LECTURE NOTES 3 FOR 247A

TERENCE TAO

1. The Hardy-Littlewood maximal inequality

Let us work in Euclidean space $\mathbf{R}^d$ with Lebesgue measure; we write $|E|$ instead of $\mu(E)$ for the Lebesgue measure of a set E. For any $x \in \mathbf{R}^d$ and $r > 0$ let $B(x, r) := \{y \in \mathbf{R}^d : |x - y| < r\}$ denote the open ball of radius r centred at x. Thus for instance $|B(x, r)| \sim_d r^d$. For any $c > 0$, we use $cB(x, r) = B(x, cr)$ to denote the dilate of $B(x, r)$ around its centre by c.

For any $r > 0$, we define the averaging operators A_r on $\mathbf{R}^d$ for any locally integrable f by

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

It is not hard to see that these averages A_r are well-defined, and are even continuous functions, for locally integrable f.

From Schur's test or Young's inequality we know that these A_r are contractions on every $L^p(\mathbf{R}^d)$, $1 \leq p \leq \infty$:

$$\|A_r f\|_{L^p(\mathbf{R}^d)} \leq \|f\|_{L^p(\mathbf{R}^d)}.$$

Thus the averages $A_r f$ are uniformly bounded in *size* as r varies. The fundamental *Hardy-Littlewood maximal inequality* asserts that they are also uniformly bounded in *shape*:

Proposition 1.1 (Hardy-Littlewood maximal inequality)**.** *We have*

$$\|\sup_{r>0} |A_r f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}$$

for all $1 < p \leq \infty$ *and any* $f \in L^p(\mathbf{R}^d)$, *and also*

$$\|\sup_{r>0} |A_r f|\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}$$

for any $f \in L^1(\mathbf{R}^d)$.

The sublinear operator

$$Mf := \sup_{r>0} A_r |f|$$

is known as the *Hardy-Littlewood maximal operator*. It is easy to see that the above proposition is equivalent to the assertion that the Hardy-Littlewood maximal operator is weak-type $(1, 1)$ and strong-type (p, p) for all $1 < p \leq \infty$. Note that it is not strong-type $(1,1)$; indeed, if f is any non-trivial function, then we easily verify the pointwise bound $Mf(x) \gtrsim_f \langle x\rangle^{-d}$, which ensures that Mf is not in $L^1_x(\mathbf{R}^d)$.

Remark 1.2. Dimensional analysis (analysing how f and Mf react under dilation of the domain $\mathbf{R}^d$ by a scaling parameter λ) shows that no weak or strong type (p,q) estimates are available in the off-diagonal case $p \neq q$.

As this proposition is so fundamental we shall give several proofs of it. We begin with the classical proof, starting with some standard qualitative reductions. Firstly we may easily reduce to f being non-negative. A monotone convergence argument also lets us restrict to functions f which are bounded and have compact support. Also, $A_r f$ is continuous in r, so we may restrict r to a countable dense set (such as the positive rationals); another monotone convergence argument then lets us restrict r to a finite set. Of course, our bounds need to be uniform in this set, as well as being uniform in the boundedness and support of f.

It is obvious that M is bounded on $L^\infty(\mathbf{R}^d)$ (indeed, it is a contraction on this space). So it suffices by Marcinkiewicz interpolation to prove the weak-type $(1,1)$ inequality; by homogeneity (and the preceding reductions) it thus suffices to show that

$$|\{\sup_{r>0} A_r f \geq 1\}| \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}$$

for any non-negative bounded compactly supported f, where we are implicitly restricting r to a finite set.

Let us denote the set on the left-hand side by E; our hypotheses on f and r easily ensure that E is a compact set[1]. By construction, we thus see that for any $x \in E$ there exists a radius $r(x) > 0$ such that f is locally large compared to the ball $B(x, r(x))$:

$$|B(x,r(x))| < \int_{B(x,r(x))} f(y)\, dy. \tag{1}$$

On the other hand, what we want to show is that f is globally large compared to E:

$$|E| \lesssim_d \int_{\mathbf{R}^d} f(y)\, dy. \tag{2}$$

Since the compact set E is covered by the balls $B(x,r(x))$, and hence by finitely many of these balls, things look quite promising. However, there is one remaining issue, which is that these balls could overlap quite heavily, preventing us from summing (1) to get (2). Fortunately there is a very simple algorithm which extracts out from any collection of overlapping balls, a collection of non-overlapping balls which manages to capture a significant fraction of the original collection in measure:

Lemma 1.3 (Wiener's Vitali-type covering lemma)**.** *Let $B_1, \dots, B_N$ be a finite collection of balls. Then there is a subcollection $B_{n_1}, \dots, B_{n_k}$ of* disjoint *balls such*

[1]Alternatively, one can work with the uncountable sup $\sup_{r>0}$, but instead replace E by an arbitrary compact subset K of itself, and then take suprema in K at the end (noting that Lebesgue measure is a Radon σ-finite Radon measure).

that

$$\sum_{i=1}^{k} |B_{n_i}| = |B_{n_1} \cup \ldots \cup B_{n_k}| \geq 3^{-d} |B_1 \cup \ldots \cup B_N|.$$

Proof We can order $B_1, \ldots, B_N$ in decreasing order of size. Now we select the disjoint balls $B'_1, \ldots, B'_N$ by the greedy algorithm, picking the largest balls we can at each stage. Namely, for $i = 1, 2, \ldots$ we choose B_{n_i} to be the first ball which is disjoint from all previously selected balls $B_{n_1}, \ldots, B_{n_{i-1}}$ (thus for instance n_1 must equal 1), until we run out of balls. Clearly this gives us a family of disjoint balls. Now observe from construction that each ball B_m in the original collection is either a ball B_{n_i} in the subcollection, or else intersects a ball B_{n_i} in the subcollection of equal or larger radius. In either case we see from the triangle inequality that B_m is contained in $3B_{n_i}$. In other words,

$$B_1 \cup \ldots \cup B_N \subseteq 3B_{n_1} \cup \ldots \cup 3B_{n_k}$$

and so

$$|B_1 \cup \ldots \cup B_N| \leq 3^d \sum_{i=1}^{k} |B_{n_i}| = 3^d |B_{n_1} \cup \ldots \cup B_{n_k}|$$

and the claim follows. ■

From the covering lemma it is easy to conclude (2). Indeed, since E is covered by finitely many of the balls $B(x, r(x))$, the covering lemma gives us finitely many *disjoint* balls $B(x_i, r(x_i))$, $i = 1, \ldots, k$ such that

$$|\sum_{i=1}^{k} B(x_i, r(x_i))| \geq 3^{-d} |E|$$

and then on summing (1) we get (2) (with an explicit constant of 3^d).

Remarks 1.4. Under mild assumptions one can generalise the covering lemma to infinite families of balls without difficulty. One can also replace balls by similar objects, such as cubes; the main property that one needs is that if two such objects overlap, then the smaller one is contained in some dilate of the larger. This is a fairly general property, and for instance holds for metric balls on a measure space with some doubling property $\mu(B(x, 2r)) = O(\mu(B(x, r)))$, but it fails for very thin or eccentric sets such as long tubes, rectangles, annuli, etc. Indeed, understanding the maximal operator for these more geometrically complicated objects is still a major challenge in harmonic analysis, leading to important open conjectures such as the Kakeya conjecture.

Let us now give a slightly different proof of the above inequality, replacing balls by the slightly simpler structure of "dyadic cubes".

Definition 1.5 (Dyadic cube)**.** A *dyadic cube* in $\mathbf{R}^d$ of *generation* n is a set of the form

$$Q = Q_{n,k} = 2^n (k + [0, 1)^d) = \{2^n (k + x) : x \in [0, 1)^d\}$$

where n is an integer and $k \in \mathbf{Z}^d$.

The crucial property of dyadic cubes is the *nesting property*: if two dyadic cubes overlap, then one must contain the other. This leads to

Lemma 1.6 (Dyadic Vitali-type covering lemma)**.** *Let $Q_1, \ldots, Q_N$ be a finite collection of dyadic cubes. Then there is a subcollection $Q_{n_1}, \ldots, Q_{n_k}$ of* disjoint *cubes such that*

$$Q_{n_1} \cup \ldots \cup Q_{n_k} = Q_1 \cup \ldots \cup Q_N.$$

Proof Take the Q_{n_i} to be the maximal dyadic cubes in $Q_1, \ldots, Q_N$ - the cubes which are not contained in any other cubes in this collection. The nesting property then ensures that they are disjoint and cover all of $Q_1, \ldots, Q_N$ between them. ∎

If we then define the *dyadic maximal function*

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

where Q ranges over the dyadic cubes which contain x, then the same argument as before then gives the *dyadic Hardy-Littlewood maximal inequality*

$$\|M_\Delta f\|_{L^{1,\infty}(\mathbf{R}^d)} \leq \|f\|_{L^1(\mathbf{R}^d)} \tag{3}$$

(with no constant loss whatsoever!) which then leads via Marcinkiewicz interpolation to

$$\|M_\Delta f\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}$$

for $1 < p \leq \infty$.

We can rewrite the dyadic maximal inequality in another way. Let $\mathcal{B}_n$ be the σ-algebra generated by the dyadic cubes of generation n, then

$$\mathbf{E}(f|\mathcal{B}_n)(x) = \fint_Q f(y)\, dy$$

where Q is the unique dyadic cube of generation n which contains x. The dyadic Hardy-Littlewood maximal inequality is then equivalent to the assertion that

$$\|\sup_n |\mathbf{E}(f|\mathcal{B}_n)|\|_{L^{1,\infty}(\mathbf{R}^d)} \leq \|f\|_{L^1(\mathbf{R}^d)}$$

and thus

$$\|\sup_n |\mathbf{E}(f|\mathcal{B}_n)|\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}$$

for $1 < p \leq \infty$.

Observe that if $x \in Q$, then there is a ball $B(x,r)$ centred at x which contains Q of comparable volume: $|B(x,r)| \sim_d Q$. Because of this, one easily obtains the pointwise inequality

$$M_\Delta f(x) \lesssim_d Mf(x)$$

and so the dyadic inequality follows (up to constants) from the non-dyadic one. The converse pointwise inequality is not true (test it with $d = 1$ and $f = 1_{[0,1]}$, for instance). However, a slightly modification of this inequality is true, thanks to the $1/3$-*translation trick* of Michael Christ. We first explain this trick in the context of the unit interval $[0,1]$.

Lemma 1.7. *Let $I \subset [0,1]$ be a (non-dyadic) interval. Then there exists an interval J which is either a dyadic interval, or a dyadic interval translated by $1/3$, such that $I \subset J$ and $|J| \lesssim |I|$.*

The only significance of $1/3$ is that its binary digit expansion oscillates between 0 and 1. Note that the claim is false without the $1/3$ shifts; consider for instance the interval $[0.5-\varepsilon, 0.5+\varepsilon]$ for some very small ε, which straddles a certain "discontinuity" in the standard dyadic mesh. The point is that the dyadic mesh and the $1/3$-translate of the dyadic mesh do not share any discontinuities.

We leave the proof of the above simple lemma to the reader. For intervals larger than $[0,1]$, a shift by $1/3$ is not enough; consider for instance what happens to the interval $(-1,1)$. Instead, we have to shift by $4^\infty/3$, which of course does not make sense as a real number. However, it does make sense in some formal 2-adic sense (as the doubly infinite binary string $\ldots 10101.010101\ldots$) which is good enough to define shifted dyadic meshes.

Definition 1.8 (Dyadic meshes)**.** We define $\mathcal{D}^1_0$ to be the collection of all dyadic intervals in $\mathbf{R}$. We define $\mathcal{D}^1_{4^\infty/3}$ to be the collection of all intervals of the form $I + 4^N/3$, where I is a dyadic interval at some generation n and N is any integer greater than or equal to n (note that the exact choice of N is irrelevant). If $\alpha = (\alpha_1, \ldots, \alpha_d) \in \{0, 4^\infty/3\}^d$, we let $\mathcal{D}^d_\alpha = \mathcal{D}^1_{\alpha_1} \times \ldots \times \mathcal{D}^1_{\alpha_d}$ be the collection of cubes formed by the Cartesian product of intervals from $\mathcal{D}^1_{\alpha_1}, \ldots, \mathcal{D}^1_{\alpha_d}$.

By modifying the above lemma one then quickly deduces

Lemma 1.9. *Let $B \subset \mathbf{R}^d$ be a ball. Then there exists $\alpha \in \{0, 4^\infty/3\}^d$ and a shifted dyadic cube $Q \in \mathcal{D}^d_\alpha$ such that $B \subset Q$ and $|Q| \lesssim_d |B|$.*

This in turn leads to the pointwise inequality bounding the dyadic maximal function by the ordinary one:

$$Mf(x) \lesssim_d \sup_{\alpha \in \{0, 4^\infty/3\}^d} M_{\Delta,\alpha} f(x)$$

where $M_{\Delta,\alpha}$ is the shifted dyadic maximal function

$$M_{\Delta,\alpha} f(x) := \sup_{Q \in \mathcal{D}^d_\alpha : Q \ni x} \fint_Q |f|.$$

A routine modification of the proof of the dyadic maximal inequality (or translating this inequality by $4^N/3$ and taking limits as $N \to \infty$) shows that each of the $M_{\Delta,\alpha}$ are individually of weak-type $(1,1)$, and bounded on L^p for $1 < p \leq \infty$. Since there are only $O_d(1)$ many choices of α, we can then deduce the usual Hardy-Littlewood maximal inequality from the dyadic one.

Remark 1.10. What is going on here is that there are two ways to view the real line $\mathbf{R}$. One is the "Euclidean" way, with the usual group structure and metric. The other is the "Walsh" or "dyadic" way, in which we identify $\mathbf{R}$ with the *Cantor group*

$$\mathbf{R}_\Delta := \{(a_n)_{n \in \mathbf{Z}} \in (\mathbf{Z}/2\mathbf{Z})^{\mathbf{Z}} : a_n = 0 \text{ for sufficiently large } n\}$$

via the binary representation, $\sum_n a_n 2^{-n} \equiv (a_n)_{n \in \mathbf{Z}}$ (identifying $\mathbf{Z}/2\mathbf{Z}$ with $(0,1)$, and ignoring the measure zero sets of terminating decimals where the binary representation is not unique). The group structure is now the one inherited from the Cantor group; in the binary representation, the Cantor-Walsh addition law $x +_\Delta y$ is the same as ordinary addition $x + y$ but where we neglect to carry bits. The usual archimedean metric $d(x, y) = |x - y|$ is replaced by the non-archimedean metric $d_\Delta(x, y)$, defined by

$$d_\Delta((a_n)_{n \in \mathbf{Z}}, (b_n)_{n \in \mathbf{Z}}) := \sup\{2^n : a_n \neq b_n\}.$$

With this metric, the dyadic intervals become the metric balls.

In the above arguments we obtained L^p bounds by first proving weak $(1,1)$ bounds and then interpolating. It is natural to ask whether such bounds can be obtained directly. The answer is yes, but it is surprisingly more difficult to do so. Let us give two such approaches, a *Bellman function* approach and a TT^* *method approach*, which are themselves powerful methods which apply to many other problems as well.

We begin with the Bellman function approach. This method works primarily for dyadic model operators, such as M_Δ, though it can also work for very geometric operators as well (using geometric averaging operators such as heat kernels in place of the dyadic averaging operators). For simplicity let us just work in one dimension (though it is possible to use rearrangement and space-filling curves to deduce the higher-dimensional case from the one-dimensional case), and consider the task of establishing

$$\|M_\Delta f\|_{L^p(\mathbf{R})} \lesssim_p \|f\|_{L^p(\mathbf{R})}$$

for some fixed $1 < p \leq \infty$ (such as $p = 2$).

The idea is to work by *induction on scales* - in other words, to induct on the number of generations. To do this we need a “base case”, so we perform some qualitative reductions. Fix $1 < p < \infty$ (the case $p = \infty$ being trivial). By a monotone convergence argument we may restrict attention only to those intervals I of length larger than 2^{-N}, so long as our estimates are uniform in N. By rescaling (replacing $f(x)$ by $f(2^N x)$) we can reduce to the case $N = 0$. Let us write $M_{\geq 1}$ for the dyadic maximal function restricted to intervals of length at least 1. By a monotone convergence argument[2] we can also assume that f is supported on a dyadic interval I of some length $2^n \geq 1$, and also we may restrict $M_{\geq 1} f$ to that interval. We can also take f to be non-negative. Our task is now to show that there exists a constant C_p such that

$$⨍_I (M_{\geq 1} f(x))^p \, dx \leq C_p ⨍_I f(x)^p \, dx \tag{4}$$

whenever $f : I \to \mathbf{R}^+$. (We make the constant C_p explicit here because of the induction that we shall shortly use.) Of course the point is that C_p is independent of n, I, and f.

[2]There is a slight problem because we cannot represent $\mathbf{R}$ as the monotone limit of dyadic intervals. However we can do this for $\mathbf{R}^+$ and $\mathbf{R}^-$ separately, and then add up, noting that the dyadic maximal function is localised to each of these half-lines (e.g. if f is supported on $\mathbf{R}^+$ then so is $M_{\geq 1} f$).

We make a small but useful remark: once f and $M_{\geq 1} f$ are both restricted to I, the only dyadic intervals J which are relevant in the definition of $M_{\geq 1} f$ are those which are contained in I (including I itself). Intervals which are disjoint from I play no role, and intervals which contain I give a worse average than that arising from I itself.

The idea is to prove this by induction on the generation n of I. Our first approach will not quite work, but a subtle modification of it will.

When $n = 0$ the claim is trivial (as long as $C_p \geq 1$), because for I a dyadic interval of generation 0 and $x \in I$ we have

$$M_{\geq 1} f(x) = ⨍_I f(y)\ dy \leq (⨍_I f(x)^p)^{1/p}.$$

Now let $n \geq 1$, and let us see whether we can deduce the n case from the $n-1$ case *without causing any deterioration in the constant* C_p. (Using various applications of the triangle inequality it is not hard to get from $n-1$ to n, replacing C_p with something worse like $2C_p + 1$, but this is not going to iterate into something independent of n.)

Let us split the dyadic interval I of generation n into two "children" $I = I_l \cup I_r$ of generation $n-1$ (l and r stand for left and right). This also causes a split $f = f_l + f_r$. By induction hypothesis we have

$$⨍_{I_l} (M_{\geq 1} f_l(x))^p\ dx \leq C_p ⨍_{I_l} f_l(x)^p\ dx$$
$$⨍_{I_r} (M_{\geq 1} f_r(x))^p\ dx \leq C_p ⨍_{I_r} f_r(x)^p\ dx.$$

and we also trivially have

$$⨍_I f(x)^p\ dx = \frac{1}{2}(⨍_{I_l} f_l(x)^p\ dx + ⨍_{I_l} f_l(x)^p)\ dx.$$

Now if we were lucky enough to have the pointwise estimates

$$M_{\geq 1} f(x) \leq M_{\geq 1} f_l(x) \text{ when } x \in I_l$$

and similarly for I_r, then we could simply average the two induction hypotheses and be done. However, this is not quite the case: the correct relationship between the maximal function of f and of f_l, f_r is that

$$M_{\geq 1} f(x) = \max(M_{\geq 1} f_l(x), ⨍_I f) \text{ when } x \in I_l$$

and similarly when $x \in I_r$. This causes a problem. If we estimate the max by a sum and use triangle type inequalities, we will eventually get a bound such as (4) but with C_p replaced by $O_p(C_p + 1)$, which is not acceptable for iteration purposes. So we have to somehow keep the max with the constant $⨍_I f$ with us in the induction argument. This eventually forces us to change the induction hypothesis (4), replacing the left-hand side $⨍_I M_{\geq 1} f(x)^p\ dx$ by the more general $⨍_I \max(M_{\geq 1} f(x), A)^p\ dx$ for some arbitrary $A > 0$. Given our knowledge that max and addition are comparable up to constants, we know that (4) is equivalent to the estimate

$$\fint_I \max(M_{\geq 1}f(x), A)^p \, dx \leq C_p'(\fint_I f(x)^p \, dx + A^p)$$

up to changes in the constant C_p. But perhaps this estimate has a better chance of being proven by induction. The key recursive inequality is now that

$$\max(M_{\geq 1}f(x), A) = \max(M_{\geq 1}f_l(x), \max(A, \fint_I f)) \text{ when } x \in I_l$$

and similarly for I_l. One can try the induction strategy again, but one sees that the inability to efficiently control $\fint_I f$ in terms of $\int_I f(x)^p \, dx$ and A is a serious problem. (Hölder's inequality of course gives $\fint_I f$ bounded by $(\fint_I f^p)^{1/p}$, but this turns out to be insufficient.) Because of this, we have no choice but to also throw in the average $B = \fint_I f$ into the induction hypothesis somehow.

Let us formalise this as follows. Given any parameters $D, B, A > 0$, let $V_n(A, B, D)$ denote the *cost function*

$$V_n(A, B, D) := \sup\{\fint_I \max(M_{\geq 1}f(x), A)^p \, dx : \fint_I f^p = D; \fint_I f = B; |I| = 2^n\}$$

where f is understood to be nonnegative and supported on a dyadic interval I of generation n. Note that Hölder's inequality shows that $V_n(A, B, D) = -\infty$ when $B^p > D$, since in this case the supremum is over an empty set. Our task is thus to show that

$$V_n(A, B, D) \lesssim_p D + A^p \tag{5}$$

uniformly in n and in A, B, D. As we said earlier, the B parameter is not obviously necessary yet, but will become so when we try to perform the induction, as it tracks a certain finer property of the function f which needs to be managed in order to prevent the constants from blowing up. The base case when $|I| = 1$ is again trivial; the issue is to pass from fine scales to coarse scales without destroying the boundedness of implicit constant.

Now the recursive inequality can be turned into an inequality for V_n. Suppose that f attains the supremum (or comes within an epsilon of it). We have

$$\fint_{I_l} f_l^p = D - \delta; \quad \fint_{I_r} f_r^p = D + \delta$$

for some $|\delta| \leq D$. Similarly

$$\fint_{I_l} f_l = B - \beta; \quad \fint_{I_r} f_r = B + \beta$$

for some $|\beta| \leq B$. Then we have

$$\max(M_{\geq 1}f(x), A) = \max(M_{\geq 1}f_l(x), \max(A, B))$$

for $x \in I_l$, and similarly for $x \in I_r$; by construction we thus have

$$\fint_I \max(M_{\geq 1}f(x), A)^p \leq \frac{1}{2}(V_{n-1}(\max(A, B), B-\beta, D-\delta) + V_{n-1}(\max(A, B), B+\beta, D+\delta)).$$

Taking suprema, we obtain the recursive inequality

$$V_n(A,B,D) \leq \sup_{|\beta|\leq B, |\delta|\leq D} \frac{1}{2}(V_{n-1}(\max(A,B), B-\beta, D-\delta) + V_{n-1}(\max(A,B), B+\beta, D+\delta)).$$

(In fact, this is an equality - why?) On the other hand, V_0 can be computed directly as

$$V_0(A,B,D) = \max(A,B)^p$$

when $B^p \leq D$, and $V_0(A,B,D) = -\infty$ otherwise. In principle, this gives us a complete description of $V_n(A,B,D)$, which should allow one to determine the truth or falsity of (5). Note that we have reduced the problem from one involving an unknown function f (which has infinitely many degrees of freedom) to one involving just three scalar parameters A, B, D, except that we have a different cost function at every scale. However, suppose that we can devise a *Bellman function* $\Phi(A,B,D)$ with the property that

$$V_0(A,B,D) \lesssim \Phi(A,B,D) \lesssim_p D + A^p \tag{6}$$

for all A, B, D, and such that Φ obeys the inequality

$$\Phi(A,B,D) \geq \frac{1}{2}(\Phi(\max(A,B), B-\beta, D-\delta) + \Phi(\max(A,B), B+\beta, D+\delta)) \tag{7}$$

whenever $|\beta| \leq B$ and $|\delta| \leq D$. Then an induction will show that

$$V_n(A,B,D) \lesssim \Phi(A,B,D) \lesssim_p D + A^p$$

for all n (and note that the implied constants here are *uniform* in n), thus proving (5).

Thus the whole task is reduced to a freshman calculus problem, namely to find a function of three variables obeying the bounds (6) and (7). The difficulty of course, is to find the function; verifying the properties is then routine. This "hunt for a Bellman function" is surprisingly subtle, and requires one to choose a surprisingly non-trivial choice of Φ.

The condition (7) resembles a concavity condition, and so it is natural to try to find choices which are concave in some of the variables. An initial candidate is the function $\Phi(A,B,D) = D$, which certainly obeys (7) and the upper bound of (6) (and which, in fact, ultimately corresponds to the original estimate (4) before we threw in the other parameters B, A); unfortunately it does not obey the lower bound in (6), in the case where A is large compared with B and $D^{1/p}$. So we need to tweak this function somewhat. The first step is to improve the concavity by exploiting the fact that the function only needs to be non-trivial on the region $B^p \leq D$. One can exploit this by using the candidate function $\Phi(A,B,C) = D - \frac{1}{2}B^p$ (say). This still obeys (7), but now with a bit of a gain when β is large due to the *strict* concavity of $-B^p$. (One cannot play similar games with the D parameter as the upper and lower bounds in (6) force linear-type behaviour in D.) But we still have not fixed the problem that Φ is not as large as $V_0(A,B,D)$ when A is large. The solution is to use the Bellman function

$$\Phi(A,B,C) = D + A^p f(B/A)$$

where $f(x)$ is a concave function which is positive for small x and is equal to $-\frac{1}{2}x^p$ for $x > 1/10$ (say). Thus we have $\Phi = D - \frac{1}{2}B^p$ when $A < 10B$, but Φ becomes as large as $D + A^p$ or so when A exceeds $10B$. This lets us verify both sides of (6), so one only needs to verify (7). If $A < 5B$ then the claim follows from the concavity of $D - \frac{1}{2}B^p$ (which does not depend on A); when $A \geq 5B$ the claim follows from the concavity of Φ in the B variable and in the D variable.

Remark 1.11. Bellman function methods are in principle the sharpest and most powerful technique to prove estimates. However they are restricted to dyadic settings (or very geometric continuous settings), and are very delicate due to the need to establish very subtle concavity properties.

Finally, we present the "TT^* approach" to the Hardy-Littlewood maximal inequality. This approach works for both the dyadic and non-dyadic maximal functions, but is restricted to establishing L^2 boundedness (this is a fundamental limitation of the TT^* method, that at least one of the spaces involved has to be a Hilbert space). The basic idea is rather than prove a boundedness result directly on the maximal operator M (i.e. an estimate of the form $M = O(1)$), we prove an estimate on the square of this operator (roughly speaking, we prove an estimate of the form $M^2 = O(M)$). This is an example of a powerful strategy to understand an operator by raising it to a higher power, and hoping to exploit some self-cancellation. Note that we do not have to cancel the entire power (which would roughly speaking correspond to proving a bound of the form $M^2 = O(1)$); any nontrivial cancellation at all is exploitable.

Let us work with the non-dyadic maximal function, thus we wish to show that

$$\| \sup_{r>0} A_r f \|_{L^2(\mathbf{R}^d)} \lesssim_d \|f\|_{L^2(\mathbf{R}^d)}$$

for all non-negative f. As before we may restrict r to range in a finite set $\mathcal{R}$, provided our bounds are independent of the choice of $\mathcal{R}$. Note that because each A_r is already bounded on L^2, the maximal operator is now already bounded with *some* finite operator norm. Let D denote the optimal such norm, thus D is the best constant for which

$$\| \sup_{r\in\mathcal{R}} A_r f \|_{L^2(\mathbf{R}^d)} \leq D \|f\|_{L^2(\mathbf{R}^d)}. \tag{8}$$

We know D is finite; our objective is to obtain the bound $D = O_d(1)$. We will do this by controlling D in a nontrivial way in terms of itself, and in particular by controlling the "square" of the maximal operator by the maximal operator.

Observe that (8) is equivalent to the uniform linearised estimate

$$\|A_{r(x)} f(x)\|_{L^2_x(\mathbf{R}^d)} \leq D \|f\|_{L^2(\mathbf{R}^d)}$$

for all measurable functions $r : \mathbf{R}^d \to \mathcal{R}$.

Let us fix the function r. Then we can define the linear operator T_r by

$$T_r f(x) := A_{r(x)} f(x) = \frac{1}{|B(x, r(x))|} \int_{\mathbf{R}^d} 1_{|x-y|\leq r(x)} f(y)\, dy.$$

We can thus interpret D as the largest L^2 operator norm of the T_r:

$$D = \sup_r \|T_r\|_{L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)}.$$

To compute this operator norm we make the following observation.

Lemma 1.12 (TT^* identity)**.** *Let $T : H \to X$ be a continuous map from a Hilbert space to a normed vector space, and let $T^* : X^* \to H$ be its adjoint. Then*

$$\|T\|_{H \to X} = \|T^*\|_{X^* \to H} = \|TT^*\|_{X^* \to X}^{1/2}.$$

Proof The first identity is just duality. Then we have

$$\|TT^*\|_{X^* \to X} \leq \|T\|_{H \to X} \|T^*\|_{X^* \to H} = \|T^*\|_{X^* \to H}^2$$

which gives the lower bound in the second identity. For the upper bound, observe that for any $f \in X^*$ that

$$\|T^* f\|_H^2 = \langle f, TT^* f \rangle \leq \|f\|_{X^*} \|TT^* f\|_X \leq \|f\|_{X^*}^2 \|TT^*\|_{X^* \to X};$$

taking square roots gives the upper bound as desired. ■

In light of this identity we know that

$$D^2 = \sup_r \|T_r T_r^*\|_{L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)}.$$

Now let us take a look at what $T_r T_r^*$ is. Observe that T_r is an integral operator with kernel

$$K(x, y) := \frac{1}{|B(x, r(x))|} 1_{|x-y| \leq r(x)}.$$

Thus the adjoint is given by

$$T_r^* g(y) = \int_{\mathbf{R}^d} \frac{1}{|B(x, r(x))|} 1_{|x-y| \leq r(x)} g(x)\ dx$$

and then TT^* is given by

$$T_r T_r^* g(x') = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \frac{1}{|B(x, r(x))||B(x', r(x'))|} 1_{|x-y| \leq r(x)} 1_{|x'-y| \leq r(x')} g(x)\ dy dx.$$

Note that the y integral can be computed fairly easily. First we observe that the y integral vanishes unless $|x - x'| \leq r(x) + r(x')$, and in the latter case it enjoys a bound of $O_d(\min(r(x), r(x'))^d)$. Also, $|B(x, r(x))| \sim_d r(x)^d$ and $|B(x', r(x'))| \sim_d r(x')^d$. Putting this together we see that

$$|T_r T_r^* g(x')| \lesssim_d \int_{\mathbf{R}^d} 1_{|x-x'| \leq r(x) + r(x')} \frac{1}{\max(r(x), r(x'))^d} |g(x)|\ dx.$$

It is natural to split this integral into the regions $r(x) \leq r(x')$ and $r(x) \geq r(x')$, leading to the bound

$$|T_r T_r^* g(x')| \lesssim_d \int_{\mathbf{R}^d} 1_{|x-x'| \leq 2r(x)} \frac{1}{r(x)^d} |g(x)|\ dx + \int_{\mathbf{R}^d} 1_{|x-x'| \leq 2r(x')} \frac{1}{r(x')^d} |g(x)|\ dx.$$

Comparing this with the formulae for T_r and $T_{r'}$, we obtain the interesting pointwise inequality

$$|T_r T_r^* g(x')| \lesssim_d T_{2r}|g|(x') + T_{2r}^*|g|(x')$$

where $2r$ is the function $x \mapsto 2r(x)$, of course. On the other hand, a scaling argument gives

$$\|T_{2r}\|_{L^2(\mathbf{R}^d)\to L^2(\mathbf{R}^d)} \leq D$$

and hence we conclude from the triangle inequality that

$$\|T_r T_r^*\|_{L^2(\mathbf{R}^d)\to L^2(\mathbf{R}^d)} \lesssim_d D;$$

taking suprema in r we conclude that $D^2 \lesssim_d D$, and hence (since D is known to be finite) $D \lesssim_d 1$, and the claim follows.

The Hardy-Littlewood maximal function directly bounds averages on balls and cubes, but it also controls several other types of averages as well. For instance, we have the pointwise inequality

$$\frac{1}{r^d}\int_{\mathbf{R}^d} \frac{1}{\langle (x-y)/r\rangle^{d+\varepsilon}} |f(y)|\ dy \lesssim_{d,\varepsilon} Mf(x) \tag{9}$$

for any locally integrable f, any $x \in \mathbf{R}^d$, and any $\varepsilon > 0$; this can be achieved by dividing the y integral into dyadic shells $2^n r \leq |x-y| < 2^{n+1} r$ for $n \geq 0$, as well as the ball $|x-y| < r$, and we leave the computation to the reader.

The proof of the Hardy-Littlewood maximal inequality extends to the more general setting of *homogeneous spaces*. These are measure spaces $(X, \mathcal{B}, \mu)$ with a metric d, such that the open balls are measurable with positive finite measure, and that one has the doubling property

$$\mu(B(x, 2r)) \lesssim \mu(B(x,r))$$

for any x, r. Then the Vitali-type covering lemma extends without difficulty to this setting and yields the maximal inequality

$$\|\sup_{r>0} ⨍_{B(x,r)} |f(y)|\ d\mu(y)\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$$

and hence

$$\|\sup_{r>0} ⨍_{B(x,r)} |f(y)|\ d\mu(y)\|_{L^p(X)} \lesssim_p \|f\|_{L^p(X)}$$

for any $1 < p \leq \infty$.

In particular we obtain the discrete inequality

$$\|\sup_{N>0} \frac{1}{N}\sum_{n=1}^N f(m+n)\|_{l^{1,\infty}_m(\mathbf{Z})} \leq \|f\|_{l^1(\mathbf{Z})}; \tag{10}$$

(which we need in our applications to ergodic theory below), while on the torus $\mathbf{R}/\mathbf{Z}$ with the usual Lebesgue measure we have

$$\|\sup_{r>0} \frac{1}{2r}\int_{|x-y|\leq r} |f(y)|\ dy\|_{L^p(\mathbf{R}/\mathbf{Z})} \lesssim_p \|f\|_{L^p(\mathbf{R}/\mathbf{Z})} \tag{11}$$

for $1 < p \leq \infty$ (which we will need for our applications to complex analysis).

Remark 1.13. One could try playing with T^*T instead of TT^* here, but that turns out to not work very well. The problem is that the linearised maximal operator is only well-behaved in one variable, and the TT^* method manages to play the two well-behaved variables against each other; going the other way one achieves no obvious cancellation.

2. Some consequences of the maximal inequality

The Hardy-Littlewood maximal inequality is the underlying quantitative estimate which powers many qualitative pointwise convergence results. A basic example is

Theorem 2.1 (Lebesgue differentiation theorem)**.** *Let f be locally integrable. Then we have the pointwise convergence* $\lim_{r \to 0} A_r f(x) = f(x)$ *for almost every x. In fact we have the stronger estimate*

$$\lim_{r \to 0} ⨍_{B(x,r)} |f(y) - f(x)| \, dy = 0 \tag{12}$$

for almost every x.

Before we prove the theorem, let us make some remarks. Firstly, from the identity

$$A_r f(x) - f(x) = ⨍_{B(x,r)} (f(y) - f(x)) \, dy$$

and the triangle inequality it is clear that the second estimate implies the first. Secondly, observe that f is continuous at x if and only if the worst-case local fluctuation goes to zero

$$\lim_{r \to 0} \sup_{y \in B(x,r)} |f(y) - f(x)| = 0.$$

Thus the differentiation theorem is asserting that locally integrable functions are *almost everywhere continuous on the average*, in that the average-case local fluctuation goes to zero. This is a manifestation of one of Littlewood's three principles, namely that *measurable functions are almost continuous*. If x is such that (12) holds, we say that x is a *Lebesgue point* of f, thus for locally integrable f, almost every point x is a Lebesgue point.

Proof It suffices to establish the result for x constrained to a large ball $B(0, R)$, where $R > 0$ is arbitrary. The claim is obvious if f vanishes on $B(0, 2R)$, so by linearity we may assume that f is supported on $B(0, 2R)$; in particular f now lies in $L^1(\mathbf{R}^d)$. Thus it suffices to establish the claim when $f \in L^1(\mathbf{R}^d)$.

By the preceding discussion we already know that the claim is true for the continuous compactly supported functions $C_c(\mathbf{R}^d)$, which is a dense subclass of $L^1(\mathbf{R}^d)$. To pass from the dense subclass to the full class we use the Hardy-Littlewood maximal inequality now lets us reduce matters to verifying the claim on a dense subset of $L^1(\mathbf{R}^d)$. To see this, suppose that we already have proven the claim on a dense class. Then, given any $f \in L^1(\mathbf{R}^d)$ we can write f as the limit in L^1 of a sequence f_n in this dense class; by refining these sequence to make the L^1 convergence sufficiently fast (e.g. $\|f_n - f\|_{L^1(\mathbf{R}^d)} \leq 2^{-n}$) and using Markov's inequality and the Borel-Cantelli lemma, we can also ensure that f_n converges to f pointwise almost everywhere. Now from the Hardy-Littlewood maximal inequality, we know that $\| \sup_{r>0} A_r |f_n - f| \|_{L^{1,\infty}(\mathbf{R}^d)}$ converges to zero, thus $\sup_{r>0} A_r |f_n - f|$ converges to zero in measure. By the triangle inequality this implies that $\sup_{r>0} |A_r f_n - A_r f|$ converges in measure also. Thus (again passing to a rapidly converging subsequence as necessary) we see that $\sup_{r>0} |A_r f_n(x) - A_r f(x)|$ converges to zero for almost every x. This uniform-in-r convergence lets one deduce the convergence of $A_r f$ from the convergence of $A_r f_n$.

Finally, by taking a dense class such as the Schwartz class (or even the continuous functions in $L^1(\mathbf{R}^d)$) we easily verify the convergence. ∎

Remark 2.2. Because the proof used a density argument, it offers no quantitative rate for the speed of convergence. Indeed the convergence can be arbitrarily slow[3]. Consider for instance the function $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = 1$ when $0 < x < 1$ and the integer part of Nx is even, and $f(x) = 0$ otherwise, where N is a large integer. Then we see that for $0 < x < 1$, $A_r(x)$ will stay close to $1/2$ for quite a while (basically for all $r \gg 1/N$), and only at scales $1/N$ or less will it begin to "decide" to converge to either 0 or 1. Modifying this type of example (e.g. by a "Weierstrass example" formed by summing together a geometrically decaying series of such oscillating functions, with N going to infinity), one can concoct functions whose convergence of $A_r f$ to f is arbitrarily slow. (One can obtain quantitative rates by enforcing regularity conditions on the function, which essentially compactifies the space of functions that one is working with; we will see some examples of this later.)

Now let us explore applications to ergodic theory. In particular we wish to investigate limits of the form

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f$$

for various functions f and various "shift operators" T.

Let us first look at an abstract setting, in which T is a unitary operator on a Hilbert space.

Theorem 2.3 (Von Neumann ergodic theorem)**.** *Let $T : H \to H$ be a unitary operator on a Hilbert space. Then for any $f \in H$, the limit $\lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^n f$ exists in the strong H topology.*

Proof Let us first argue formally. We have

$$\frac{1}{N} \sum_{n=1}^{N} T^n f = \frac{T + \ldots + T^N}{N} f.$$

Formally, we have the geometric series formula

$$\frac{T + \ldots + T^N}{N} f = \frac{T^{N+1} - T}{N(T-1)} f$$

[3]This is ultimately due to the implicit hypothesis that f is measurable, which is itself a qualitative assumption that offers no explicit bounds. More quantitative versions of measurability, e.g. quantifying the extent to which a measurable set can be approximated by elementary sets, can lead to more explicit bounds.

which looks like it converges to zero, unless $T-1$ fails to be invertible; but on the other hand when $T=1$ $\frac{T+\ldots+T^N}{N}$ is just 1, which of course converges to the identity. So we seem to have covered two extreme cases.

Now let us make the above argument rigorous. If f is T-invariant, thus $Tf=f$, then $T^n f = f$ and it is clear that $\lim_{n\to\infty} \frac{1}{N}\sum_{n=1}^N T^n f$ converges to f. If on the other hand f is a T-difference, $f = Tg - g$ for some $g \in H$, then we verify the telescoping identity $\sum_{n=1}^N T^n f = T^{N+1} g - Tg$. Since unitary operators preserve the norm, we thus see from the triangle inequality that $\lim_{n\to\infty} \frac{1}{N}\sum_{n=1}^N T^n f$ converges to zero. Also, since the averaging operators $\frac{1}{N}\sum_{n=1}^N T^n$ are uniformly bounded in H (by the triangle inequality), we see that any strong limit of T-differences also has the property that $\lim_{n\to\infty} \frac{1}{N}\sum_{n=1}^N T^n f$ exists and equals zero.

To summarise so far, we have two closed subspaces of H for which we know convergence. The first is $H^T := \{f \in H : Tf = f\}$, the invariant space of T; here the limit converges to the identity. The other is $\overline{\{Tg - g : g \in H\}}$, the closure of the T-differences; here the limit conveges to zero. These two spaces turn out to be orthogonal complements. To see this, first observe that they are orthogonal: if $Tf = f$, then $\langle f, Tg\rangle = \langle Tf, Tg\rangle = \langle f, g\rangle$ by unitarity, and hence f is orthogonal to every T-difference $Tg-g$ and hence by continuity of inner product is orthogonal to $\overline{\{Tg - g : g \in H\}}$. To show orthogonal complement, it then suffices to show that any vector f orthogonal to $\overline{\{Tg - g : g \in H\}}$ is invariant. But then f is orthogonal to $Tf - f$:

$$\langle f, Tf - f\rangle = 0.$$

We rewrite the left-hand side as

$$\langle f, Tf\rangle - \langle f, f\rangle = \langle f, Tf\rangle - \langle Tf, Tf\rangle = -\langle Tf - f, Tf\rangle$$

and on conjugating and subtracting we conclude that

$$\langle Tf - f, Tf - f\rangle = 0$$

and thus $Tf = f$ as claimed. ■

Note that the above argument in fact shows the stronger claim that $\lim_{n\to\infty} \frac{1}{N}\sum_{n=1}^N T^n f$ converges to the orthogonal projection of f to H^T.

Now we move to a more specialised setting, that of a measure-preserving system.

Definition 2.4 (Measure-preserving system)**.** A *measure-preserving system* $(X, \mathcal{B}, \mu, T)$ is a probability space $(X, \mathcal{B}, \mu)$ (thus $\mu(X) = 1$) together with a bi-measurable bijection $T : X \to X$ (thus T and T^{-1} are both measurable) such that T is measure-preserving, thus $\mu(T(E)) = \mu(E)$ for all $E \in \mathcal{B}$.

Example 2.5 (Circle shift)**.** *Let $X = \mathbf{R}/\mathbf{Z}$ be the standard circle with the usual Borel σ-algebra and Lebesgue measure, and let $Tx := x + \alpha$ for some $\alpha \in \mathbf{R}$, which may be either rational or irrational.*

Remark 2.6. Many of the results here hold under more relaxed assumptions on T, but we will not attempt to optimise the hypotheses here.

The shift T on the base space X, $x \mapsto T(x)$, induces a shift on sets $E \in \mathcal{B}$, $E \mapsto T(E)$, and then also induces a map on measurable functions $f : X \to \mathbf{C}$ by $Tf := f \circ T^{-1}$. The use of T^{-1} is natural since it ensures that $T1_E = 1_{TE}$ and $Tf(Tx) = f(x)$.

Proposition 2.7 (Mean ergodic theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, and let $f \in L^p(X)$ for some $1 \leq p < \infty$. Then the sequence $\frac{1}{N}\sum_{n=1}^N T^n f$ converges strongly in $L^p(X)$.*

Proof The operator T is an isometry on L^p, and so by the triangle inequality the averaging operators $\frac{1}{N}\sum_{n=1}^N T^n$ are uniformly bounded in L^p. Thus it suffices to prove the claim for a dense subclass of L^p; we shall pick $L^\infty(X)$. In this class, which is embedded into the Hilbert space $L^2(X)$, we already know from the von Neumann ergodic theorem (and the fact that T is a unitary operator on $L^2(X)$) that the averages are convergent in $L^2(X)$ norm. But they are also uniformly bounded in $L^\infty(X)$. So the claim follows from the log-convexity of L^p norms (for $p \geq 2$) or by Hölder's inequality and the finite measure of X (for $p < 2$). ∎

Problem 2.8. Show by example that the mean ergodic theorem fails for $p < 1$ and for $p = \infty$.

Now we study the pointwise convergence problem. The key quantitative estimate needed is the following analogue of the Hardy-Littlewood maximal inequality.

Theorem 2.9 (Hardy-Littlewood maximal inequality for measure-preserving systems)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. Then we have*

$$\| \sup_{N>0} \frac{1}{N}\sum_{n=1}^N T^n f\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$$

and

$$\| \sup_{N>0} \frac{1}{N}\sum_{n=1}^N T^n f\|_{L^p(X)} \lesssim_p \|f\|_{L^p(X)}$$

for all $1 < p \leq \infty$.

Proof The claim is trivial when $p = \infty$, so once again the task is to prove the pointwise estimate. By monotone convergence it suffices to show that

$$\| \sup_{1 \leq N < N_0} \frac{1}{N}\sum_{n=1}^N T^n f\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$$

uniformly in $N_0 > 1$.

The idea is to lift up to a space where T can be modeled by the integer shift $n \mapsto n + 1$, at which point we can apply (10). Fix N_0, and let $[2N_0]$ be the finite set $[2N_0] := \{0, \dots, 2N_0 - 1\}$, endowed with the discrete σ-algebra and uniform measure, so that $X \times [2N_0]$ is a probability space. Inside this space we have $X \times [N_0]$, where $[N_0] := \{0, \dots, N_0 - 1\}$. We define the function $F : X \times [2N_0] \to \mathbf{C}$ by

$$F(x, n) := T^n f(x).$$

From (10) we have

$$\| \sup_{1 \le N < N_0} \frac{1}{N} \sum_{n=1}^N F(x, n+m) \|_{l^{1,\infty}_m([N_0])} \lesssim \|F(x,m)\|_{l^1_m([2N_0])}$$

for all $x \in X$; writing out what the $l^{1,\infty}$ norm means, and integrating in X using Fubini's theorem, we conclude

$$\| \sup_{1 \le N < N_0} \frac{1}{N} \sum_{n=1}^N F(x, n+m) \|_{L^{1,\infty}_{x,m}(X \times [N_0])} \lesssim \|F\|_{L^1(X \times [2N_0])}.$$

The right-hand side is just $\|f\|_{L^1(X)}$. As for the left-hand side, observe that

$$\sup_{1 \le N < N_0} \frac{1}{N} \sum_{n=1}^N F(x, n+m) = \sup_{1 \le N < N_0} \frac{1}{N} \sum_{n=1}^N T^n f(T^{-m} x)$$

and so on writing out the $L^{1,\infty}$ norm we see that the left-hand side is comparable to

$$\| \sup_{1 \le N < N_0} \frac{1}{N} \sum_{n=1}^N T^n f(x) \|_{L^{1,\infty}(X)}$$

and the claim follows. ■

Remark 2.10. Note how a maximal inequality for the integers was "transferred" to a maximal inequality on arbitrary measure preserving systems. There are several abstract *transference principles* which generalise this type of phenomenon. In particular, the above argument is in fact a special case of the *Calderón transference principle.*

Now we can present the analogue of the Lebesgue differentiation theorem for measure-preserving systems.

Proposition 2.11 (Pointwise ergodic theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, and let $f \in L^1(X)$. Then the sequence $\frac{1}{N} \sum_{n=1}^N T^n f$ converges pointwise almost everywhere.*

Proof By repeating the argument in the Lebesgue differentiation theorem more or less verbatim (using the above maximal inequality in place of the Hardy-Littlewood inequality) it suffices to verify the claim for a dense subclass of $L^1(X)$, such as $L^2(X)$. Since the $L^2(X)$ norm controls the L^1 norm, it suffices to do so for a dense subclass of $L^2(X)$.

Now we repeat the proof of the von Neumann ergodic theorem. For the invariant part $L^2(X)^T$ of $L^2(X)$, the pointwise convergence is obvious. Also, for functions of the form $Tg - g$ with $g \in L^\infty(X)$, the convergence is also obvious. But these functions are clearly dense in $\{Tg - g : g \in L^2(X)\}$ and hence in $\overline{\{Tg - g : g \in L^2(X)\}}$. Since this space and $L^2(X)^T$ are orthogonal complements of $L^2(X)$, we thus have demonstrated convergence of a dense subclass of $L^2(X)$ and thus of $L^1(X)$, as desired. ■

Remark 2.12. Notice how the strategy of establishing convergence splits into two independent parts - obtaining convergence on a dense subclass, and then establishing some harmonic analysis estimate to pass to the general case. This is not the only way to achieve convergence results. Later on we shall see a variation-norm approach which relies purely on harmonic analysis estimates to obtain convergence (foregoing the dense subclass). In the opposite direction, there are more dynamical approaches (which we do not cover here) which forego the harmonic analysis component of the argument, relying instead on analysing the dynamics of the measure-preserving system by other means (such as measure-theoretic or topological means). It is not fully understood to what extent these different techniques complement each other.

By combining the Lebesgue differentiation theorem with the Radon-Nikodym theorem, one easily obtains

Theorem 2.13 (One-dimensional Radamacher differentiation theorem)**.** *If $f : \mathbf{R} \to \mathbf{C}$ is Lipschitz, then it is differentiable almost everywhere, its derivative lies in $L^\infty(\mathbf{R})$, and we have the fundamental theorem of calculus $f(y) - f(x) = \int_x^y f'(z)\, dz$ for all $x, y \in \mathbf{R}$.*

Proof We may reduce to the case when f is real. The Riemann-Stietjes measure df is easily seen to be absolutely continuous with respect to Lebesgue measure, and is thus of the form $df = gdx$ for some locally integrable g, thus $f(y) - f(x) = \int_x^y g(z)\, dz$ by the Radon-Nikodym theorem. The Lebesgue differentiation theorem then shows that f' exists and is equal to g at every Lebesgue point of g (see Q7), and the claim follows. (The boundedness of $g = f'$ is clear from the Lipschitz nature of f.) ■

Remark 2.14. There are other ways to prove this theorem that do not require Radon-Nikodym differentiation, which we shall encounter later.

This one-dimensional theorem implies a multi-dimensional analogue:

Theorem 2.15 (Radamacher differentiation theorem)**.** *If $f : \mathbf{R}^d \to \mathbf{C}$ is Lipschitz, then f is differentiable almost everywhere, thus for almost every x there exists a vector $\nabla f(x)$ such that*

$$\lim_{h \to 0} \frac{|f(x+h) - f(x) - \nabla f(x) \cdot h|}{|h|} = 0.$$

Proof For simplicity of notation we take $d = 2$, although the general case is similar. Theorem 2.13 (and Fubini's theorem) shows that the partial derivatives $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}$ exist almost everywhere and are bounded. This gives us a candidate gradient ∇f defined almost everywhere. The remaining challenge is to show total differentiability. In other words, for any $\varepsilon > 0$, we need to show that for almost every x, we have

$$f(x+h) - f(x) = \nabla f(x) \cdot h + O(\varepsilon |h|) \tag{13}$$

whenever $|h|$ is sufficiently small depending on x. Note that the control of the partial derivatives only achieve this when h is a multiple of e_1 or e_2.

Fix $\varepsilon > 0$. We will need to make things slightly more quantitative. For any $n \geq 1$ and $i = 1, 2$, let $X_{i,n}$ denote the set

$$X_{i,n} := \{x \in \mathbf{R}^2 : |f(x + h_i e_i) - f(x) - \frac{\partial f}{\partial x_i}(x) h_i| \leq \varepsilon |h_i| \text{ whenever } |h_i| \leq \frac{1}{n}\}.$$

Then the almost everywhere existence of $\frac{\partial f}{\partial x_i}$ implies that the $X_{i,n}$ are measurable and increase to $\mathbf{R}^2$ as $n \to \infty$ (neglecting sets of measure zero). In particular, almost every point lies in infinitely many of the $X_{i,n}$. Because of this, we know that almost every point is a horizontal and vertical Lebesgue point of $\frac{\partial f}{\partial x_i}$ and of $1_{X_{i,n}}$ for infinitely many n. (We say that (x, y) is a horizontal Lebesgue point of F if $⨍_{-r}^{r} |F(x+t, y) - F(x, y)|\, dt \to 0$ as $r \to 0$, and similarly define a vertical Lebesgue point of F.) In particular, for almost every point there is an n such that one lies in $X_{i,n}$, and is a horizontal and vertical Lebesgue point of $\frac{\partial f}{\partial x_i}$ and of $1_{X_{i,n}}$ for $i = 1, 2$. From this it is not hard to show (13) for a large density set of h near x (say of density $1 - \varepsilon^2$), basically by decomposing $f(x + h_1 e_1 + h_2 e_2) - f(x) = [f(x + h_1 e_1) - f(x)] + [f(x + h_1 e_1 + h_2 e_2) - f(x + h_1 e_1)]$ and using the Lebesgue point properties to ensure that $\frac{\partial f}{\partial x_2}(x + h_1 e_1)$ is usually close to $\frac{\partial f}{\partial x_2}(x)$; we omit the details. The remaining exceptional values of h can then be dealt with by locating a nearby non-exceptional value of h and using the Lipschitz property. ■

3. H^p FUNCTIONS

Another classical application of the Hardy-Littlewood maximal function lies in obtaining a satisfactory theory of (complex) H^p functions on the complex disk $\mathcal{D} := \{z : |z| < 1\}$ for $1 < p \leq \infty$. (The theory for $p \leq 1$ is more subtle and will not be dealt with here.)

Let $f : \mathcal{D} \to \mathbf{C}$ be a holomorphic function. From residue calculus we have

$$2f(z) - f(0) = \frac{1}{2\pi i} \int_{|\zeta|=r} \frac{f(\zeta)(\zeta + z)}{\zeta(\zeta - z)} d\zeta$$

and

$$f(0) = \frac{1}{2\pi i} \int_{|\zeta|=r} \frac{f(\zeta)(r^2/\zeta + \bar{z})}{\zeta(r^2/\zeta - \bar{z})} d\zeta$$

when $|z| < r$, and so on averaging, and noting that $\frac{r^2/\zeta + \bar{z}}{r^2/\zeta - \bar{z}}$ is the complex conjugate of $\frac{\zeta + z}{\zeta - z}$ when $|\zeta| = r$, we obtain

$$f(z) = \int_{|\zeta|=r} f(\zeta) \mathrm{Re}(\frac{\zeta + z}{\zeta - z}) \frac{d\zeta}{2\pi i \zeta}$$

or in other words, if we set $f_r : \mathbf{R}/\mathbf{Z} \to \mathbf{C}$ to be the function $f_r(\theta) := f(re^{2\pi i \theta})$, we have the *reproducing formula*

$$f_s = f_r * K_{s/r}$$

whenever $0 \leq s < r < 1$ and K_t is the *Poisson kernel*

$$K_t(\beta) = \mathrm{Re}(\frac{e^{2\pi i \beta} + t}{e^{2\pi i \beta} - t}) = \frac{1 - t^2}{1 + t^2 - 2t \cos 2\pi\beta}.$$

The kernel K_t is clearly non-negative; applying the above identity to $f \equiv 1$ (or computing directly) we see also that K_t has mass 1:

$$\|K_t\|_{L^1(\mathbf{R}/\mathbf{Z})} = \int_{\mathbf{R}/\mathbf{Z}} K_t \, d\theta = 1.$$

From Young's inequality we thus obtain that for $1 \leq p \leq \infty$, the L^p norm of f is increasing on circles:

$$\|f_s\|_{L^p(\mathbf{R}/\mathbf{Z})} \leq \|f_t\|_{L^p(\mathbf{R}/\mathbf{Z})}.$$

(Note that the $p = \infty$ case is just the maximum principle.) In particular, the quantity

$$\|f\|_{H^p(\mathcal{D})} := \lim_{t \to 1^-} \|f_t\|_{L^p_\theta(\mathbf{R}/\mathbf{Z})}$$

always exists. We say that f is an H^p *function* if f is analytic on $\mathcal{D}$ and the norm $\|f\|_{H^p(\mathcal{D})}$ is finite. It is not hard to see that $H^p(\mathcal{D})$ becomes a normed vector space (and that $H^p(\mathcal{D}) \supset H^q(\mathcal{D})$ whenever $p < q$); a little more work shows that it is in fact complete.

Problem 3.1. For any $n \in \mathbf{Z}$ and $g \in L^1(\mathbf{R}/\mathbf{Z})$, define the *Fourier coefficient* $\hat{g}(n)$ of g by the formula

$$\hat{g}(n) := \int_{\mathbf{R}/\mathbf{Z}} g(\theta) e^{-2\pi i n\theta} \, d\theta.$$

Show that

$$\hat{K}_t(n) = t^{|n|}$$

for every $0 \leq t < 1$ and $n \in \mathbf{Z}$, and that

$$\hat{f}_s(n) = (s/r)^{|n|} \hat{f}_r(n)$$

for every $f \in \mathcal{D}$, $0 \leq s < r < 1$ and $n \in \mathbf{Z}$. Also show that $\hat{f}_r(n) = 0$ for all $n \neq 0$, and that

$$K_t * K_{t'} = K_{tt'}$$

for all $0 \leq t, t' \leq 1$. (Note that these identities can be proven either by complex analysis methods or by Fourier analysis methods; it is instructive to prove them both ways and compare results.)

The kernel K_t can be easily verified to obey the bounds

$$|K_t(\theta)| \lesssim \frac{1/(1-t)}{\langle \mathrm{dist}(0,\theta)/(1-t) \rangle^2}.$$

This is enough to obtain the pointwise estimate

$$|f * K_t| \lesssim Mf$$

where

$$Mf(\theta) := \sup_{r > 0} \frac{1}{2r} \int_{\mathrm{dist}(\alpha,\theta) < r} |f(\alpha)| \, d\alpha$$

is the Hardy-Littlewood maximal function of f. From last week's notes we conclude that for any $1 \leq p \leq \infty$ and $f \in L^p(\mathbf{R}/\mathbf{Z})$, that $f * K_t$ converges to f in $L^p(\mathbf{R}/\mathbf{Z})$ norm, and by modifying the proof of the Lebesgue differentiation theorem we also see that it converges pointwise almost everywhere.

The definition of H^p strongly suggests (but does not immediately prove) that if f lies in H^p, then f_t should converge to some sort of limit f_1 as $t \to 1^-$. This is indeed the case:

Theorem 3.2. *Let $f \in H^p(\mathcal{D})$ for $1 < p < \infty$. Then f_t converges both in $L^p(\mathbf{R}/\mathbf{Z})$ norm and pointwise almost everywhere to a limit $f_1 \in L^p(\mathbf{R}/\mathbf{Z})$.*

Remark 3.3. The theorem fails for $p \leq 1$, as can be seen by explicitly computing using the function $f(z) := \frac{1+z}{1-z}$. When $p = \infty$, the pointwise almost everywhere claim is still true, but the convergence is not.

Proof Let us first demonstrate weak convergence. For any $g \in L^{p'}(\mathbf{R}/\mathbf{Z})$ and $0 < t < 1$, we observe that

$$\langle f_r, g * K_t \rangle = \langle f_r * K_t, g \rangle = \langle f_{rt}, g \rangle = \langle f_t * K_r, g \rangle \tag{14}$$

where $\langle f, g \rangle := \int_{\mathbf{R}/\mathbf{Z}} f(\theta)\overline{g(\theta)}\, d\theta$ is the inner product on $\mathbf{R}/\mathbf{Z}$ (here we use the fact that K_t is real and even). Since $f_t * K_r$ converges in L^p norm to f_t as $r \to 1$, we thus see that $\langle f_r, g * K_t \rangle$ converges to a limit as $r \to 1$. Since $g * K_t$ converges in $L^{p'}$ norm to g as $t \to 1$, we conclude (from Hölder's inequality and the uniform L^p bound on f_r) that $\langle f_r, g \rangle$ also converges to a limit as $r \to 1$, thus f_r converges weakly. But as L^p is reflexive, the closed unit ball is also weakly closed, and we conclude that f_r converges weakly to f_1 for some $f_1 \in L^p(\mathbf{R}/\mathbf{Z})$, thus

$$\lim_{r \to 1} \langle f_r, g \rangle = \langle f_1, g \rangle$$

for all $g \in L^{p'}$. Replacing g by $g * K_t$ and using (14) we conclude

$$\lim_{r \to 1} \langle f_t * K_r, g \rangle = \langle f_1, g * K_t \rangle,$$

and hence (since $f_t * K_r$ converges to f_t in L^p norm)

$$\langle f_t, g \rangle = \langle f_1, g * K_t \rangle = \langle f_1 * K_t, g \rangle.$$

This is for all $g \in L^{p'}$, thus we have

$$f_t = f_1 * K_t.$$

Thus f_t converges pointwise almost everywhere and in L^p norm to f_1. ■

4. The Calderón-Zygmund decomposition

For a function $f : X \to \mathbf{C}$ on an abstract measure space, and any threshold $\lambda > 0$ one has the basic decomposition

$$f = f 1_{|f| \leq \lambda} + f 1_{|f| > \lambda}$$

into a "good" $g := f 1_{|f| \leq \lambda}$ piece (bounded by λ), and a "bad" piece $b := f 1_{|f| > \lambda}$ (larger than λ, but at least of small support). If for instance $f \in L^1(X)$, then we have the bounds

$$\|g\|_{L^1(X)} \leq \|f\|_{L^1(X)}; \quad \|g\|_{L^\infty(X)} \leq \lambda$$

for the good piece and

$$\|b\|_{L^1(X)} \leq \|f\|_{L^1(X)}; \quad \mu(\mathrm{supp}(b)) \leq \frac{\|f\|_{L^1(X)}}{\lambda}$$

for the bad piece (the last inequality being Markov's inequality). Thus both pieces inherit the L^1 bound of the whole, but enjoy an additional useful property. This type of decomposition underlies tools such as the layer cake decompositions and the real interpolation method.

For arbitrary measure spaces X with no additional structure, this type of decomposition is about the best one can do to split f into "good" and "bad" parts. But if f has more structure - in particular some sort of metric or dyadic structure compatible with the measure - then one can do better, in particular one can select the bad piece b to be "locally oscillating" in a certain sense. This principle, which uses the same ideas that underlie the Hardy-Littlewood maximal inequality, is formalised in the *Calderón-Zygmund decomposition*, a fundamental tool in the Calderón-Zygmund theory of singular integrals.

To motivate the lemma let us first give a simple one-dimensional version.

Lemma 4.1 (Rising sun lemma)**.** *Let $I \subset \mathbf{R}$ be a bounded interval, and let $f : I \to \mathbf{R}$ be integrable, and let $\lambda \in \mathbf{R}$ be such that $⨍_I f \leq \lambda$. Then there exist an at most countable family $(I_\alpha)_{\alpha \in A}$ of disjoint open intervals in I such that*

- *$f(x) \leq \lambda$ for almost every $x \in I \backslash \bigcup_{\alpha \in A} I_\alpha$;*
- *$⨍_{I_\alpha} f = \lambda$ for every $\alpha \in A$.*

Proof By subtracting λ from f we may normalise $\lambda = 0$. We may take I to be half-open $I = [x_-, x_+)$. Define $F : I \to \mathbf{R}$ by $F(x) := -\int_{x_-}^x f(y)\, dy$. Then F is a continuous function which starts at zero and ends up positive. Define a "maximal" version $F^* : I \to \mathbf{R}$ of F by $F^*(x) := \sup_{x_- \leq y \leq x} F(y)$, then F^* is continuous non-decreasing which starts at zero and ends up positive. Let $\Omega := \{x \in I : F^*(x) > F(x)\}$, then Ω is an open subset of I, thus $\Omega = \bigcup_{\alpha \in A} I_\alpha$ for some at most countable set of intervals I_α. At the endpoints of each interval we have $F^*(x) = F(x)$, while on the interior of these intervals F^* is necessarily constant (why?). This gives the second desired property. For the first, one observes from Q7 that at almost every point where $f(x) < 0$, we have F differentiable with $F' < 0$, which implies that $x \in \Omega$, and the claim follows. ■

As a corollary we have

Corollary 4.2 (One-dimensional rising sun lemma)**.** *Let the hypotheses be as in the above lemma. Then we have a decomposition $f = g + \sum_{\alpha \in A} b_\alpha$, where $g : I \to \mathbf{R}$ is bounded above by λ with $\int_I g = \int_I f$, and for each $\alpha \in A$, b_α is supported in I_α and has mean zero. Furthermore, if $\lambda > 0$ and f is non-negative, then $\sum_\alpha |I_\alpha| \leq \frac{\|f\|_{L^1(I)}}{\lambda}$ and $\|b_\alpha\|_{L^1} \leq 2\lambda |I_\alpha|$.*

Proof We set $b_\alpha := (f - \lambda) 1_{I_\alpha}$, and set g equal to f outside of $\bigcup_\alpha I_\alpha$ and equal to λ inside $\bigcup_\alpha I_\alpha$. To prove the last two claims, we have

$$\int_{I_\alpha} f = \lambda |I_\alpha|$$

so on summing and using the non-negativity of f

$$\int_I f \geq \lambda \sum_\alpha |I_\alpha|$$

from which the former claim follows. Finally, from the definition of b_α and the triangle inequality

$$\|b_\alpha\|_{L^1} \leq \int_{I_\alpha} f + \int_{I_\alpha} \lambda = 2\lambda |I_\alpha|$$

which is the latter claim. ∎

The proof of the rising sun lemma relies crucially on the ordered nature of the real line. For more general spaces, such as Euclidean spaces, it is preferable to use arguments that rely instead on metric or dyadic structure. We begin with a dyadic version.

Proposition 4.3 (Dyadic Calderón-Zygmund decomposition)**.** *Let $f \in L^1(\mathbf{R}^d)$ and $\lambda > 0$. Then there exists a decomposition $f = g + \sum_{Q \in \mathcal{Q}} b_Q$, where the "good" function g obeys the bounds*

$$\|g\|_{L^1(\mathbf{R}^d)} \leq \|f\|_{L^1(\mathbf{R}^d)}; \quad \|g\|_{L^\infty(\mathbf{R}^d)} \leq 2^d \lambda,$$

Q ranges over a disjoint family $\mathcal{Q}$ of dyadic cubes, and each b_Q is supported on Q, has mean zero, and has the L^1 bound

$$\|b_Q\|_{L^1(\mathbf{R}^d)} \leq 2^{d+1} \lambda |Q|.$$

Furthermore, we have the inclusions

$$\bigcup_{Q \in \mathcal{Q}} Q = \{M_\Delta f > \lambda\} \subset \{Mf \gtrsim \lambda\}.$$

In particular from (3) *we have $\sum_{Q \in \mathcal{Q}} |Q| \leq \frac{\|f\|_{L^1(\mathbf{R}^d)}}{\lambda}$.*

Proof The arguments here will closely resemble that used to prove (3). Let us say that a cube Q is *bad* if $⨍_Q |f| > \lambda$, and *good* otherwise. Call a bad cube Q *maximal* if Q is bad, but no cube strictly containing Q is bad, and let $\mathcal{Q}$ be the collection of all maximal bad cubes, thus $\mathcal{Q}$ is a collection of disjoint dyadic cubes, and we also clearly have $M_\Delta f \geq \lambda$ on each bad cube. From the hypothesis $f \in L^1(\mathbf{R}^d)$ and monotone convergence we see that all cubes with sufficiently large side-length are automatically good. Thus every bad cube is contained in a maximal bad cube.

Now let Q be a maximal bad cube. The parent Q' of Q (i.e. the unique dyadic cube of twice the side-length containing Q) is good, hence $⨍_{Q'} |f| \leq \lambda$. Since $|Q'| = 2^d |Q|$, we conclude that $⨍_Q |f| \leq 2^d \lambda$. Now set

$$b_Q := (f - ⨍_Q f) 1_Q$$

and

$$g = f(1 - 1_{\bigcup_{Q \in \mathcal{Q}} Q}) + \sum_Q (⨍_Q f) 1_Q.$$

All the desired properties are then easily verified, except perhaps for the claim $\|g\|_{L^\infty(\mathbf{R}^d)} \leq 2^d\lambda$. This claim is clear on each separate cube Q; the only difficulty is to show that g is bounded by $2^d\lambda$ outside of $\bigcup_{Q\in\mathcal{Q}} Q$. But by construction we see that if $x \notin \bigcup_{Q\in\mathcal{Q}} Q$ and $Q \ni x$, then Q is good, and $|⨍_Q g| = |⨍_Q f| \leq \lambda$. On the other hand, the (dyadic version of the) Lebesgue differentiation theorem shows that $⨍_Q f$ converges to $f(x)$ for almost every x as $|Q| \to 0$. Thus $|g(x)| \leq \lambda$ for almost every such x, and we are done. ■

Problem 4.4. Let $\mathcal{B}$ be the σ-algebra generated by the cubes in $\mathcal{Q}$, and the Borel subsets of $\mathbf{R}^d \backslash \bigcup_{Q\in\mathcal{Q}} Q$. Show that $g = \mathbf{E}(f|\mathcal{B})$.

Note that the total bad set $\bigcup_{Q\in\mathcal{Q}} Q$ is given by the level set $\{M_\Delta f > \lambda\}$ of the dyadic maximal function. This suggests an alternate approach to the Calderón-Zygmund decomposition, in which one starts by identifying the total bad set (in this case, $\{M_\Delta f > \lambda\}$), and then decomposes it into cubes or balls. A prototype decomposition of this type is

Proposition 4.5 (Dyadic Whitney decomposition)**.** *Let $\Omega \subsetneq \mathbf{R}^d_+$ be an open set. Then there exists a decomposition $\Omega = \bigcup_{Q\in\mathcal{Q}} Q$, where Q ranges over a family $\mathcal{Q}$ of disjoint dyadic cubes, and for each Q in this family $\mathcal{Q}$, the parent Q' of Q is not contained in Ω.*

Proof Define $\mathcal{Q}$ to be the set of all dyadic cubes Q in Ω which are maximal with respect to set inclusion; the claim follows from the nesting property. (The condition $\Omega \subsetneq \mathbf{R}^d_+$ is needed to ensure that every cube is contained in a maximal cube; the open-ness is to get every point of Ω contained in at least one cube.) ■

The property that Q' is not contained in Ω implies the bounds

$$0 \leq \operatorname{dist}(Q, \mathbf{R}^d_+ \backslash \Omega) \leq \operatorname{diam}(Q).$$

In many applications we need to complement the upper bound with a non-trivial lower bound. This can be done with a little more effort:

Proposition 4.6 (Whitney decomposition)**.** *Let $\Omega \subsetneq \mathbf{R}^d$ be an open set and let $K \geq 1$. Then there exists a decomposition $\Omega = \bigcup_{Q\in\mathcal{Q}} Q$, where Q ranges over a family $\mathcal{Q}$ of disjoint dyadic cubes, and for each Q in this family $\mathcal{Q}$, we have $\operatorname{dist}(Q, \mathbf{R}^d \backslash \Omega) \sim K \operatorname{diam}(Q)$.*

Proof Let $\mathcal{Q}'$ denote those dyadic cubes Q in Ω such that

$$K \operatorname{diam}(Q) \leq \operatorname{dist}(Q, \mathbf{R}^d \backslash \Omega) \leq 5K \operatorname{diam}(Q).$$

It is not hard to show that these cubes cover Ω; indeed, given any x in Ω, all one needs to do is to locate a cube Q containing x with diameter between $\operatorname{dist}(x, \mathbf{R}^d \backslash \Omega)/4K$ and $\operatorname{dist}(x, \mathbf{R}^d \backslash \Omega)/2K$. The cubes are not disjoint; however if one lets $\mathcal{Q} \subset \mathcal{Q}'$ be those cubes in $\mathcal{Q}'$ which are maximal with respect to set inclusion, then the claim follows from the nesting property. ■

Note that if K is large, the cubes in this above decomposition have the property that nearby cubes Q, Q' (in the sense that $\operatorname{dist}(Q,Q') \lesssim \operatorname{diam}(Q)+\operatorname{diam}(Q')$) have comparable diameter, thanks to the triangle inequality

$$|\operatorname{dist}(Q,\mathbf{R}^d\backslash\Omega) - \operatorname{dist}(Q',\mathbf{R}^d\backslash\Omega)| \leq \operatorname{dist}(Q,Q') + \operatorname{diam}(Q) + \operatorname{diam}(Q').$$

Since every cube is contained in a ball of comparable radius (with constants depending on d), we then conclude

Proposition 4.7 (Whitney decomposition for balls)**.** *Let $\Omega \subsetneq \mathbf{R}^d$ be an open set and let $K \geq 1$. Then one can cover Ω by balls B such that $\operatorname{dist}(B,\mathbf{R}^d\backslash\Omega) \sim_d K\operatorname{diam}(B)$, and such that each point in Ω is contained in at most $O_d(1)$ balls.*

These decompositions can be used to prove some minor variants of the Calderón-Zygmund decomposition, which we will not describe in detail here.

5. Exercises

- Q1 (Hardy-Littlewood maximal inequality for filtrations) Let $(X,\mathcal{B},\mu)$ be a measure space, and let $\mathcal{B}_n$ be an increasing sequence of σ-finite σ-algebras in $\mathcal{B}$ (thus $\mathcal{B}_n \subset \mathcal{B}_{n+1}$ for all $n \in \mathbf{Z}$). Show that
$$\|\sup_n |\mathbf{E}(f|\mathcal{B}_n)|\|_{L^{1,\infty}(X,\mathcal{B},\mu)} \leq \|f\|_{L^1(X,\mathcal{B},\mu)}$$
and hence
$$\|\sup_n |\mathbf{E}(f|\mathcal{B}_n)|\|_{L^p(X,\mathcal{B},\mu)} \lesssim_p \|f\|_{L^p(X,\mathcal{B},\mu)}$$
for all $1 < p \leq \infty$ and all $\mathcal{B}$-measurable f for which the right-hand side is finite. (Hint: use monotone convergence to reduce to finitely many $\mathcal{B}_n$. Reduce further to the case when the $\mathcal{B}_n$ are countably generated (by using the level sets of the $\mathbf{E}(f|\mathcal{B}_m)$ for rational intervals). Reduce further still to the case where the $\mathcal{B}_n$ are finitely generated, i.e. finite. Now adapt the dyadic argument. There are also simpler arguments which do not require all of these reductions.) This inequality is also known as *Doob's inequality*, and implies in particular that $\mathbf{E}(f|\mathcal{B}_n)$ converges pointwise a.e. to $\mathbf{E}(f|\bigvee_n \mathcal{B}_n)$ whenever $f \in L^1(X,\mathcal{B},\mu)$.
- Q2 (Baby Besicovitch covering lemma) Let $I_1,\ldots,I_N$ be a collection of intervals on the real line. Show that there exist a subcollection $I_{n_1},\ldots,I_{n_k}$ such that $I_{n_1} \cup \ldots \cup I_{n_k} = I_1 \cup \ldots \cup I_N$, and such that every point $x \in \mathbf{R}$ belongs to at most $O(1)$ of the intervals $I_{n_1},\ldots,I_{n_k}$. What is the best explicit bound for $O(1)$ you can get?
- Q3. Show that if $f : \mathbf{R} \to \mathbf{C}$ is supported on $[0,1]$, then
$$\|Mf\|_{L^1([0,1])} \lesssim \|f\|_{L \log L([0,1])}.$$
(Hint: use Q10 from last week's notes.)
- Q4. For any locally integrable $f : \mathbf{R}^d \to \mathbf{C}$, let $M_\Box f$ denote the rectangular maximal function
$$M_\Box f(x) := \sup_{R \ni x} \fint_R |f(y)|\, dy$$

where R ranges over all rectangles with sides parallel to the coordinate axes which contain x. Show that $M_\Box$ is bounded on L^p for all $1 < p \leq \infty$. (Hint: prove by induction, controlling the d-dimensional rectangular maximal function by the $d-1$-dimensional "horizontal" rectangular maximal function, applied to a one-dimensional "vertical" maximal function. A certain amount of application of the Fubini-Tonelli theorem may be needed.) Show by example that $M_\Box$ is *not* of weak-type $(1,1)$ in dimensions $d \geq 2$.

- Q5 (Hedberg's inequality). Let $1 \leq p < \infty$, $0 < \alpha < d/p$, and let f be locally integrable on $\mathbf{R}^d$. Establish *Hedberg's inequality*

$$\int_{\mathbf{R}^d} \frac{|f(y)|}{|x-y|^{d-\alpha}}\ dy \lesssim_{d,\alpha,p} \|f\|_{L^p(\mathbf{R}^d)}^{\frac{\alpha p}{d}} (Mf(x))^{1-\frac{\alpha p}{d}}$$

for all $y \in \mathbf{R}^d$. (Hint: use symmetries to normalise as many quantities as you can, and then divide the integral either dyadically, or into a region near x and a region away from x, estimating the two regions differently.) Use Hedberg's inequality and the Hardy-Littlewood maximal inequality to obtain another proof of the Hardy-Littlewood-Sobolev theorem from last week's notes.
- Q6 (Lebesgue points). Let f be a locally integrable function on $\mathbf{R}^d$. Call a point $x \in \mathbf{R}^d$ a *Lebesgue point* if there exists a number c such that $\lim_{r \to 0} ⨍_{B(x,r)} |f(y) - c|\ dy = 0$. Show that almost every point is a Lebesgue point, and that c is equal to $f(x)$ almost everywhere.
- Q7 (Fundamental theorem of calculus) Let $f : \mathbf{R} \to \mathbf{C}$ be locally integrable, and let $F(x) := \int_0^x f(y)\ dy$ (with the usual convention that $\int_0^{-x} = -\int_{-x}^0$). Show that F is differentiable at every Lebesgue point of f, and that $F' = f$ almost everywhere.
- Q8. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. Let $\mathcal{B}^T = \{E \in \mathcal{B} : T(E) = E\}$ be the elements of $\mathcal{B}$ which are T-invariant (up to sets of measure zero, of course). Show that for any $f \in L^p(\mathcal{B})$ and $1 \leq p < \infty$, the averages $\frac{1}{N}\sum_{n=1}^N T^n f$ converge in L^p norm and pointwise almost everywhere to $\mathbf{E}(f|\mathcal{B}^T)$. In particular, when X is *ergodic* (thus the only invariant sets have zero measure or full measure), conclude that $\frac{1}{N}\sum_{n=1}^N T^n f$ converges pointwise and in L^p norm to $\int_X f\ d\mu$.
- Q9 (Poincare recurrence theorem). Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, and let $f \geq 0$ be non-negative with $\int_X f\ d\mu > 0$. Show that $\int_X f T^n f\ d\mu > 0$ for infinitely many $n \geq 1$.
- Q10 (Heat kernels) For any $t > 0$ and any $f \in L^p(\mathbf{R}^d)$ for some $1 \leq p \leq \infty$, define the *heat kernel* $e^{t\Delta} f$ by

$$e^{t\Delta} f(x) := \frac{1}{(4\pi t)^{d/2}} \int_{\mathbf{R}^d} e^{-|x-y|^2/4t} f(y)\ dy.$$

Show that if $1 \leq p < \infty$ and $f \in L^p(\mathbf{R}^d)$, then $e^{t\Delta} f$ converges both pointwise and in norm to f as $t \to 0$.
- Q11 (Fatou's theorem) Let $f \in H^p(\mathcal{D})$ for some $1 < p \leq \infty$, and let f_1 be as in Theorem 3.2. Show that for almost every $\theta \in \mathbf{R}/\mathbf{Z}$, we have

$$\lim_{n \to \infty} f(z_n) = f_1(\theta)$$

whenever $z_n \in \mathcal{D}$ is a sequence of points converging to $e^{2\pi i\theta}$ *non-tangentially* in the sense that $\angle(e^{2\pi i\theta} - z_n), e^{2\pi i\theta}$ is uniformly bounded away from $\pi/2$. (Hint: first reduce to a fixed angle of non-tangentiality (e.g. all angles less than pi/2 - 1/n), and then build an appropriate "non-tangential maximal function", formed by taking suprema over all points in a sector with apex $e^{2\pi i\theta}$ and this fixed angle of non-tangentiality. Use the kernel bounds to bound this maximal function by the Hardy-Littlewood maximal operator.)

- Q12 Let $f \in L^1(\mathbf{R}^d)$, and let B be a ball such that $Mf \geq \lambda$ at every point of B. Show that $Mf \gtrsim_d \lambda$ at every point of $2B$.
- Q13 (Relationship between dyadic and non-dyadic Hardy-Littlewood maximal inequalities) Let $f : \mathbf{R}^d \to \mathbf{C}$ be locally integrable. Establish the pointwise bound

$$|\{M_\Delta f \geq C_d \lambda\}| \leq |\{Mf \geq \lambda\}| \lesssim_d |\{M_\Delta f \geq c_d \lambda\}|$$

for some $c_d > 0$ depending only on d.

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

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