounded function) for any simple function f of finite measure support; or if X, Y are topological spaces (with μ_X, μ_Y Radon measures) and K is locally integrable, then Tf makes sense (as a locally integrable function) for all continuous, compactly supported functions, basically thanks to the Fubini-Tonelli theorem.

Remark 5.1. If X, Y are finite sets with counting measure, then K is just the matrix associated to the linear transformation T_K.

A fundamental problem in harmonic analysis is to determine conditions on the kernel K which will ensure that T is of strong-type (p,q) (or maybe weak-type or restricted weak-type), and if so, to estimate the constant. In general, these problems are very difficult. However, if one is mapping from L^1 into a Banach space, or mapping from a Banach space into L^∞, the problem becomes much simpler. We illustrate this with the L^p spaces, $1 \leq p \leq \infty$:

[5]Unfortunately, the word "kernel" is also used for the null space of T; in cases where this could lead to confusion the full terminology "integral kernel" is of course recommended instead.

[6]Later on we shall study *singular integrals*, in which K is a distribution which typically does not obey any absolute integrability condition, but nevertheless for which the integral operator can make sense due to cancellation inherent in the kernel.

Proposition 5.2 (Mapping from L^1). *Let $1 \leq q \leq \infty$ and $K : X \times Y \to \mathbf{C}$. Suppose also that $\|K(x,\cdot)\|_{L^q(Y)}$ is uniformly bounded. Then the operator* (4) *is of strong-type* $(1,q)$ *(and absolutely convergent for all $L^1(X)$), and furthermore*

$$\|T_K\|_{L^1(X)\to L^q(Y)} = \sup_{x\in X} \|K(x,\cdot)\|_{L^q(Y)} \tag{5}$$

(as always, supremum is understood to be essential supremum).

Proof For any $f \in L^1(X)$, Minkowski's inequality[7]

$$\begin{aligned} \| \int_X |K(x,y)||f(x)| \, d\mu_X(x)\|_{L^q_y(Y)} &\leq \int_X \| |K(x,y)||f(x)| \, \|_{L^q_y(Y)} \, d\mu_X(x) \\ &\leq \int_X |f(x)| \, d\mu_X(x) \sup_{y\in Y} \| |K(x,y)| \, \|_{L^q_y(Y)} \\ &< \infty \end{aligned}$$

and so in particular we see that $Tf(y)$ is absolutely convergent for almost every y, and from the triangle inequality we also conclude that

$$\|T_K\|_{L^1(X)\to L^q(Y)} \leq \sup_{x\in X} \|K(x,\cdot)\|_{L^q(Y)}.$$

To get the converse inequality, we observe from limiting arguments that we may take K to be a simple function on $X \times Y$; indeed we may go further and take K to be a product simple function, i.e. a linear combination of tensor products of indicator functions on X and Y separately. In such a case, we can partition X into finitely many sets of positive measure, on each of which K is independent of x (excluding sets of measure zero of course). On one of these sets, call it A, the supremum $\|K(x,\cdot)\|_{L^q(Y)}$ is attained. If we then test T_K against $1_{A'}$ for some $A' \subset A$ of finite measure, we obtain the claim. ■

Remark 5.3. If $\|K(x,\cdot)\|_{L^q(Y)}$ is unbounded, then (5) strongly suggests that T is not of strong type $(1,q)$. However proving this rigorously is remarkably difficult, because the lack of even qualitative control on K (other than measurability) defeats the use of most limiting arguments. Nevertheless, for heuristic purposes at least, the above proposition gives a necessary and sufficient condition to map from L^1 to L^q for any $q \geq 1$. A similar argument works with L^q replaced by other Banach spaces, but it unfortunately does *not* work for quasi-normed spaces such as $L^{1,\infty}$; indeed, weak-type $(1,1)$ estimates can be remarkably delicate to establish (or disprove).

There is a dual statement:

Proposition 5.4 (Mapping into L^∞). *Let $1 \leq p \leq \infty$ and $K : X \times Y \to \mathbf{C}$. Suppose also that $\|K(\cdot,y)\|_{L^{p'}(X)}$ is uniformly bounded. Then the operator T_K defined in* (4) *is of strong-type* (p,∞) *(and absolutely convergent for all $f \in L^p(X)$), and furthermore*

$$\|T_K\|_{L^p(X)\to L^\infty(Y)} = \sup_{x\in X} \|K(x,\cdot)\|_{L^{p'}(X)} \tag{6}$$

[7]This can be deduced from the triangle inequality by a monotone convergence argument, approximating $|K|$ and $|f|$ from below by simple functions. The corresponding claim for K and f then follows from a similar dominated convergence argument.

(as always, supremum is understood to be essential supremum).

We leave the proof as an exercise to the reader; it is similar to the previous proposition but relies on Hölder's inequality (and its converse) rather than Minkowski's inequality.

The above propositions give useful necessary and sufficient criteria for strong-type (p,q) bounds when $p=1$ or $q=\infty$. For other cases, it does not seem possible to find such a simple criterion for strong-type (p,q), although when the kernel K is *non-negative* one can at least handle the $p=\infty$ or $q=1$ cases easily:

Problem 5.5. Let $K: X \times Y \to \mathbf{R}^+$ be non-negative and $1 \leq q \leq \infty$. Then T_K is of strong type (∞, q) if and only if $\int_X K(x,y)\ d\mu_X(x)$ lies in $L^q(Y)$, and in fact in this case we have

$$\|T_K\|_{L^\infty(X) \to L^q(Y)} = \| \int_X K(x,y)\ d\mu_X(x) \|_{L^q_y(Y)}.$$

Dually, if $1 \leq p \leq \infty$, then T_K is of strong type $(p,1)$ if and only if $\int_Y K(x,y)\ d\mu_X(y)$ lies in $L^{p'}(X)$, and in fact in this case we have

$$\|T_K\|_{L^p(X) \to L^1(Y)} = \| \int_Y K(x,y)\ d\mu_Y(y) \|_{L^{p'}_y(Y)}.$$

By combining Propositions 5.2, 5.4 with the Riesz-Thorin theorem we can obtain a useful test for L^p boundedness:

Theorem 5.6 (Schur's test)**.** *Suppose that $K: X \times Y \to \mathbf{C}$ obeys the bounds*

$$\int_X |K(x,y)|\ d\mu_X(x) \leq A \text{ for almost every } y \in Y$$

and

$$\int_Y |K(x,y)|\ d\mu_Y(y) \leq B \text{ for almost every } x \in X$$

for some $0 < A, B < \infty$. Then for every $1 \leq p \leq \infty$, the integral operator T_K in (4) *is well-defined (in the sense that the integral is absolutely integrable for almost every y) for all $f \in L^p(X)$, with*

$$\|T_K f\|_{L^p(Y)} \leq A^{1/p'} B^{1/p} \|f\|_{L^p(X)}.$$

Proof The statement of the theorem is invariant[8] under multiplying either μ_X or μ_Y by a positive constant; hence we may normalise $A = B = 1$. By the triangle inequality (replacing K and f with $|K|$ and $|f|$ respectively) it suffices to show the claim when K and f are non-negative. By monotone convergence we may then take f to be a simple function with finite measure support. More monotone convergence then allows us to reduce to the case where X and Y have finite measure. In particular from Fubini's theorem we now know that K is absolutely integrable, and

[8]Note how we are exploiting the *non*-assumption that the domain (X, μ_X) and range (Y, μ_Y) are *not* required to be equal. Thus we see that generalising a proposition may in fact make it easier to prove, as it can introduce more symmetries or other structural features.

so the form $\int_Y T_K f g \, d\mu_Y$ makes sense and is absolutely integrable for all simple f, g.

From the preceding two propositions we know that the claim is true for $p = 1$ and $p = \infty$, and the general case follows from the Riesz-Thorin theorem. (One can also use real interpolation coupled with the tensor power trick.) ■

We can give an alternate proof of Schur's test, which is more elementary (using real convexity rather than complex analyticity) but relies on a non-obvious trick, as follows. Again we normalise $A = B = 1$. As the cases $p = 1, \infty$ are trivial (or can be obtained by limiting arguments) we assume $1 < p < \infty$. By duality and monotone convergence it suffices to show that

$$\int_X \int_Y |K(x,y)||f(x)||g(y)| \, d\mu_X(x) d\mu_Y(y) \leq \|f\|_{L^p(X)} \|g\|_{L^{p'}(Y)}$$

for all simple functions f, g. We can normalise $\|f\|_{L^p(X)} = \|g\|_{L^{p'}(Y)} = 1$. We now use the weighted arithmetic mean-geometric mean inequality

$$|f(x)||g(y)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{p'}|g(y)|^{p'}$$

which if we rewrite it as

$$e^{\frac{1}{p}\log|f(x)|^p + \frac{1}{p'}\log|g(y)|^{p'}} \leq \frac{1}{p} e^{\log|f(x)|^p} + \frac{1}{p'} e^{\log|g(y)|^{p'}}$$

is seen to just be the convexity of the exponential function in disguise. We conclude that

$$\int_X \int_Y |K(x,y)||f(x)||g(y)| \, dx dy \leq \frac{1}{p} \int_X \int_Y |K(x,y)||f(x)|^p \, d\mu_X(x) d\mu_Y(y) + \frac{1}{p'} \int_X \int_Y |K(x,y)||g(y)|^p \, d\mu_X(x) d\mu_Y$$

Computing the first term by integrating in y first, and the second term by integrating in x first, we obtain the claim.

Remark 5.7. A variant on this elementary approach is to factor $|K(x,y)||f(x)||g(y)|$ as $|K(x,y)|^{1/p}|f(x)|$ and $|K(x,y)|^{1/p'}|g(y)|$ and use Hölder's inequality, estimating these two factors in $L^p(X \times Y)$ and $L^{p'}(X \times Y)$ respectively. (We thank Kenley Jung for pointing out this simple argument.)

Problem 5.8. By observing what happens when one multiplies K, μ_X, or μ_Y by positive constants, show that the hypotheses of Schur's test cannot be used to deduce a strong type (p,q) (or even a restricted weak-type (p,q)) bound on T when $p \neq q$.

Remark 5.9. An illustrative case of Schur's test is as follows: let K be an $n \times n$ matrix whose entries are non-negative, and whose row and column sums are all bounded by A. Then the operator norm of K is also bounded by A.

Schur's test can be sharp. Suppose that X, Y have finite measure, that K is non-negative, and that the hypotheses of Schur's test are satisfied exactly, in the sense that

$$\int_X |K(x,y)| \, d\mu_X(x) = A$$

for all $y \in Y$ and

$$\int_Y |K(x,y)|\ d\mu_Y(y) = B$$

for all $x \in X$. Then Fubini's theorem gives the identity

$$A\mu_Y(Y) = B\mu_X(X).$$

Also, we see that

$$T_K 1_X = A 1_Y$$

and so if we set $f = 1_X$ then we easily compute that

$$\|T_K f\|_{L^p(Y)} = A^{1/p'} B^{1/p} \|f\|_{L^p(X)}$$

for all $1 \leq p \leq \infty$, so Schur's test is completely sharp in this case. From this analysis we see that we expect Schur's test to be sharp when (a) there is no oscillation in the kernel K, and (b) K is distributed homogeneously in x and y, in the sense that the marginal integrals $\int_X |K(x,y)|\ d\mu_X(x)$ and $\int_Y |K(x,y)|\ d\mu_Y(y)$ behave like constant functions. When the latter condition is not true, then sometimes a weighted version of Schur's test is more effective (see Q1); alternatively, a decomposition of the operator into components, on each of which Schur's test can be applied, and which can be summed in some efficient manner, is often a better approach.

When K oscillates, one usually does not expect Schur's test to be efficient. For instance, consider the Fourier transform operator $f \mapsto \hat{f}$ on $\mathbf{R}^d$, which is an integral operator with kernel $K(x,y) = e^{2\pi i x \cdot y}$. The values of A and B for this kernel are infinite, yet Plancherel's theorem (which we shall review below) shows that this operator is bounded in $L^2(\mathbf{R}^d)$. Handling oscillatory integrals is in fact a major challenge in harmonic analysis, requiring tools such as almost orthogonality and various wave packet decompositions; we shall return to these issues in later notes.

Remark 5.10. There is in fact a sense in which the weighted version of Schur's test is always sharp for non-negative kernels $K \geq 0$, provided one chooses the weights optimally. Indeed, suppose that $\|T_K\|_{L^p(X) \to L^p(Y)} = A$, and furthermore that this bound is attained by some non-zero f, thus $\|T_K f\|_{L^p(Y)} = A\|f\|_{L^p(X)}$. Since K is non-negative, we can assume without loss of generality that f is also non-negative; we can normalise $\int_X f^p\ d\mu_X = 1$, thus

$$\int_Y (T_K f)^p\ d\mu_Y = A^p.$$

Elementary calculus of variations using Lagrange multipliers then reveals that

$$T_K^*((T_K f)^{p-1}) = \lambda f^{p-1}$$

for some $\lambda \geq 0$; integrating this against f reveals that $\lambda = A^p$. If we write $w = f$ and $v = (T_K f)^{p-1}$ we thus have

$$\int_X K(x,y) w(x)\ d\mu_X(x) = v(y)^{1/(p-1)}$$

and

$$\int_Y K(x,y) v(y)\ d\mu_Y(y) = A^p w(x)^{p-1}.$$

It is then not difficult to formulate a weighted version of Schur's test which gives $\|T_K\|_{L^p(X)\to L^p(Y)} \leq A$. Thus we see that by choosing the weights to be associated to the extremal functions for T, Schur's test does not lose any constants whatsoever.

6. Young's inequality

We have the following variant of Schur's test.

Proposition 6.1 (Generalised fractional integration)**.** *Suppose that $K : X\times Y \to \mathbf{C}$ obeys the bounds*

$$\|K(\cdot, y)\|_{L^{r,\infty}(X)} \lesssim A \text{ for almost every } y \in Y$$

and

$$\|K(x, \cdot)\|_{L^{r,\infty}(Y)} \lesssim A \text{ for almost every } x \in X$$

for some $1 < r < \infty$ and $A > 0$. Then the operator T_K defined in (4) is of weak type $(1, r)$ with norm $O_r(A)$ and of restricted type (r', ∞) with norm $O_r(A)$. For any $1 < p < q < \infty$ and $1 \leq s \leq \infty$ with $\frac{1}{p} + \frac{1}{r} = \frac{1}{q} + 1$, T_K maps $L^{p,s}(X)$ to $L^{q,s}(Y)$ with norm $O_{r,p,q,s}(A)$. In particular, T_K is of strong-type (p, q) with norm $O_{r,p,q}(A)$.

Proof By splitting K into real and imaginary parts, and then into positive and negative parts, we may assume that K is real and non-negative. We may also restrict the functions f we are testing to also be real and non-negative, for similar reasons. By multiplying K by a constant we may normalise $A = 1$.

From the dual characterisation of $L^{r,\infty}(X)$ (Problem 6.9 of last week's notes) we know that the $L^{r,\infty}(X)$ quasi-norm is equivalent to a norm. One can then modify the proof of Proposition 5.2 to obtain the weak type $(1, r)$. A similar argument (modifying Proposition 5.4) gives us the restricted type (r', ∞). The Lorentz claims then follow from Marcinkiewicz interpolation, and the final strong-type claim follows by setting $s = p$ and recalling that $L^{q,p}(Y)$ embeds into L^q. ■

A virtually identical argument (using complex interpolation instead of real) gives

Proposition 6.2 (Generalised Young's inequality)**.** *Let $1 \leq p, q, r \leq \infty$ be such that $\frac{1}{p} + \frac{1}{r} = \frac{1}{q} + 1$, and suppose that $K : X \times Y \to \mathbf{C}$ obeys the bounds*

$$\|K(\cdot, y)\|_{L^r(X)} \leq A \text{ for almost every } y \in Y$$

and

$$\|K(x, \cdot)\|_{L^r(Y)} \leq A \text{ for almost every } x \in X$$

for som $A > 0$. Then the operator T defined in (4) is of strong-type (p, q) with norm at most A.

We leave the proof to the reader as an exercise.

A particular case of Proposition 6.1 is the following inequality, which is of fundamental importance in PDE.

Corollary 6.3 (Hardy-Littlewood-Sobolev fractional integration inequality)**.** *Let* $d \geq 1$, $0 < s < d$, *and* $1 < p < q < \infty$ *be such that*

$$\frac{d}{q} = \frac{d}{p} - s.$$

Then we have

$$\| \int_{\mathbf{R}^d} \frac{f(x)}{|x-y|^{d-s}}\ dx \|_{L^q_y(\mathbf{R}^d)} \lesssim_{d,s,p,q} \|f\|_{L^p_x(\mathbf{R}^d)}$$

for all $f \in L^p_x(\mathbf{R}^d)$.

Indeed, one simply sets $K(x,y) := \frac{1}{|x-y|^{d-s}}$ and $r := \frac{d}{d-s}$ and verifies the hypotheses of Proposition 6.1. The reason for the terminology "fractional integration" will be explained later in this course.

The quantity $\int_{\mathbf{R}^d} \frac{f(x)}{|x-y|^{d-s}}\ dx$ is of course a convolution of f with the convlution kernel $\frac{1}{|x|^{d-s}}$. Convolutions can in fact be studied on more general groups than $\mathbf{R}^d$. Suppose we have a domain X which is both a measure space, $X = (X, \mathcal{B}, \mu)$, and a multiplicative[9] group. We assume that the measure structure and group structure are compatible in the following senses. First, we assume that the group operations $(x,y) \mapsto xy$ and $x \mapsto x^{-1}$ are measurable, and that the translations $x \mapsto xy$, $x \mapsto yx$ and reflection $x \mapsto x^{-1}$ are all measure preserving[10]. We can then define (formally, at least), the convolution $f * g$ of two functions $f, g : X \to \mathbf{C}$ by

$$f * g(x) := \int_X f(y) g(y^{-1}x)\ d\mu(y) = \int_X f(xy^{-1}) g(y)\ d\mu(y).$$

This operation is bilinear and (formally) associative (the association being justified from the Fubini-Tonelli theorem when f, g are either both non-negative or both absolutely integrable), but is only commutative when the underlying group X is also. The convolution algebra is a continuous version of the *group algebra* $\mathbf{C}X$ of X, which corresponds to the case when μ is counting measure and all functions are restricted to have finite support.

From Proposition 6.2 and the hypotheses on μ we quickly obtain

Theorem 6.4 (Young's inequality)**.** *Let* X *be as above, and let* $1 \leq p, q, r \leq \infty$ *be such that* $\frac{1}{p} + \frac{1}{r} = \frac{1}{q} + 1$. *Then for any* $f \in L^p(X)$ *and* $g \in L^r(X)$, *the convolution* $f * g$ *is absolutely convergent for almost every* x *and*

$$\|f * g\|_{L^q(X)} \leq \|f\|_{L^p(X)} \|g\|_{L^r(X)}.$$

Remark 6.5. This looks very similar to Hölder's inequality, but note the "+1" in the scaling condition $\frac{1}{p} + \frac{1}{r} = \frac{1}{q} + 1$ (which, by the way, can be easily seen to

[9] We use multiplication whenever we do not wish to assume that the group is commutative. Of course, many important examples, such as Euclidean space $\mathbf{R}^d$, are additive (and thus commutative) groups.

[10] This is for instance the case if X is a Lie group, with μ a bi-invariant Haar measure. Such measures exist whenever X is unimodular, which in particular occurs when X is compact or abelian. Another example is if μ is counting measure and X is at most countable (for σ-finiteness).

be necessary by the symmetry of multiplying the measure μ by a constant). This shows that convolutions *raise* the integrability exponent, in contrast to products which *lower* it. (This will be consistent with the Hausdorff-Young inequality in the next section.)

Remark 6.6. For some groups, Young's inequality is sharp. For instance, if X contains a measurable subgroup H of positive finite measure, then $1_H * 1_H = \mu(H)1_H$, and one verifies that Young's inequality is sharp for all p, q, r with this example. However, for other groups such as $\mathbf{R}^d$, which do not contain subgroups of positive finite measure, one can lower the constant below 1. We shall establish this in later notes. On the other hand, when f, g are non-negative, Young's inequality at L^1 is an equality:

$$\|f * g\|_1 = \|f\|_1 \|g\|_1.$$

This follows easily from the Fubini-Tonelli theorem.

Young's inequality does not directly imply the Hardy-Littlewood-Sobolev inequality (the problem being that $\frac{1}{|x|^{d-s}}$ lies in weak $L^{d/(d-s)}$ but not strong $L^{d/(d-s)}$). However it does imply the following variant.

Corollary 6.7 (Non-endpoint Hardy-Littlewood-Sobolev). *Let $d \geq 1$, $s < d$, and $1 \leq p \leq q \leq \infty$ be such that*

$$\frac{d}{q} < \frac{d}{p} - s.$$

Then we have

$$\| \int_{\mathbf{R}^d} \frac{f(x)}{\langle x-y \rangle^{d-s}}\ dx \|_{L^q_y(\mathbf{R}^d)} \lesssim_{d,s,p,q} \|f\|_{L^p_x(\mathbf{R}^d)}$$

for all $f \in L^p_x(\mathbf{R}^d)$. Here $\langle x \rangle := (1+|x|^2)^{1/2}$ is the Japanese bracket *of x.*

One corollary of Young's inequality is the norm convergence of approximations to the identity.

Theorem 6.8 (Norm convergence of approximations to the identity). *Let $\phi \in L^1(\mathbf{R}^d)$ be such that $\int_{\mathbf{R}^d} \phi(x)\ dx = 1$, and for each $\varepsilon > 0$ let $\phi_\varepsilon(x) := \varepsilon^{-d}\phi(\frac{x}{\varepsilon})$. Then for any $1 < p \leq \infty$ and $f \in L^p(\mathbf{R}^d)$, the functions $f * \phi_\varepsilon$ converge to f in the $L^p(\mathbf{R}^d)$ topology.*

Proof The key observation here is the quantitative estimate

$$\|f * \phi_\varepsilon\|_{L^p(\mathbf{R}^d)} \leq \|f\|_{L^p(\mathbf{R}^d)} \|\phi_\varepsilon\|_{L^1(\mathbf{R}^d)} = \|f\|_{L^p(\mathbf{R}^d)} \|\phi\|_{L^1(\mathbf{R}^d)}.$$

This quantitative bound lets us make a number of qualitative reductions. First we may restrict ϕ to a dense subclass of $L^1(\mathbf{R}^d)$, such as the test functions $C^\infty_0(\mathbf{R}^d)$. Indeed the general case then follows from a limiting argument taking advantage of the linearity of $f * \phi_\varepsilon$ in ϕ, together with the above bound. A similar argument lets us also restrict f to be in $C^\infty_0(\mathbf{R}^d)$. But then one easily verifies from the identity

$$f * \phi_\varepsilon(x) = \int_{\mathbf{R}^d} f(x-y)\phi(y)\ dy = f(x) + \int_{\mathbf{R}^d} (f(x-y) - f(x))\phi(y)\ dy$$

that $f * \phi_\varepsilon$ converges uniformly to f. Since $f * \phi_\varepsilon$ also vanishes outside a fixed compact set for ε sufficiently small, $f * \phi_\varepsilon$ also converges to f in $L^p(\mathbf{R}^d)$ norm, and we are done. ■

Problem 6.9. Show that the theorem fails for $p = \infty$, but is true when $L^\infty(\mathbf{R}^d)$ is replaced by the closed subspace $C^0(\mathbf{R}^d)$ of bounded continuous functions.

Problem 6.10. If $1 \leq p \leq \infty$, $f \in L^p(\mathbf{R}^d)$, and $g \in L^{p'}(\mathbf{R}^d)$, show that $f * g$ is continuous and decays to zero at infinity.

7. Hausdorff-Young

We will now very quickly introduce the Fourier transform on $\mathbf{R}^d$ in order to demonstrate a classical application of interpolation theory, namely the *Hausdorff-Young inequality*. In this section we shall present this Fourier transform in a rather unmotivated way; much later in this course we shall systematically study the Fourier transform on various groups in a more unified context.

For any $f \in L^1_x(\mathbf{R}^d)$, we define the Fourier transform $\mathcal{F}f = \hat{f} : \mathbf{R}^d \to \mathbf{C}$ by

$$\mathcal{F}f(\xi) = \hat{f}(\xi) := \int_{\mathbf{R}^d} f(x) e^{-2\pi i x \cdot \xi}\ dx.$$

Note that while f is a function of "position" $x \in \mathbf{R}^d$, it is useful to think of $\hat{f}$ as a function of "frequency" $\xi \in \mathbf{R}^d$, which should be kept separate from position[11] It is clear that the Fourier transform $\mathcal{F}$ is a bounded linear operator from $L^1_x(\mathbf{R}^d)$ to $L^\infty_x(\mathbf{R}^d)$. The Fourier transform enjoys the easily verified symmetries

$$\begin{aligned} \mathcal{F}\,\mathrm{Trans}_{x_0} &= \mathrm{Mod}_{-x_0}\,\mathcal{F} \\ \mathcal{F}\,\mathrm{Mod}_{\xi_0} &= \mathrm{Trans}_{\xi_0}\,\mathcal{F} \\ \mathcal{F}\,\mathrm{Dil}^p_\lambda &= \mathrm{Dil}^{p'}_{\lambda^{-1}}\,\mathcal{F} \end{aligned}$$

where Trans_{x_0} for $x_0 \in \mathbf{R}^d$ is the spatial translation operator

$$\mathrm{Trans}_{x_0} f(x) := f(x - x_0),$$

Mod_{ξ_0} for $\xi_0 \in \mathbf{R}^d$ is the frequency modulation operator

$$\mathrm{Mod}_{x_0} f(x) := e^{2\pi i x \cdot \xi_0} f(x),$$

and Dil^p_λ for $\lambda > 0$ and $1 \leq p \leq \infty$ is the L^p-normalised dilation operator

$$\mathrm{Dil}^p_\lambda f(x) := \frac{1}{\lambda^{d/p}} f(\frac{x}{\lambda}).$$

[11] To be truly finicky, the frequency variable should live in the dual space $(\mathbf{R}^d)^*$ of $\mathbf{R}^d$; this is of course canonically identifiable with $\mathbf{R}^d$ once one imposes the Euclidean inner product $x \cdot \xi$ on $\mathbf{R}^d$, but one could also work on a more abstract finite-dimensional vector space without a preferred inner product, in which case no canonical identification is available. It is occasionally useful to work in the latter abstract setting, in order to more easily exploit the $GL(\mathbf{R}^d)$ symmetry, but in many applications the Euclidean structure is already being exploited (e.g. through the use of balls, or of the magnitude function $x \mapsto |x|$), and so there is little point in trying to remove that structure from the Fourier transform.

Thus the Fourier transform interchanges translation with modulation, and reverses dilation. There is in fact a more general symmetry

$$\mathcal{F}\operatorname{Dil}^p_U = \operatorname{Dil}^{p'}_{(U^*)^{-1}} \mathcal{F}$$

for any invertible linear transformation $U : \mathbf{R}^d \to \mathbf{R}^d$ (i.e. $U \in GL(\mathbf{R}^d)$), where

$$\operatorname{Dil}^p_U f(x) := \frac{1}{|\det U|^{1/p}} f(U^{-1}x).$$

Thus, for instance, the Fourier transform commutes with orthogonal transformations such as rotations and reflections. Finally, we observe the commutation relations

$$\mathcal{F}\frac{\partial}{\partial x_j} = 2\pi i \xi_j \mathcal{F}$$

$$\mathcal{F} 2\pi i x_j = -\frac{\partial}{\partial \xi_j}\mathcal{F}$$

which can be viewed as infinitesimal versions of the translation and modulation symmetries. One consequence of these symmetries is the observation that $\mathcal{F}$ maps Schwartz functions to Schwartz functions, continuously in the Schwartz topology.

The adjoint Fourier transform $\mathcal{F}^*$ is clearly given by

$$\mathcal{F}^* g(x) = \check{g}(x) := \int_{\mathbf{R}^d} g(\xi) e^{2\pi i x\cdot\xi}\ dx$$

and is well defined for $g \in L^1(\mathbf{R}^d)$ at least. Being the conjugate of the Fourier transform as well as the adjoint, it obeys essentially the same symmetries:

$$\mathcal{F}^* \operatorname{Trans}_{x_0} = \operatorname{Mod}_{x_0} \mathcal{F}^*$$

$$\mathcal{F}^* \operatorname{Mod}_{-\xi_0} = \operatorname{Trans}_{\xi_0} \mathcal{F}^*$$

$$\mathcal{F}^* \operatorname{Dil}^p_\lambda = \operatorname{Dil}^{p'}_{\lambda^{-1}} \mathcal{F}^*$$

$$\mathcal{F}^* \frac{\partial}{\partial \xi_j} = -2\pi i x_j \mathcal{F}^*$$

$$\mathcal{F}^* 2\pi i \xi_j = \frac{\partial}{\partial x_j}\mathcal{F}^*.$$

In particular the compositions $\mathcal{F}^*\mathcal{F}$ and $\mathcal{F}\mathcal{F}^*$, which maps Schwartz functions with Schwartz functions continuously in the Schwartz topology, commute with translations, modulations, dilations, multiplication by linear functions, and derivatives. One also easily computes that the standard Gaussian $e^{-\pi|x|^2}$ is left unaffected by $\mathcal{F}$ and $\mathcal{F}^*$, and hence by $\mathcal{F}^*\mathcal{F}$ and $\mathcal{F}\mathcal{F}^*$. One then easily deduces that $\mathcal{F}^*\mathcal{F}$ and $\mathcal{F}\mathcal{F}^*$ leaves invariant all polynomial multiples of $e^{-\pi|x|^2}$, which happen to be dense in the Schwartz space (essentially by the Weierstrass approximation theorem). Thus we obtain the *Fourier inversion formula*

$$\mathcal{F}^*\mathcal{F} f = f; \quad \mathcal{F}\mathcal{F}^* g = g$$

for all Schwartz f and g. In other words, $\mathcal{F}$ is formally unitary. As a corollary we obtain the *Parseval identity*[12]

$$\langle f, g\rangle = \langle \mathcal{F}^*\mathcal{F} f, g\rangle = \langle \mathcal{F} f, \mathcal{F} g\rangle$$

[12] Also known as the *Plancherel identity*.

for all Schwartz f, g. Specialising to the diagonal case $f = g$ we obtain the *Plancherel identity*

$$\|\mathcal{F}f\|_{L^2_\xi(\mathbf{R}^d)} = \|f\|_{L^2_x(\mathbf{R}^d)},$$

which allows us by density to extend the Fourier transform uniquely to a continuous linear unitary transformation on $L^2_x(\mathbf{R}^d)$. The transform $\mathcal{F}$ now maps $L^1_x(\mathbf{R}^d)$ to $L^\infty_x(\mathbf{R}^d)$ and $L^2_x(\mathbf{R}^d)$ to $L^2_\xi(\mathbf{R}^d)$, both with norm[13] 1. Riesz-Thorin interpolation then yields the *Hausdorff-Young inequality*

$$\|\mathcal{F}f\|_{L^{p'}_\xi(\mathbf{R}^d)} \leq \|f\|_{L^p_x(\mathbf{R}^d)} \tag{7}$$

for all $1 < p < 2$ and simple functions f with compact support, and (thus by density) we may also extend $\mathcal{F}$ uniquely and continuously to $L^p(\mathbf{R}^d)$. (For $p > 2$, the Fourier transform of an $L^p(\mathbf{R}^d)$ function can still be defined as a distribution, but it need not correspond to a locally integrable function.)

Problem 7.1. Show that up to constant multiplication, the Fourier transform $\mathcal{F}$ is the only continuous map from Schwartz space to itself which obeys the translation and modulation symmetries.

Problem 7.2. Obtain the Hausdorff-Young inequality using real interpolation and the tensor power trick.

The Hausdorff-Young inequality (7) inequality gives an upper bound of 1 for the operator norm $\|\mathcal{F}\|_{L^p_x(\mathbf{R}^d) \to L^{p'}_\xi(\mathbf{R}^d)}$. In the converse direction, we have:

Problem 7.3. Show that if f is a modulated gaussian function, thus $f(x) := ce^{2\pi i x \cdot \xi_0} e^{-A(x-x_0)\cdot(x-x_0)}$ for some positive definite operator $A : \mathbf{R}^d \to \mathbf{R}^d$, some $c > 0$, and $x_0, \xi_0 \in \mathbf{R}^d$, then for all $1 \leq p \leq \infty$ we have

$$\|\mathcal{F}f\|_{L^{p'}_\xi(\mathbf{R}^d)} = \frac{p^{1/2p}}{(p')^{1/2p'}} \|f\|_{L^p_x(\mathbf{R}^d)}.$$

(Hint: use symmetries to reduce to the case of the standard gaussian $f(x) = e^{-\pi|x|^2}$.)

This gives a lower bound of $\frac{p^{1/2p}}{(p')^{1/2p'}}$ for the operator norm. This bound turns out to be sharp (a famous result of Beckner) but we will not demonstrate it here.

One can ask whether the Fourier transform has any other L^p mapping properties, i.e. to determine all the (p, q) for which $\mathcal{F}$ is strong type (or restricted weak type, etc.). The scale invariance of the Fourier transform shows that the duality condition $q = p'$ is necessary for any of these type properties to hold. The Hausdorff-Young inequality also shows that this necessary condition is sufficient (for any of the types) for $p \leq 2$. Unfortunately the inequality fails for $p > 2$. In fact the failure is quite severe and can be demonstrated in a number of ways. One is by *Littlewood's principle*, which asserts that on non-compact domains (such as $\mathbf{R}^d$), it is not possible for an operator with any sort of translation symmetry to map high-exponent Lebesgue spaces to low-exponent spaces. ("The higher exponents are always on the left".) We give one rigorous formulation of this principle in Q12.

[13] This is one reason why we normalise the Fourier transform by placing the 2π in the exponent.

That principle does not directly apply here, but a variant of it will suffice here. Let N be a large integer, let $v \in \mathbf{R}^d$ have large magnitude, and consider the function

$$f(x) := \sum_{n=1}^{N} e^{2\pi i x \cdot nv} g(x - nv)$$

where g is the standard Gaussian $g(x) := e^{-\pi|x|^2}$. Then the symmetries of the Fourier transform show that

$$\hat{f}(\xi) := \sum_{n=1}^{N} e^{-2\pi i \xi \cdot nv} g(\xi - nv) = \overline{f(\xi)}.$$

For any $1 \leq p \leq \infty$, it is not hard to show that

$$\|f\|_{L^p(\mathbf{R}^d)} \sim_p N^{1/p}$$

if N and $|v|$ are sufficiently large depending on p. (Hint: establish the upper bound at $p = 1$ and $p = \infty$, which gives the upper bound for all intermediate p also by log-convexity. The lower bound can be established by considering the value of f near integer multiples of v, or else by computing the L^2 norm by hand and using log-convexity and the upper bounds to then obtain lower bounds elsewhere.) In particular

$$\|\mathcal{F}f\|_{L^{p'}(\mathbf{R}^d)} \sim_p N^{\frac{1}{p'} - \frac{1}{p}} \|f\|_{L^p(\mathbf{R}^d)}.$$

If $p > 2$, then the exponent of N here is positive; since N can be made arbitrarily large, we see that $\mathcal{F}$ cannot be of strong type (p, p') when $p > 2$.

Problem 7.4. Show also that $\mathcal{F}$ cannot be of restricted weak type (p, p') for $p > 2$.

Remark 7.5. The Hausdorff-Young inequality maps spaces at L^2 or below to spaces at L^2 or above. It is very difficult to map back to the original space (in part due to the Littlewood principle) so in practice the Hausdorff-Young inequality is not very useful for the type (p, p) of operators connected to the Fourier transform (such as Fourier multipliers), except of course in the important case $p = 2$. The Hausdorff-Young inequality is also a prototype of a *Fourier restriction estimate* (also known as a *restriction theorem*), of which more will be said later.

8. Kernel truncation

We continue our study of integral operators (4). Intuitively, the smaller one makes the kernel K, the easier it should be to bound it. One basic result in this direction is:

Theorem 8.1 (Positive domination). *Let $K : X \times Y \to \mathbf{C}$, $K' : X \times Y \to \mathbf{R}^+$ be kernels, and let T_K, $T_{K'}$ be the corresponding integral operators. Suppose that we have the pointwise bound $|K| \leq K'$. Then if $T_{K'}$ is of strong type (p, q), then T_K is also (and the integral is a.e. absolutely convergent); in fact we have the bound $\|T_K\|_{L^p(X) \to L^q(Y)} \leq \|T_{K'}\|_{L^p(X) \to L^q(Y)}$.*

Proof From the triangle inequality we have the pointwise bound $|T_K(f)| \leq T_{K'}(|f|)$. The claim follows. ∎

Note that similar results also hold for weak type, restricted type, etc. because of the monotonicity property of the underlying function spaces.

In particular, when K is non-negative, and Ω is any subset of $X \times Y$, we see that T_{K1_Ω} has a smaller (L^p, L^q) operator norm than T_K; thus truncation always reduces the operator norm. However, this property is not quite true for signed kernels, the problem being that the boundedness of T might arise from delicate cancellations within the kernel which might be destroyed by truncation.

Problem 8.2. Let $X = Y = \mathbf{R}$, let K be the Fourier kernel $K(x, y) := e^{-2\pi ixy}$, and let $\Omega \subset \mathbf{R} \times \mathbf{R}$ be those pairs (x, y) for which $e^{-2\pi ixy}$ has positive real part. Show that even though T_K is of strong type $(2, 2)$, the operator T_{K1_Ω} is not. (Hint: compute a lower bound for the real part of $\langle T_{K1_\Omega} 1_I, 1_I \rangle$ for a large interval I.)

On the other hand, certain truncations are acceptable for signed kernels. For instance, if $p = 1$ or $q = \infty$ then all truncations contract the operator norm, thanks to Propositions 5.2, 5.4. Secondly, it is clear that if Ω is a product set $\Omega = A \times B$, then from the[14] identity

$$T_{K1_\Omega}(f) = 1_B T_K(1_A f) \tag{8}$$

we see that if T_K is of strong-type (p, q) then T_{K1_Ω} is also, with the expected comparison principle $\|T_{K1_\Omega}\|_{L^p(X)\to L^q(Y)} \leq \|T_K\|_{L^p(X)\to L^q(Y)}$. Also note in this case that T_{K1_Ω} can also be viewed as an operator from $L^p(A)$ to $L^p(B)$, and

$$\|T_{K1_\Omega}\|_{L^p(X)\to L^q(Y)} = \|T_{K1_\Omega}\|_{L^p(A)\to L^q(Y)} = \|T_{K1_\Omega}\|_{L^p(A)\to L^q(B)}.$$

Let us write the identity (8) in operator form as

$$T_{K1_\Omega} = 1_B T_K 1_A$$

where we now think of 1_A, 1_B not as functions, but rather as the multiplier operators $f \mapsto 1_A f$, $f \mapsto 1_B f$. Motivated by this identity, let us now study other ways in which an operator T can be multiplied on the left and right by indicator functions.

Problem 8.3 (Block diagonal operators). Suppose Q is at most countable, and we have disjoint subsets $(X_n)_{n\in Q}$, $(Y_n)_{n\in Q}$ of X and Y respectively. For each $n \in Q$, let $T_n : L^p(X_n) \to L^q(Y_n)$ be a bounded linear operator for some $1 \leq p, q \leq \infty$, and define the block diagonal[15]

$$T := \sum_{n\in Q} 1_{Y_n} T_n 1_{X_n}.$$

Here we identify functions on X_n (resp. Y_n) with functions on X (resp. Y) by extending by zero.

[14]Here we are implicitly assuming that there is enough regularity in K and f that the expressions are well-defined; we gloss over this subtlety here.

[15]Strictly speaking, "block diagonal" refers to the case when $X = Y$ and $X_n = Y_n$. A more appropriate term here would be "block permutation", but we prefer block diagonal as it is more familiar (and covers the most important cases).

- If $q \geq p$, show that
$$\|T\|_{L^p(X)\to L^q(Y)} = \sup_n \|T_n\|_{L^p(X_n)\to L^q(Y_n)}.$$
- If $q < p$, show that
$$\|T\|_{L^p(X)\to L^q(Y)} = \|\|T_n\|_{L^p(X_n)\to L^q(Y_n)}\|_{l^r_n}$$
where r is such that $\frac{1}{p} + \frac{1}{r} = \frac{1}{q}$.

Thus the behaviour of the block-diagonal operator T is determined entirely by the behaviour of the components T_n. In particular, if $q \geq p$, T is of strong type (p,q) if and only if the T_n are uniformly of strong type (p,q).

Corollary 8.4 (Truncation to a block diagonal). *Let Q, $(X_n)_{n\in Q}$, $(Y_n)_{n\in Q}$ be as above, and let $T : L^p(X) \to L^q(Y)$ be a bounded linear operator for some $1 \leq p \leq q \leq \infty$. Let T_Δ be the block-diagonal restriction of T,*
$$T_\Delta := \sum_{n\in Q} 1_{Y_n} T 1_{X_n}.$$
Then T_Δ is also of type (p,q) and $\|T_\Delta\|_{L^p(X)\to L^q(Y)} \leq \|T\|_{L^p(X)\to L^q(Y)}$.

Proof Apply the above problem with $T_n := 1_{Y_n} T 1_{X_n}$. ■

Thus we see that we may freely restrict the kernel of an integral operator T_K to a block-diagonal set such as $\bigcup_n X_n \times Y_n$ without increasing the operator norm, as long as we respect Littlewood's principle (keeping the higher exponent on the left).

A similar argument works for "almost block diagonal" operators[16]:

Problem 8.5 (Almost block diagonal operators). Let $(X_n)_{n\in Q}$, $(Y_m)_{m\in Q'}$ be disjoint subsets of X and Y respectively, where Q, Q' are at most countable. Let $\sim$ be a relation between Q and Q' which has "bounded degree" in the sense that for each $n \in Q$ there are at most $O(1)$ values of $m \in Q'$ for which $n \sim m$, and conversely for each $m \in Q'$ there are at most $O(1)$ values of $n \in Q$ for which $n \sim m$. Let $1 \leq p \leq q \leq \infty$, and for each (n,m) with $n \sim m$, let $T_{n,m} : L^p(X_n) \to L^q(Y_m)$ be a bounded linear operator, and let T be the sum
$$T := \sum_{(n,m)\in Q\times Q' : n\sim m} 1_{Y_m} T_{n,m} 1_{X_n}.$$
(Note that the bounded degree hypothesis ensures that this sum is well-defined.) Then for any $A > 0$ the following are equivalent up to changes in the implied constants.

- $\|T\|_{L^p(X)\to L^q(Y)} \lesssim_{p,q} A$.
- $\|T_{n,m}\|_{L^p(X_n)\to L^q(Y_m)} \lesssim_{p,q} A$ for all n, m with $n \sim m$.

[16]Again, a better term here would be "almost block permutation" operators.

Remark 8.6. The point about block-diagonal or almost block-diagonal operators are that they are *local*; the influence of the subdomain X_n on the operator T is limited to a bounded number of subranges Y_m. Thus we see that for local operators, global boundedness and local boundedness are essentially equivalent. One can now obtain an analogue of Corollary 8.4, namely that one can restrict a kernel to an almost block diagonal region $\bigcup_{n \sim m} X_n \times Y_m$ without increasing the $(L^p(X), L^q(Y))$ norm by more than a multiplicative constant; we leave the details to the reader.

Having considered truncations to "diagonal" regions, let us now consider truncations to "upper diagonal" regions. Here it turns out we need to strengthen the Littlewood condition $p \leq q$ to a strict inequality $p < q$ to obtain good results (otherwise we lose a logarithm, see Q14).

Theorem 8.7 (Christ-Kiselev lemma, finite version)**.** *Let $N \geq 1$, and suppose we have partitions $X = \bigcup_{n=1}^N X_n$, $Y = \bigcup_{n=1}^N Y_n$. Let $1 \leq p < q \leq \infty$, and let $T : L^p(X) \to L^q(Y)$ be a bounded linear operator. Let T' be the modified operator*

$$T' := \sum_{1 \leq n \leq m \leq N} 1_{Y_m} T 1_{X_n}.$$

Then T' is also bounded from $L^p(X)$ to $L^q(Y)$, and we have

$$\|T'\|_{L^p(X) \to L^q(Y)} \lesssim_{p,q} \|T\|_{L^p(X) \to L^q(Y)}.$$

Proof We will prove this by induction on N; however, one must be careful because of the use of the Vinogradov notation $\lesssim_{p,q}$ in the conclusion. To do things rigorously, let $A_{p,q}$ be a large constant depending only on p and q to be chosen later. We claim that for every N, we have the bound

$$\|T'\|_{L^p(X) \to L^q(Y)} \leq A_{p,q} \|T\|_{L^p(X) \to L^q(Y)}.$$

The claim is trivial for $N = 1$ (if $A_{p,q}$ is large enough) so suppose inductively that $N > 1$ and the claim has already been proven for smaller values of N. We may normalise $\|T\|_{L^p(X) \to L^q(Y)} = 1$. Let us also choose $f \in L^p(X)$, normalised so that $\|f\|_{L^p(X)} = 1$. Our task is to show that

$$\|T' f\|_{L^q(Y)} \leq A_{p,q}.$$

The idea is to divide-and-conquer f in an intelligent fashion, adapted to both f and the blocks $X_1, \dots, X_n$. Consider the sequence of numbers $\|f 1_{X_1 \cup \ldots \cup X_n}\|_{L^p(X)}^p$ for $n = 0, 1, \dots, N$. This sequence increases from 0 to 1, so we can find $0 \leq n_0 < N$ such that

$$\|f 1_{X_1 \cup \ldots \cup X_{n_0}}\|_{L^p(X)}^p \leq \frac{1}{2} < \|f 1_{X_1 \cup \ldots \cup X_{n_0+1}}\|_{L^p(X)}^p$$

and hence

$$\|f 1_{X_1 \cup \ldots \cup X_{n_0}}\|_{L^p(X)}, \|f 1_{X_{n_0+2} \cup \ldots \cup X_N}\|_{L^p(X)} \leq \frac{1}{2^{1/p}}.$$

Applying the induction hypothesis, we conclude that

$$\|T'(f 1_{X_1 \cup \ldots \cup X_{n_0}})\|_{L^q(Y_1 \cup \ldots \cup Y_{n_0})}, \|T'(f 1_{X_{n_0+2} \cup \ldots \cup X_N})\|_{L^q(Y_{n_0+2} \cup \ldots \cup Y_N)} \leq \frac{1}{2^{1/p}} A_{p,q}.$$

On the other hand, we have

$$\|T'(f 1_{X_1 \cup \ldots \cup X_{n_0}})\|_{L^q(Y_{n_0+1} \cup \ldots \cup Y_N)} = \|T(f 1_{X_1 \cup \ldots \cup X_{n_0}})\|_{L^q(Y_{n_0+1} \cup \ldots \cup Y_N)} = O(1)$$

and

$$\|T'(f1_{X_{n_0+1}})\|_{L^q(Y)} \leq \|T(f1_{X_{n_0+1}})\|_{L^q(Y)} = O(1)$$

by the normalisations on T and f and the definition of T'. Also, $T'(f1_{X_{n_0+2}\cup\ldots\cup X_N})$ vanishes outside of $Y_{n_0+2}\cup\ldots\cup Y_N$. From the triangle inequality we conclude that

$$\|T'f\|_{L^q(Y_1\cup\ldots\cup Y_{n_0})}, \|T'f\|_{L^q(Y_{n_0+1}\cup\ldots\cup Y_N)} \leq \frac{1}{2^{1/p}}A_{p,q} + O(1)$$

and thus

$$\|T'f\|_{L^q(Y)} \leq \frac{2^{1/q}}{2^{1/p}}A_{p,q} + O(1).$$

If we choose $A_{p,q}$ large enough depending on p, q, then we can make the right-hand side less than $A_{p,q}$ (here we use the hypothesis $q > p$). The claim follows. ∎

In terms of kernels, the above lemma allows one to truncate the kernel to upper-diagonal regions $\bigcup_{n\leq m} X_n \times Y_m$ without losing too much in the operator norm, so long as $q > p$. The claim is unfortunately false when $p = q$ (except when $p = q = 1$ or $p = q = \infty$, where the claim is easy to verify), although producing an example will have to wait for our analysis of the Hilbert transform in next week's notes.

The fact that the partitions $X = \bigcup_n X_n$, $Y = \bigcup_n Y_n$ are arbitrary makes the Christ-Kiselev lemma quite powerful. Indeed, we can conclude

Corollary 8.8 (Christ-Kiselev lemma, maximal version)**.** *Let Q be a countable ordered set, and for each $\alpha \in Q$ let E_α be a set in X with the nesting property $E_\alpha \subset E_{\alpha'}$ whenever $\alpha < \alpha'$. Let $1 \leq p < q \leq \infty$, and let $T : L^p(X) \to L^q(Y)$ be a bounded linear operator. Define the maximal operator*

$$T_*f(y) := \sup_{\alpha\in Q} |T(f1_{E_\alpha})(y)|. \tag{9}$$

Then the sublinear operator T_* is also bounded from $L^p(X)$ to $L^q(Y)$, and in fact

$$\|T_*\|_{L^p(X)\to L^q(Y)} \lesssim_{p,q} \|T\|_{L^p(X)\to L^q(Y)}.$$

Proof We may normalise $\|T\|_{L^p(X)\to L^q(Y)} = \|f\|_{L^p(X)} = 1$. By monotone convergence we may assume that Q is finite, and by relabeling we can then take $Q = \{1, \ldots, N\}$. By adding dummy indices if necessary we may assume $E_N = X$, and we also adopt the convention $E_0 = \emptyset$. Then for each $y \in Y$ there is an $1 \leq \alpha(y) \leq N$ which attains the supremum in (9):

$$T_*f(y) = |T(f1_{E_{\alpha(y)}})(y)|,$$

and so we need to show that

$$\|T(f1_{E_{\alpha(y)}})\|_{L^q_y(Y)} \lesssim_{p,q} 1.$$

(This trick of replacing a maximal operator T_* by a linear operator $f \mapsto T(f1_{E_{\alpha(y)}})$, albeit an operator which depends on an arbitrary function α, is known as *linearisation.*) Now for each $1 \leq n \leq N$, let $X_n := E_n\backslash E_{n-1}$ and $Y_n := \{y \in Y : \alpha(y) = n\}$. Then we observe the telescoping identity

$$T(f1_{E_{\alpha(y)}}) = \sum_{1\leq n\leq m\leq N} 1_{Y_m}T(1_{X_n}f).$$

The claim now follows from Lemma 8.7. ∎

Let us give a sample application of this lemma. Given any locally integrable function $f : \mathbf{R} \to \mathbf{C}$, define the *maximal Fourier transform* $\mathcal{F}_* f$ by

$$\mathcal{F}_* f(\xi) := \sup_I |\int_I f(x) e^{-2\pi i x\xi}\ dx| = \sup_I \widehat{f 1_I}(\xi)$$

where I ranges over compact intervals in $\mathbf{R}$.

Corollary 8.9 (Menshov-Paley-Zygmund theorem, quantitative version)**.** *For any $1 \leq p < 2$ we have*

$$\|\mathcal{F}_*\|_{L^p(\mathbf{R}) \to L^{p'}(\mathbf{R})} \lesssim_p 1.$$

Proof By the triangle inequality we may restrict I to be an interval with one endpoint 0; by symmetry we may take I to have the form $[0, y]$ for some positive y. By monotone convergence we may take y to be rational. Thus we wish to show that

$$\| \sup_{y \in \mathbf{Q}^+} \widehat{f 1_{[0,y]}} \|_{L^{p'}(\mathbf{R})} \lesssim_p \|f\|_{L^p(\mathbf{R})}.$$

But this follows from the Hausdorff-Young inequality and Corollary 8.8. ■

Corollary 8.10 (Menshov-Paley-Zygmund theorem, qualitative version)**.** *If $1 \leq p < 2$ and $f \in L^p(\mathbf{R})$, then for almost every $\xi \in \mathbf{R}$ the Fourier integral $\int_{\mathbf{R}} f(x) e^{-2\pi i x \xi}\ dx$ is conditionally convergent to $\hat f(\xi)$, thus*

$$\lim_{N_+, N_- \to +\infty} \int_{-N_-}^{N_+} f(x) e^{-2\pi i x\xi}\ dx = \hat f(\xi)$$

for almost every ξ.

Proof We can rewrite the claim as

$$\| \limsup_{N_+, N_- \to +\infty} | \int_{-N_-}^{N_+} f(x) e^{-2\pi i x\xi}\ dx - \hat f(\xi)| \|_{L^{p'}_\xi(\mathbf{R})} = 0$$

The claim is trivial if f lies in $L^1(\mathbf{R}) \cap L^p(\mathbf{R})$, which is a dense subset of $L^p(\mathbf{R})$. The claim then follows from a limiting argument and Corollary 8.9 (and the Hausdorff-Young inequality). ■

Remark 8.11. The above results are also true for $p = 2$, but is much more difficult to prove, and is known as *Carleson's theorem.*

One can pass from the finite version of the Christ-Kiselev lemma to an infinite version if one has enough regularity on the kernel. For instance, we have

Problem 8.12 (Christ-Kiselev lemma, upper diagonal version). Let $K : \mathbf{R} \times \mathbf{R} \to \mathbf{C}$ be a locally integrable kernel, and suppose that the operator T_K (which is defined and locally integrable for compactly supported continuous functions $f \in C^0_c(\mathbf{R})$ at least) is bounded from $L^p(\mathbf{R})$ to $L^q(\mathbf{R})$ for some $1 \leq p < q \leq \infty$ (i.e. it is bounded on the dense subclass $C^0_c(\mathbf{R})$ and thus has a unique continuous extension). Let $K_>$ be the restriction of K to the upper diagonal $\{(s,t) \in \mathbf{R} \times \mathbf{R} : t > s\}$. Then the operator $T_{K_>}$ (which is also defined on $C^0_c(\mathbf{R})$) is also of strong type (p, q), with

$$\|T_{K_>}\|_{L^p(\mathbf{R}) \to L^q(\mathbf{R})} \lesssim_{p,q} \|T_K\|_{L^p(\mathbf{R}) \to L^q(\mathbf{R})}.$$

Remark 8.13. A "vector-valued" version of this lemma, in which the input and output functions are not complex values, but instead take values in some infinite-dimensional vector space, is very useful in the study of nonlinear dispersive PDE, and in particular in the theory of Strichartz estimates, but we will not discuss that topic further here.

In the above results we have only discussed "rough" truncations, in which a kernel K was multiplied by an indicator function 1_Ω. One can do somewhat better with "smooth" truncations, but this requires significantly more structure on the underlying domain.

Definition 8.14 (Adapted bump function)**.** Let $\Omega \subset \mathbf{R}^d$ be a bounded open set, and let $L : \mathbf{R}^d \to \mathbf{R}^d$ be an invertible affine transformation (the composition of a general linear transformation and a translation), thus $L(\Omega)$ is another bounded open set in $\mathbf{R}^d$. We say that a function $\phi : \mathbf{R}^d \to \mathbf{C}$ is a *bump function adapted to* $L(\Omega)$ if ϕ is smooth, supported in $L(\Omega)$, and we have the bounds

$$\nabla^k(\phi \circ L)(x) = O_{k,\Omega}(1)$$

for all $k \geq 0$ and $x \in \mathbf{R}^d$ (actually one just needs $x \in \Omega$).

Remark 8.15. This definition only becomes meaningful if we hold Ω fixed and let L vary. Typically Ω is a standard set, such as the unit ball, unit cube, unit cylinder, or unit annulus, so that $L(\Omega)$ describes balls, cubes, boxes, tubes, or annuli of various radius, eccentricity, and location; the point is then that the bump functions adapted to these sets are adapted uniformly in these additional parameters.

Proposition 8.16 (Smooth cutoffs)**.** *Let $K : \mathbf{R}^d \times \mathbf{R}^{d'} \to \mathbf{C}$ be a locally integrable function, and suppose that T_K is bounded from $L^p(\mathbf{R}^d)$ to $L^q(\mathbf{R}^{d'})$ for some $1 \leq p, q \leq \infty$. Let $L(\Omega) \subset \mathbf{R}^d$ and $L'(\Omega') \subset \mathbf{R}^{d'}$ be affine images of bounded open sets $\Omega \subset \mathbf{R}^d$, $\Omega' \subset \mathbf{R}^d$, and let $\phi : \mathbf{R}^d \times \mathbf{R}^{d'} \to \mathbf{C}$ be a bump function adapted to $L(\Omega) \times L'(\Omega')$. Then $T_{K\phi}$ is bounded from $L^p(\mathbf{R}^d)$ to $L^q(\mathbf{R}^{d'})$, and*

$$\|T_{K\phi}\|_{L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^{d'})} \lesssim_{\Omega,\Omega',d,d'} \|T_K\|_{L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^{d'})}$$

where we define $T_{K\phi}$ first for test functions $C^0_c(\mathbf{R}^d)$ and then extend by density.

Proof The idea is to exploit the fact that while bump functions adapted to product sets such as $L(\Omega) \times L'(\Omega')$ are not necessarily tensor products of bump functions adapted to $L(\Omega)$ and $L'(\Omega')$ separately, they can be efficiently decomposed as a convergent sum of such.

By rescaling by L and L' we may assume that these transformations are the identity. By rescaling a little more to shrink Ω and Ω' we may assume that these sets are supported in (say) the cubes $[-1/4, 1/4]^d$ and $[-1/4, 1/4]^{d'}$ respectively. In particular, ϕ is now a smooth bump function supported on $[-1/4, 1/4]^d \times [-1/4, 1/4]^{d'}$. We can identify this set with a subset of the torus $(\mathbf{R}/\mathbf{Z})^d \times (\mathbf{R}/\mathbf{Z})^{d'}$ in the usual manner, thus creating a function $\tilde{\phi} : (\mathbf{R}/\mathbf{Z})^d \times (\mathbf{R}/\mathbf{Z})^{d'} \to \mathbf{C}$ whose k^{th} derivatives are $O_{k,\Omega,\Omega'}(1)$. We now appeal to the theory of Fourier series to write

$$\tilde{\phi}(x, y) = \sum_{n \in \mathbf{Z}^d} \sum_{m \in \mathbf{Z}^{d'}} \hat{\tilde{\phi}}(n, m) e^{2\pi i n \cdot x} e^{2\pi i m \cdot y} \qquad (10)$$

where $\hat{\tilde{\phi}}(n,m)$ is the Fourier coefficient

$$\hat{\tilde{\phi}}(n,m) = \int_{(\mathbf{R}/\mathbf{Z})^d \times (\mathbf{R}/\mathbf{Z})^{d'}} \tilde{\phi}(x,y) e^{-2\pi i n \cdot x} e^{-2\pi i m \cdot y}\ dxdy.$$

The smoothness of $\tilde{\phi}$, combined with a standard integration by parts argument, shows that $\hat{\tilde{\phi}}(n,m)$ decays rapidly, thus for any k we have

$$\hat{\tilde{\phi}}(n,m) = O_{k,\Omega,\Omega'}((1+|n|+|m|)^{-k}).$$

In particular (taking k larger than $d+d'$) we obtain that these coefficients are summable:

$$\sum_{n \in \mathbf{Z}^d} \sum_{m \in \mathbf{Z}^{d'}} |\hat{\tilde{\phi}}(n,m)| \lesssim_{d,d',\Omega,\Omega'} 1.$$

Thus the summation in (10) is in fact absolutely convergent. This leads to the kernel identity

$$\phi(x,y)K(x,y) = \sum_{n \in \mathbf{Z}^d} \sum_{m \in \mathbf{Z}^{d'}} \hat{\tilde{\phi}}(n,m) e^{2\pi i n \cdot x} 1_{[-1/4,1/4]^d}(x) e^{2\pi i m \cdot y} 1_{[-1/4,1/4]^{d'}}(y) K(x,y)$$

which in turn leads to the operator identity

$$T_{K\phi} = \sum_{n \in \mathbf{Z}^d} \sum_{m \in \mathbf{Z}^{d'}} \hat{\tilde{\phi}}(n,m) 1_{[-1/4,1/4]^{d'}} e_m T_K 1_{[-1/4,1/4]^d} e_n$$

(at least when applied to test functions), where e_m is the multiplier operator $f(y) \mapsto f(y) e^{2\pi i m \cdot y}$, and similarly for e_n. Since the operators $1_{[-1/4,1/4]^d}$, e_n are contractions on $L^p(\mathbf{R}^d)$, and $1_{[-1/4,1/4]^{d'}}$, e_m are contractions on $L^q(\mathbf{R}^d)$, the claim now follows from the triangle inequality ■

One of the morals of the above proposition is that smooth bump functions, when applied to an integral kernel, have essentially a negligible impact on the behaviour of the integral operator, other than to localise that operator to the support of that bump function. In particular the precise choice of bump function used is almost always irrelevant. The boundedness of an operator is instead controlled by those features unaffected by multiplying by smooth bump functions, such as singularity or rapid oscillation.

9. Exercises

- Q1 (Weighted Schur's test) Suppose that $K : X \times Y \to \mathbf{C}$ obeys the bounds

  $$\int_X |K(x,y)| w_X(x)\ d\mu_X(x) \leq A w_Y(y) \text{ for almost every } y \in Y$$

  and

  $$\int_Y |K(x,y)| w_Y(x)\ d\mu_Y(y) \leq B w_X(x) \text{ for almost every } x \in X$$

  for some $0 < A, B < \infty$ and some strictly positive weight functions $w_X : X \to \mathbf{R}^+$, $w_Y \to \mathbf{R}^+$. Show that the integral operator (4) is bounded from $L^2(X)$ to $L^2(Y)$ with an operator norm bound $\|T\|_{L^2(X) \to L^2(Y)} \leq \sqrt{AB}$.

- Q2 (Hilbert-Schmidt norm) Suppose that $K : X \times Y \to \mathbf{C}$ is square-integrable. Show that the integral operator (4) is bounded from $L^2(X)$ to $L^2(Y)$ with operator norm bound
$$\|T\|_{L^2(X) \to L^2(Y)} \leq \|K\|_{L^2(X \times Y)}.$$
Show that equality occurs if and only if T is rank one, thus $K(x,y) = f(x)g(y)$ for some $f \in L^2(X)$, $g = L^2(Y)$. The quantity $\|K\|_{L^2(X\times Y)}$ is known as the *Hilbert-Schmidt norm* or *Frobenius norm* of T. From the above analysis of equality, we only expect the above estimate to be efficient when T is "approximately low rank" (i.e. it has some good compactness properties; more informally, this estimate is only efficient when T is only sensitive to a few degrees of freedom of the input function).
- Q3. Let X be a measure space which is *non-atomic* in the sense that given any set $E \subset X$ of positive measure, there exists a subset E' with $0 < \mu(E') < \mu(E)$. (The real line $\mathbf{R}$ is a typical example.)
  (a) Show that if E has positive measure and $0 < \alpha < \mu(E)$, then there exists a subset E' of E with $\mu(E') = \alpha$. (Feel free to use Zorn's lemma, though it is possible to establish this result using only the axiom of countable choice instead. You may wish to first establish the weaker result that there exists a subset E' with $0 < \mu(E') \leq \alpha$.)
  (b) Suppose that T is a linear operator mapping simple functions on X to functions on Y which is of restricted weak-type (p,q) for some $1 \leq q < \infty$ and $0 < p < 1$. Show that $T = 0$. Indeed, it is very rare to see any non-trivial boundedness statement originating from below the L^1 norm, especially given that such functions are usually not even locally integrable.
- Q4. Show by means of examples (or by some sort of dimensional analysis or scaling argument) that Corollary 6.3 fails when $p = 1$, when $q = \infty$, or when the condition $\frac{d}{p} = \frac{d}{q} - s$ is violated.
- Q5. (Hardy inequality, I.) Let $\mathbf{R}^+$ be given Lebesgue measure dx. Establish the one-dimensional *Hardy inequality*
$$\|\frac{1}{x}\int_0^x f(t)\ dt\|_{L^p_x(\mathbf{R}^+)} \lesssim_p \|f\|_{L^p(\mathbf{R}^+)}.$$
for all $1 < p \leq \infty$ and $f \in L^p(\mathbf{R}^+)$, and show that the inequality fails at $p = 1$. (Hint: first establish restricted weak type for $1 < p < \infty$.)
- Q6. (Hardy inequality, II.) Show that
$$\|\frac{1}{x}\int_0^x f(t)\ dt\|_{L^2_x(\mathbf{R}^+)} \leq 2\|f\|_{L^2(\mathbf{R}^+)}$$
for all $f \in L^2_x(\mathbf{R}^+)$, and that the constant 2 cannot be replaced by any smaller constant. (Hint: there are many ways to prove this; one is by Q1 with some perturbation of the weight $|x|^{-1/2}$.)
- Q7. (Hardy inequality, III.) Let $1 < p < \infty$. Show that
$$\|\frac{1}{x}\int_0^x f(t)\ dt\|_{L^p_x(\mathbf{R}^+)} \leq \frac{p}{p-1}\|f\|_{L^p(\mathbf{R}^+)}$$
for all $f \in L^p_x(\mathbf{R}^+)$, and that the constant $\frac{p}{p-1}$ cannot be replaced by any smaller constant. (You may need to develop an L^p version of Q1.)

- Q8. Let $d \geq 1$, $1 \leq p, q \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} \geq 1$, and let $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$. Let $\alpha, \beta \in \mathbf{R}$ obey the scaling condition
$$\alpha + \beta = -\frac{d}{p'} - \frac{d}{q'}.$$
  - If $\alpha > -\frac{n}{p'}$ (thus $\beta < -\frac{n}{q'}$), show that
$$\int\int_{|x| \lesssim |y|} |x|^\alpha |y|^\beta |f(x)||g(y)|\ dxdy \lesssim_{\alpha,\beta,p,q} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$
  - If $\alpha < -\frac{n}{p'}$ (thus $\beta > -\frac{n}{q'}$), show that
$$\int\int_{|y| \lesssim |x|} |x|^\alpha |y|^\beta |f(x)||g(y)|\ dxdy \lesssim_{\alpha,\beta,p,q} \|f\|_{L^p(\mathbf{R}^d)} \|g\|_{L^q(\mathbf{R}^d)}.$$
- Q9. (Weighted one-dimensional Hardy-Littlewood-Sobolev) Let $1 \leq p, q \leq \infty$, $0 < s < 1$, and $\alpha, \beta \in \mathbf{R}$ obey the conditions
$$\alpha > -\frac{1}{p'}$$
$$\beta > -\frac{1}{q'}$$
$$1 \leq \frac{1}{p} + \frac{1}{q} \leq 1 + s$$
and the scaling condition
$$\alpha + \beta - 1 + s = -\frac{n}{p'} - \frac{n}{q'},$$
with at most one of the equalities
$$p = 1, p = \infty, q = 1, q = \infty, \frac{1}{p} + \frac{1}{q} = 1 + s$$
holding. Let $f \in L^p(\mathbf{R})$ and $g \in L^q(\mathbf{R})$. Show that
$$\int_{\mathbf{R}}\int_{\mathbf{R}} \frac{|x|^\alpha |y|^\beta}{|x-y|^{1-s}} |f(x)||g(y)|\ dxdy \lesssim_{\alpha,\beta,p,q} \|f\|_{L^p(\mathbf{R})} \|g\|_{L^q(\mathbf{R})}.$$
- Q10. Let X be a measure space with $\mu(X) = 1$. Let T be a linear operator mapping simple functions on X to some normed vector space W. Suppose we have the bound
$$\|T1_E\|_W \lesssim A|E|(1 + \log \frac{1}{|E|})$$
for all sets $E \subset X$ and some $A > 0$. Conclude that
$$\|Tf\|_W \lesssim A\|f\|_{L \log L}$$
for all simple f. (Hint: first do this for quasi-step simple functions, and then perform a decomposition similar to that in the previous week's notes.) Thus in some sense $L \log L$ and "restricted $L \log L$" are equivalent.
- Q11 (Yano extrapolation theorem). Let X be a measure space with $\mu(X) = 1$. Let T be a linear operator defined on simple functions which is of strong type (p, p) for all $1 < p \leq 2$. Furthermore, assume that
$$\|Tf\|_{L^p(X)} \lesssim \frac{1}{p-1} \|f\|_{L^p(X)}$$

for all $1 < p \leq 2$ and all simple functions f. Show that we also have

$$\|Tf\|_{L^1(X)} \lesssim \|f\|_{L \log L(X)}$$

for all simple functions f. (Hint: use Q10.)

- Q12 (Littlewood's principle). Let $1 \leq q < p \leq \infty$, and suppose that $T : L^p(\mathbf{R}^d) \to L^q(\mathbf{R}^d)$ is a bounded linear operator which commutes with translations, thus $T \operatorname{Trans}_{x_0} = \operatorname{Trans}_{x_0} T$ for all $x_0 \in \mathbf{R}$. Show that T is identically zero. (Hint: if T is not identically zero, then there is a non-zero $f \in L^p(\mathbf{R}^d)$ such that Tf is also not identically zero. Use monotone convergence to make f and Tf small in a suitable sense outside of a large ball. Now investigate the $L^p(\mathbf{R}^d)$ norm of $\sum_{n=1}^N \operatorname{Trans}_{x_n} f$, where the x_n are widely separated points in space, as well as the L^q norm of the image of that function under T.) Give a counterexample to show that Littlewood's principle fails if $\mathbf{R}^d$ is replaced with a compact domain such as the torus $\mathbf{R}^d/\mathbf{Z}^d$.
- Q13. Let $\mathcal{B}'_X$ be a σ-finite sub-σ-algebra of $\mathcal{B}_X$, and similarly let $\mathcal{B}'_Y$ be a σ-finite sub-σ-algebra of $\mathcal{B}_Y$. Show that if $K : X \times Y \to \mathbf{C}$ is bounded and measurable with respect to $\mathcal{B}'_X \times \mathcal{B}'_Y$, and T_K extends to a continuous linear operator from $L^p(\mathcal{B}_X)$ to $L^q(\mathcal{B}_Y)$, then we have the

$$\|T_K\|_{L^p(\mathcal{B}_X) \to L^q(\mathcal{B}_Y)} = \|T_K\|_{L^p(\mathcal{B}'_X) \to L^q(\mathcal{B}_Y)} = \|T_K\|_{L^p(\mathcal{B}'_X) \to L^q(\mathcal{B}'_Y)}$$

for all $1 \leq p, q \leq \infty$. Thus to establish strong type (p, q) bounds on T_K on the original σ-algebra $\mathcal{B}_X$, it suffices to do so on the sub-algebra $\mathcal{B}'_X$. The qualitative hypotheses that K is bounded and T_K is continuous can be relaxed, but we shall not do so here.
- Q14. Suppose that T is a continuous linear operator from $L^p(X)$ to $L^p(Y)$ for some $1 \leq p \leq \infty$. Let $E_1 \subset \ldots \subset E_N$ be a nested sequence of sets in X for some $N > 1$, and let T_* be the maximal operator

$$T_* f(y) := \sup_{1 \leq n \leq N} |T(f 1_{E_n})|.$$

Show that

$$\|T_*\|_{L^p(X) \to L^p(Y)} \lesssim \log N \|T_K\|_{L^p(X) \to L^p(Y)}.$$

- Q15 (Radamacher-Menshov inequality). Let $f_1, \ldots, f_N$ be an orthonormal set of functions in $L^2(X)$ for some $N > 1$. Show that

$$\| \sup_{1 \leq n \leq N} | \sum_{m=1}^n f_m | \|_{L^2(X)} \lesssim N^{1/2} \log N.$$

(Hint: use Q14 somehow.) Comparing this with what Pythagoras's theorem gives for $\sup_{1 \leq n \leq N} \| \sum_{m=1}^n f_m \|_{L^2(X)}$, we see that putting the sup inside only costs at most $\log N$.

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

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