1. Lecture 4

1.1. Square functions, paraproducts, Khintchin's inequality.

The dyadic Littlewood Paley square function of a function f is defined as

$$Sf(x) := (\sum_I \frac{|\langle f, h_I\rangle|^2}{|I|} 1_I(x))^{1/2}$$

where the summation goes over all dyadic intervals I. To write this square function in more compressed form, let E_k denote the averaging operator on dyadic intervals of length 2^k and $\Delta_k = E_{k-1} - E_k$ the orthogonal projection onto the span of Haar functions at level 2^k. Then

$$Sf = (\sum_I |\langle f, h_I\rangle h_I|^2)^{1/2} = (\sum_k |\Delta_k f|^2)^{1/2}$$

A fairly immediate corollary of the techniques discussed in the last lecture is

Theorem 1.1. *For all* $1 < p < \infty$

$$\|Sf\|_p \leq \|f\|_p$$

Proof: For $p = 2$ this estimate is immediate from Bessel's inequality:

$$\|Sf\|_2^2 = \int \sum_I \frac{|\langle f, h_I\rangle|^2}{|I|} 1_I(x)\, dx$$

$$= \sum_I \frac{|\langle f, h_I\rangle|^2}{|I|} \int 1_I(x)\, dx = \sum_I |\langle f, h_I\rangle|^2 \leq \|f\|_2^2$$

Indeed, the L^2 bound for the square function is essentially a reformulation of Bessel's inequality for the orthonormal basis H_I. Observe how the square function is a way to encode which coefficients play a role at which points x.

To see bounds for $p < 2$ we employ Calderón Zygmund as in the case of Haar multipliers. The important observation is that if b_I is supported on I and has mean zero, then $Sb_I(x) = 0$ for $x \notin I$. Standard CZ decomposition technique then proves a weak type 1 bound for S, which by interpolation yields the L^p bounds for $1 < p < 2$.

To see bounds for $p > 2$ we can consider the sharp function of $(Sf)^2$. Note that for any dyadic interval J we have

$$\int_J (Sf)^2(x) - [(Sf)^2]_J\, dx$$

$$\leq \int_J \sum_{I \subset\neq J} \frac{|\langle f, h_I\rangle|^2}{|I|} 1_I(x)\, dx \leq \|f 1_J\|_2^2$$

To pass form first to second line, we have used that intervals $J \supset J$ only contribute a constant to the integrand, and we have used that for any L^2 function on J we have by the Pythagorean theorem

$$\|g - [g]_J\|_{L^2(J)} \le \|g\|_{L^2(J)}$$

because $g - [g]_J$ and $[g]_J$ are orthogonal. In the second line we have simply used the known result in L^2. We obtain that the sharp function of $(Sf)^2$ is bounded by $(M_2 f)^2$ which proves the desired estimates for $p > 2$. □

Unlike for linear operators such as Haar multipliers, where the bounds $p < 2$ are equivalent for the bounds $p > 2$, in the case of square functions the two regions are independent. However, duality serves to yield converse bounds:

Corollary 1.2. *We have for* $1 < p < \infty$

$$\|f\|_p \lesssim \|Sf\|_p \lesssim \|f\|_p$$

Proof: The second inequality is simply the square function estimate form before. To see the first inequality, we employ duality. Let $g = |f|^p/f$ (defined to be 0 wherever $f(x) = 0$)

$$\|f\|_p^p = \int f(x)g(x)\,dx$$

$$= \sum_I \langle f, h_I\rangle \langle h_I, g\rangle$$

$$= \int \sum_I \Big(\frac{\langle f, h_I\rangle}{\sqrt{|I|}} 1_I(x)\Big)\Big(\frac{\langle h_I, g\rangle}{\sqrt{|I|}} 1_I(x)\Big)\,dx$$

$$\le \int Sf(x)Sg(x)\,dx$$

$$\le \|Sf(x)\|_p \|Sg(x)\|_{p'} \lesssim \|Sf(x)\|_p \|g(x)\|_{p'}$$

Dividing by $\|g\|_{p'} = \|f\|_p^{p-1}$ proves the desired estimate. □

The corollary can not be extended to $p \le 1$s. However, the square function norm can be a good replacement for the norm of the function when extending other inequalities. For example we have

Proposition 1.1 (Burkholder, Davis, Gundy)**.** *For* $0 < p < \infty$ *we have for* $r > 2$

$$\| \|E_k f\|_{V^r(k)} \|_p \lesssim \|Sf\|_p \quad .$$

For $p > 1$, this is simply Lépingle's estimate by the previous corollary. For $p \leq 1$, the inequality is not true with the $\|f\|_p$ on the right-hand-side.

One application of square functions is towards paraproduct theory. The standard dyadic paraproduct is a bilinear operator defined by

$$P(f,g) = \sum_k (E_k f)(\Delta_k g)$$

where E_k and Δ_k are as above. The name paraproduct stems from a decomposition of the point-wise product into two paraproducts plus a simpler term:

$$fg = P(f,g) + P(g,f) + \sum_k (\Delta_k)(f \Delta_k g)$$

To see this identity we use the splitting

$$f = \sum_k \Delta_k f$$

and observe

$$fg = (\sum_k \Delta_k f)(\sum_j \Delta_j g)$$

$$= \sum_{k<j} (\Delta_k f)(\Delta_j g) + \sum_{k>j} (\Delta_k f)(\Delta_j g) + \sum_{k=j} (\Delta_k f)(\Delta_j g)$$

The last three terms are exactly the ones in the claimed splitting of the product. For example we have

$$\sum_k (\Delta_k f) \sum_{j:k<j} \Delta_j g = \sum_k (\Delta_k f)(E_k g) \quad .$$

Sometimes the terms in the paraproduct decomposition are referred to as the high-low , low-high, and low-low terms, indicating which frequencies of f and g are combined.

Lemma 1.3. *For $1 < p, q < \infty$ and $1/r = 1/p + 1/q$ we have*

$$\|P(f,g)\|_r \leq C_{p,q} \|f\|_p \|g\|_q \quad .$$

Note that r may be below 1, we have $1/2 < r < \infty$. The analogous bounds for the point-wise product are simply Hölder's inequality. The low-low term can be estimated

$$\| \sum_k \Delta_k f \Delta_k g \|_r$$

$$\leq \|(Sf)(Sg)\|_r \leq \|Sf\|_p \|Sg\|_q \lesssim \|f\|_p \|g\|_q$$

Here we have used Cauchy Schwarz in the first inequality, and then Hölder's inequality and the square function bounds. These bounds show that the bounds for $P(f,g)$ and $P(g,f)$ are equivalent and we have in this sense symmetry in the two arguments of the paraproduct.

Proof of Lemma:

We first consider the case $r > 1$.

We note that $E_k f$ is constant on each dyadic interval I of length 2^k, while $\Delta_k g$ is a multiple of the corresponding Haar function on each such interval I. Hence the product $E_k f \Delta_k g$ is also a multiple of the corresponding Haar function on I. We then have

$$\Delta_k((E_k f)(\Delta_k g)) = (E_k f)(\Delta_k g)$$

0 and by the (corollary to the) square function estimate

$$\|P(f,g)\|_r \leq \|(\sum_k |E_k f \Delta_k g|)^2)^{1/2}\|_r \quad .$$

We then use the point-wise bound

$$(\sum_k |E_k f \Delta_k g|)^2)^{1/2}$$

$$\leq (\sup_k |E_k f|)(\sum_k |\Delta_k g|)^2)^{1/2} = (Mf)(Sg)$$

The desired estimate now follows by Hölder's inequality and estimates for the Hardy Littlewood maximal function and the square function.

To pass to smaller values of r, we use Calderón Zygmund decomposition in the bilinear setting. Beginning with an estimate

$$L^p \times L^q \to L^r$$

we will deduce a weak type bound

$$L^1 \times L^q \to \mathrm{weak} L^{q/(q+1)}$$

i.e.,

$$|\{x : |P(f,h)| > \lambda\}| \lesssim \lambda^{-q/(q+1)}(\|f\|_1 \|h\|_q)^{q/(q+1)} \quad .$$

We may assume $\|f\|_1 = 1$ and $\|h\|_q = 1$, and let $\lambda > 0$. Split f into good and bad part along the set

$$\{x : Mf(x) > \lambda^{q/(q+1)}\}$$

Then we have by the weak 1 bound for the Hardy Littlewood maximal operator

$$|E| \lesssim \lambda^{-q/(q+1)}$$

We again see $P(b,h)(x) = 0$ for $x \notin E$, hence it sufficies to estimate $P(g,h)$ by the known estimate. By homogeneity, the correct power of λ

has to pop out of the argument. Marcinkiewicz interpolation produces strong type estimates for $p > 1$.

The same argument also works to lower the second exponent. □

Remark: We note in passing the "twisted paraproduct" which is considerably harder to estimate. For functions in two variable, let $E_{k,i}$, $\Delta_{k,i}$ for $i = 1, 2$ denote the operators as above acting only on the i-th variable, while fixing the other variable. Then we have the twisted paraproduct

$$\sum_k E_{k,1} f \Delta_{k,2} g$$

which does satisfy some L^p bounds as in Hölder's inequality, but this is much harder to prove. The currently known proof uses Bellman function techniques discussed further below.

Another application of the square function estimate is Khintchin's inequality. It is the special case of the square function estimate when all Haar coefficients of the same scale are equal.

Corollary 1.4 (Khintchin's inequality)**.** *Assume f is a linear combination of Haar functions supported in $[0, 1)$ and*

$$|\langle f, h_I\rangle| = |\langle f, h_J\rangle| =: a_k 2^{k/2}$$

whenever $|I| = |J| = 2^k$. Then we have for $0 < p < q < \infty$

$$\|f\|_p \sim \|f\|_q$$

Proof: Note that since we are on $[0, 1)$, we have by Hölder

$$\|f\|_p \le \|f\|_q \quad .$$

It remains to show the converse estimate.

We first consider $p > 1$. We have

$$\|f\|_p \sim \|Sf\|_p = (\int |\sum_k |a_k|^2|^{p/2} \, dx)^{1/p} = (\sum_k |a_k|^2)^{1/2}$$

and the right hand side is independent of p.

To pass to $p < 1$, we not the following trick. Let $\alpha + \beta = 1$ and $1/r + 1/s = 1$. Then

$$\int |f|^p \, dx = \int |f|^{\alpha p} |f|^{\beta p} \, dx \le \|f^{\alpha p}\|_r \|f^{\beta p}\|_s = \|f\|_{\alpha r p}^{\alpha p} \|f\|_{\beta s}^{\beta p}$$

Now note that we can have $\alpha r < 1 < \beta s$. Assume we already know equivalence of $\|f\|_p$ and $\|f\|_{\beta s p}$. Then we can conclude

$$\|f\|_p^p \lesssim \|f\|_{\alpha r p}^{\alpha p} \|f\|_p^{\beta p}$$

This is the comparibility of $\|f\|_p$ with $\|f\|_{\alpha rp}$, after division of by the correct power of $\|f\|_p$ and raising to the correct power. This improves the previous range of exponent, and one can reach the whole range $0 < p < q < \infty$.

□

In the martingale picture we discussed before, Khintchin's inequality has an important probabilistic interpretation. We may interpret each Haar function as a coin flip with certain value attached to it. This is a sequence of coin flips, with values dependent on the previous coin flips. If the values of all Haar coefficients at level k are equal, this means that the k-th coin flip is a random variable independent of the previous flips. Thus for a sequence a_k and independent random variables $\epsilon_k \in \{-1, 1\}$ Khintchin's inequality (or rather the proof of it above) becomes

$$E(|\sum \epsilon_k a_k|)^p \sim (\sum_k |a_k|^2)^{p/2} \quad .$$

It is instructive to observe that one can use Khintchine's inequality to deduce bounds for the square function from Haar multiplier bounds. Since we already know these bounds and in fact used the square function bounds to deduce Khintchine's inequality, this is not a proof within our line of reasoning, merely an observation that is used similarly in other contexts.

We consider a random Haar multilier

$$T_\epsilon = \sum_I \epsilon_I \langle f, h_I \rangle h_I$$

where the ϵ_I are independent random variables taking values ± 1 with probability $1/2$ each. L^p bounds for Haar multiliers imply

$$\|T_\epsilon f\|_p^p \lesssim \|f\|_p^p$$

uniformly in ϵ. Integrating over the probability space (taking expectation) gives

$$E(\int |T_\epsilon f|^p)\, dx) \lesssim \|f\|_p^p$$

Butr the left-hand-side is equal to

$$= \int E(|T_\epsilon f|^p)\, dx$$

$$= \int E(|\sum_I \epsilon_I \langle f, h_I \rangle h_I|^p)\, dx$$

$$\sim \int (\sum_I |\langle f, h_I \rangle h_I|^2)^{p/2}\, dx = \|Sf\|_p^p$$

where the $\sim$ comes from Khintchin.

Another very similar application concerns vector valued inequalities. If T is any linear operator that satisfies an a priori bound

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

Then we have the vector valued bound for sequences of functions f_i:

$$\|(\sum |Tf_i|^2)^{1/2}\|_p \lesssim \|(\sum_i |f_i|^2)^{1/2}\|_p$$

For the proof, introduce random signs ϵ_i and observe that the original inequality yields for every instance

$$\|T(\sum \epsilon_i f_i)\|_p \lesssim \|(\sum_i \epsilon_i f_i\|_p$$

Using linearity gand raising to power p

$$\| \sum T(\epsilon_i f_i)\|_p^p \lesssim \|(\sum_i \epsilon_i f_i\|_p^p$$

Integrating over the probability space gives

$$E(\| \sum T(\epsilon_i f_i)\|_p^p) \lesssim E(\|(\sum_i \epsilon_i f_i\|_p^p)$$

This becomes the vector valued inequality with Khintchin.

Finally, random signs in connection with Khintchin are often used to prove counterexamples top a priori bounds for operators.Often one builds up a counterexample from small pieces, which one has to add up to obtain the counterexample. Often one does not have good control of the sign of the little pieces, and in principle the built up of a large quantity may be hampered by cancellation between the pieces. Often the problem is solved by introducing random signs of the pieces, and use Khintchin to see that the expectation of the sum is large. But if the expectation is large, there has to be some instance for which the sum is large. It is called the probabilistic method, note that is does not construct the exact example, but merely shows existence.

1.2. **John Nirenberg inequality, Bellman functions.** An important tool in dyadic analysis is Bellman functions. We demonstrate it with a proof of the following theorem (which alternatively can be proved with a stopping time argument)

Theorem 1.5 (John-Nirenberg)**.** *There are constants C and c such that the following holds. Assume that f is a finite linear combination* *of Haar functions and define*

$$\|f\|_{BMO} := \sup_I \left(\frac{1}{|I|}\int_I |f(x)-[f]_I|^2\,dx\right)^{1/2}$$

If f is supported on J, we have

$$|\{x : |f(x)| > \lambda\}| \le C|I|e^{-c\lambda/\|f\|_{BMO}}$$

Note this inequality is stronger (possibly up to constants) than any weak type p inequality for $p < \infty$.

Proof:

Note that the statement of the proposition is for a specific interval J. The idea of Bellman function proofs is induction by scales, which means that one assumes the inequality for the two dyadic children of J and uses that information to prove the inequality on J.

Induction beginning is usually settled with a soft argument. In our case we can begin the induction with scale smaller than the smallest scale of Haar functions in the linear combination of f, in this case it turns out we only need the induction hypothesis for the function 0.

The subtlety is that such induction argument often (in particular here) does not work with the actual inequality that is desired, rather one has to find a very carefully chosen equivalent inequality for which the induction works. The functional that does the trick or some part of it is called the Bellman function. It is an elaborate art to find a Bellman function. For many problems of this type, i.e. estimating sufficiently symmetric expressions over all dyadic intervals, that if an estimate holds then a Bellman function exists. There is even an optimal Bellman function which will yield optimal constants in the desired bounds. However, it may be very hard to find the Bellman function,

We carefully rephrase the theorm in a way that we can prove it by induction.

Define

$$\eta = 20 - (x-3)^2$$

and note that for $-2 < x < 2$ we have

$$\eta(x) \sim 1, \eta'(x) \sim 1, \eta''(x) \sim -1$$

Lemma 1.6. *There are constanst A and ϵ such that the following holds. Let f be a finite linear combination of Haar functions and assume*

$$\|f\|_{BMO} = 1$$

For each dyadic interval I:

$$|\{x \in I : f(x) - [f]_I > \lambda\}| \le A|I|\eta(\frac{1}{|I|}\|f-[f]_I\|_{L^2}^2)e^{-\epsilon\lambda}$$

$$|\{x \in I : f(x) - [f]_I < -\lambda\}| \leq A|I|\eta(\frac{1}{|I|}\|f - [f]_I\|_{L^2}^2)e^{-\epsilon\lambda}$$

The two inequalities combined give an inequality for

$$|\{x \in I : |f(x) - [f]_I| > \lambda\}| \leq 2A|I|\eta(\frac{1}{|I|}\|f - [f]_I\|_{L^2}^2)e^{-\epsilon\lambda}$$

By the bounds on η this implies John Nirenberg as stated above. The proof of both inequalities is similar, thus we will only discuss the first one.

Proof of Lemma

By scaling and translating we may assume $I = [0, 1)$. We may also assume that $\lambda > 1$ or else the proof is trivial for sufficiently large A, since the set on the left-hand-side is always contained in I.

We may also assume f is supported on I and has mean zero there.

We split

$$f = f_0 + f_l + f_r = \langle f, h_I \rangle h_I + \sum_{J \in I_l} \langle f, h_J \rangle h_J + \sum_{J \in I_r} \langle f, h_J \rangle h_J$$

Pick a so that f_0 equals a on I_l and $-a$ on I_r. It is then clear that

$$|\{x \in I : |f(x)| > \lambda\}|$$

$$\leq |\{x \in I_l : f_l(x) > \lambda/2 - a\}| + |\{x \in I_r : f_r(x) > \lambda/2 + a\}|$$

Appealing to the induction hypothesis, and using monotonicity of η, we obtain

$$|\{x \in I : |f(x)| > \lambda\}| \leq A(|I|/2)[\eta(2\|f_l\|_2^2)e^{-\epsilon(\lambda-a)} + \eta(2\|f_r\|_2^2)e^{-\epsilon(\lambda+a)}]$$

It thus suffices to show

$$\eta(2\|f_l\|_2^2)e^{\epsilon a)} + \eta(2\|f_r\|_2^2)e^{-\epsilon a)} \leq 2\eta(\|f\|_2^2)$$

Using Taylor series on noting $\|f\|_2^2 \leq |I|$ we can estimate the left-hand-side by

$$\leq (\eta(2\|f_l\|_2^2) + \eta(2\|f_r\|_2^2)) + \epsilon a(\eta(2\|f_l\|_2^2) - \eta(2\|f_r\|_2^2)) + C\epsilon^2 a^2$$

$$\leq (\eta(2\|f_l\|_2^2) + \eta(2\|f_r\|_2^2)) + C\epsilon a|\|f_l\|_2^2 - \|f_r\|_2^2| + C\epsilon^2 a^2$$

where the latter was by estimates on difference quotients of η Orthogonality implies

$$\|f\|_2^2 = a^2 + \|f_l\|_2^2 + \|f_r\|_2^2$$

Concavity of η then estimates the previous display by

$$\leq 2\eta(\|f\|_2^2 - a^2) - c|\|f_l\|_2^2 - \|f_r\|_2^2|| + +C\epsilon a|\|f_l\|_2^2 - \|f_r\|_2^2| + C\epsilon^2 a^2$$

where the second term is a result of the concavity of the function η and c can be determined by Taylor expansion of η and using lower bounds on the second derivative. Using again difference quotient estimates for η we may estimate the last display by

$$\leq 2\eta(\|f\|_2^2) - ca^2 - c|\|f_l\|_2^2 - \|f_r\|_2^2|| + +C\epsilon a|\|f_l\|_2^2 - \|f_r\|_2^2| + C\epsilon^2 a^2$$

Choosing now ϵ small enough depending in A and the other various constants we estimate the last display by $2\|f\|_2^2$ as desired. Note that we use the fact

$$bd \leq b^2 + d^2$$

This completes the proof.□