Classical Fourier Analysis: Math 247B, Winter 2000

Christoph Thiele

September 21, 2000

1 The Haar basis

We shall construct a nice orthonormal basis of $L^2(\mathbf{R})$. An interval $[x, y)$ is called dyadic, if it is of the form $[2^k n, 2^k(n+1))$ with integers k and n. In other words, its length is an integral power of 2, and its distance to the origin is a multiple of its length. Each dyadic interval is closed on the lower end and open on the upper end. Let $\mathcal{I}$ denote the set of all dyadic intervals.

We list the following immediate and useful properties of dyadic intervals:

1. For each $x \in \mathbf{R}$ and $k \in \mathbf{Z}$ there is exactly one dyadic interval of length 2^k which contains x.

2. Each dyadic interval is the disjoint union of two dyadic intervals of half the length.

3. If two dyadic intervals have nonempty intersection, then one of them is contained in the other.

If I is a dyadic interval, let I_l and I_u its lower and upper half, i.e., I_l and I_u are again dyadic intervals. For each dyadic interval we define the Haar function h_I by

$$h_I = \frac{1}{\sqrt{I}}\left(1_{I_l} - 1_{I_u}\right) \quad ,$$

where 1_I denotes the characteristic function of I.

Lemma 1 *The set* $\{h_I : I \ \text{ dyadic}\}$ *is an orthonormal basis of* $L^2(\mathbf{R})$.

Proof We have

$$\langle h_I, h_I \rangle = \frac{1}{|I|}\left(\int_{I_l} 1_{I_l}(x)\, dx + \int_{I_u} 1_{I_u}(x)\, dx\right) = 1 \quad ,$$

thus all Haar functions have norm 1. Assume I and J are two different dyadic intervals. If I and J are disjoint, then $\langle h_I, h_J \rangle = 0$ because the supports of h_I and h_J are disjoint. Assume that I and J are not disjoint. By symmetry we can assume that J is contained in I. Since $I \neq J$, we conclude that $|J| < |I|$. Thus $|J| \leq |I_l|$ and $|J| \leq |I_u|$. Since J intersects at least one of I_l and I_u, we have that J is contained in I_l or I_u. Thus

$$\langle h_I, h_J \rangle = \pm \frac{1}{\sqrt{|I||J|}} \left(\int_{J_l} dx - \int_{J_u} dx \right) = 0$$

Thus we have seen that the Haar functions are pairwise orthogonal.

It remains to show that the Haarfunction are a complete orthonormal set. We shall prove more generally that the closure of the linear span of all Haar functions is dense in L^p for all $1 < p < \infty$. Fix p. By Hahn Banach we have to show that every continuous linear functional f on L^p (thus $f \in L^{p'}$) which vanishes on all Haar functions is already 0.

Fix $f \in L^{p'}$ such that $\langle f, h_I \rangle = 0$ for all I.

Let I_0 be any dyadic interval and define c by

$$\int_{I^0} f(x)\, dx = c|I^0| \quad .$$

Assume $|I| = 2^{k_0}$. For $k > 0$ let I^k be the unique dyadic interval of length 2^{k_0+k} which contains I^0. We claim that

$$\int_{I_k} f(x)\, dx = c|I_k|$$

for all $k \geq 0$ and prove this by induction. We conclude from $\langle f, h_{I^k k} \rangle = 0$ that

$$\int_{I_l^k} f(x)\, dx = \int_{I_u^k} f(x)\, dx \quad .$$

Since one of the intervals I_l^k and I_u^k is equal to I_{k-1}, both sides of this equation are equal to $c|I_{k-1}|$, which proves the induction statement for k. By Hölder's inequality we have

$$|c||I_k| = \left| \int_{I_k} f(x)\, dx \right| \leq |I_k|^{\frac{1}{p}} \|f\|_{p'} \quad .$$

By letting k tend to infinity we conclude $c = 0$.

Now let J be any open interval in $\mathbf{R}$. It is the disjoint union of the collection of maximal dyadic intervals in J. (Every point in J is contained in some dyadic interval contained in J, and two maximal such intervals are disjoint because none of them is contained in the other.). Thus we conclude

$$\int_I f(x)\, dx = 0 \quad .$$

By Lebesgue's differentiation theorem, f has an indefinite integral whose derivative is equal to f almost everywhere. But the indefinite integral is constant by the previous observation, so f is zero almost everywhere. Thus f is 0. This finishes the proof of the lemma. ∎

Thus we have also proved the following lemma

Lemma 2 *The linear span of the Haar functions is dense in $L^p(\mathbf{R})$ for $1 < p < \infty$.*

Observe that this statement is not true for $p = 1$, because taking the mean of a function is a continuous operation in $L^1(\mathbf{R})$, and all Haar functions have mean zero, whereas $L^1(\mathbf{R})$ certainly contains functions which do not have mean 0. The statement is also false for $p = \infty$, because all linear combinations of Haar functions are compactly supported, and the compactly supported functions are not dense in $L^\infty(\mathbf{R})$.

The set of Haar functions is also a minimal set of functions such that the closure of their linear span is dense in L^p. Namely, if we remove one Haar function h_I from this set, $h_I \in L^{p'}$ is a nonzero continuous linear functional that vanishes on all other Haar functions, which therefore do not span a dense subspace. In this sense $\{h_I : I \text{ dyadic}\}$ is a basis of L^p.

Let $\mathcal{H}(\mathbf{R})$ denote the linear span of the Haar functions, i.e., the set of all finite linear combinations of Haar functions. Let $f \in L^p(\mathbf{R})$ and assume $f_i \in \mathcal{H}(\mathbf{R})$ converge to f as $i \to \infty$. For each i write

$$f_i = \sum_I a_I^{(i)} h_I \quad ,$$

where $a_I^i \neq 0$ only for finitely many I. Then an easy calculation gives

$$\lim_{i\to\infty} a_I^{(i)} = \lim_{i\to\infty} \left\langle \sum_J a_J^{(i)} h_J, h_I \right\rangle = \langle f, h_I \rangle \quad .$$

Therefore we would like to write

$$f = \sum_I \langle f, h_I \rangle \, h_I \quad .$$

In L^2, this infinite sum is easily interpreted and shown to be well behaved by Hilbert space techniques, but in L^p, convergence of this infinite sum is not so clear, e.g. it may depend on the order of summation. We will come back to this issue later.

2 The Calderon Zygmund decomposition

We begin to prove estimates in L^p involving expansions of Haar functions. The operator T_ϵ in the following lemma is a discrete model for what is called a singular integral operator in Fourier analysis.

Lemma 3 *Let $\{\epsilon_I : I \text{ dyadic}\}$ be a collection of numbers bounded by 1. Then the operator*

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

mapping $\mathcal{H}(\mathbf{R})$ to itself extends to a bounded operator on $L^p(\mathbf{R})$ for all $1 < p < \infty$. More precisely, we have

$$\|T_\epsilon f\|_p \leq C_p \|f\|_p$$

with a constant C_p depending only on p.

The statement of this lemma can be expressed by saying that $\{h_I\}$ is an unconditional basis of L^p.

Proof For $p=2$ the lemma follows trivially from the fact that the Haar functions are an orthonormal basis of $L^2(\mathbf{R})$. Namely, we have

$$\|T_\epsilon f\|_2^2 = \sum_I |\epsilon_I \langle f, h_I\rangle|^2 \leq \sum_I |\langle f, h_I\rangle|^2 = \|f\|_2^2 \quad .$$

We observe that it suffices to prove the lemma for $1<p<2$. Namely, suppose it is true for some $1<p<2$, we conclude by duality and self adjointness of T_ϵ that it is also true for p':

$$\|T_\epsilon f\|_{p'} = \sup_{g\in\mathcal{H}(\mathbf{R}):\|g\|_p=1} \langle T_\epsilon f, g\rangle = \sup_{\|g\|_p=1} \sum_I \langle f, h_I\rangle \langle h_I, g\rangle$$

$$\leq \sup_{\|g\|_p=1} \|f\|_{p'} \left\| \sum_I \langle g, h_I\rangle h_I \right\|_p \leq C_p \|f\|_{p'} \quad .$$

By the Marcinkiewicz interpolation lemma (see next section and a technical remark further down) it suffices to prove the following socalled weak type 1, 1 estimate for the operator T_ϵ:

$$|\{x : |T_\epsilon f(x)| > \lambda\}| \leq C\lambda^{-1}\|f\|_1 \quad .$$

Let $\mathcal{I}'$ be the collection of all maximal dyadic intervals such that

$$\int_I |f(x)|\, dx > \lambda |I| \quad . \tag{1}$$

Then the intervals in $\mathcal{I}'$ are pairwise disjoint, because if two such intervals were not disjoint, one was contained in the other and therefore not maximal. Thus we have

$$\sum_{I\in\mathcal{I}'} |I| \leq \lambda^{-1} \sum_{I\in\mathcal{I}'} |f(x)|\, dx \leq \lambda^{-1} \int_{\mathbf{R}} |f(x)|\, dx \leq \lambda^{-1}\|f\|_1 \quad .$$

By a similar argument, we see that each interval satisfying (**??**) has length bounded by $\lambda^{-1}\|f\|_1$, and is thus contained in a maximal interval satisfying (**??**) (there are not infinite increrasing chains of such intervals).

For each $I \in \mathcal{I}'$ define

$$b_I = \sum_{J\subset I} \langle f, h_J\rangle h_J$$

Observe that b_I is supported in I and has mean 0. If J is a dyadic interval not contained in I, then h_J is constant on I, and therefore $\langle b_I, h_I\rangle = 0$. Thus $T_\epsilon b_I$ is supported in $E = \bigcup_{I\in\mathcal{I}'} I$. Define

$$g = f - \sum_{I\in\mathcal{I}'} b_I \quad ,$$

then it remains to show that

$$|\{x : |T_\epsilon g(x)| > \lambda\}| \leq C\lambda^{-1}\|f\|_1 \quad . \tag{2}$$

We shall prove that $|g|$ is bounded by 2λ. Assume this is correct, then we have

$$\|g\|_2^2 = \int |g(x)|^2 \, dx$$

$$\leq 2\lambda^2 |E| + 2\int_{\mathbf{R}\setminus E} |g(x)||\lambda| \, dx$$

$$\leq 2\lambda \|f\|_1 + 2\int_{\mathbf{R}\setminus E} |f(x)||\lambda| \, dx \leq 4\lambda \|f\|_1 \quad .$$

This gives

$$\|T_\epsilon g\|_2^2 \leq \|g\|_2^2 \leq 4\lambda \|f\|_1 \quad .$$

This implies (**??**) by Chebyshev's inequality, wich states that for any function F and any $0 < p < \infty$

$$|\{x : |F(x)| > \lambda\}| \, \lambda^p \leq \int_{\mathbf{R}} |F(x)|^p \, dx \quad .$$

To see that $|g|$ indeed is bounded by 2λ, let J be any dyadic interval. If J is not contained in any $I \in \mathcal{I}'$, then 1_J is constant on any $I \in \mathcal{I}'$. Therefore $\int_J b_I(x) \, dx$ vanishes for all $I \in \mathcal{I}'$ and we have

$$\left| \int_J g(x) \, dx \right| = \left| \int_J f(x) \, dx \right| \leq \lambda |J|$$

by definition of $\mathcal{I}'$. If J is contained in some $I \in \mathcal{I}'$, then, because g is constant on I and equal to the average of f over I,

$$\left| \int_J g(x) \, dx \right| = \frac{|J|}{|I|} \left| \int_I f(x) \, dx \right| \leq \frac{|J|}{|I|} \int_I |f(x)| \, dx$$

Let I' be the dyadic interval of length $2|I|$ containing I. Then we have, again by definition of $\mathcal{I}'$:

$$\left| \int_J g(x) \, dx \right| \leq \frac{|J|}{|I|} \int_{I'} |f(x)| \, dx \leq \frac{|J|}{|I|} \lambda |I'| = 2\lambda |J| \quad .$$

Thus the average of f over any dyadic interval is bounded by 2λ. Since $f \in \mathcal{H}(\mathbf{R})$, this proves that f is bounded by 2λ.

This completes the proof of this lemma, up to the Marcinkiewicz interpolation theorem which will be proved in the next section.

Technical remark: In order to apply the Marcinkiewicz interpolation operator, we need to extend the operator to a slightly larger space than $\mathcal{H}(\mathbf{R})$. Let $\Delta(\mathbf{R})$ denote the linear span of the characteristic functions of all dyadic intervals. It is easy to see that $\Delta(\mathbf{R}) = \mathcal{H}(\mathbf{R}) + \mathbf{R}1_{[0,1)} + \mathbf{R}1_{[-1,0)}$. We easily observe that for any numbers $\epsilon_I \in [-1, 1]$ the sum

$$\sum_I \epsilon_I \left\langle 1_{[0,1)}, h_I \right\rangle h_I$$

converges pointwise regardless of the order of summation and similar for the function $1_{[-1,0)}$. This defines $T_\epsilon(f)$ for $f \in \Delta(\mathbf{R})$ as a measurable function.

Since $\Delta(\mathbf{R}) \subset L^2(\mathbf{R})$, T_ϵ satisfies the L^2 bound on $\Delta(\mathbf{R})$. We prove that T_ϵ satisfies the weak type $1,1$ estimate on $\Delta(\mathbf{R})$. First we observe

$$\left|T_\epsilon(1_{[0,1)})\right| \leq \sum_{k>0} 2^{-k} 1_{[0,2^k)} \leq \frac{1}{|x|} \quad ,$$

so the weak type estimate is satisfied for functions in the space $\mathbf{R}1_{[0,1)}$. Moreover we have for any $f \in \mathcal{H}(\mathbf{R})$:

$$\left|\left\{|T_\epsilon(f + c1_{[0,1)})| > \lambda\right\}\right| \leq \left|\left\{|T_\epsilon f| > \frac{\lambda}{2}\right\}\right| + \left|\left\{|cT_\epsilon 1_{[0,1)}| > \frac{\lambda}{2}\right\}\right|$$

$$\leq C\lambda^{-1}\|f\|_1 + C\lambda^{-1}c \leq c\lambda^{-1}\max(\|f\|_1, c) \leq C\|f + c1_{[0,1)}\|_1 \quad .$$

Here the last inequality follows from the fact that f has mean zero and thus

$$|c| = \int_0^\infty (f(x) + c1_{[0,1)}(x))\, dx \leq \|f + c1_{[0,1)}\|_1$$

$$\|f\|_1 = 2\int_{cf>0} |f(x)|\, dx \leq 2\int_{cf>0} |f(x) + c1_{[0,1)}(x)|\, dx \leq 2\|f + c1_{[0,1)}\|_1 \quad .$$

Similarly we proceed with functions in $\mathbf{R}1_{[-1,0)}$ and thus conclude the weak type $1,1$ estimate for all of $\Delta(\mathbf{R})$. Thus we can apply the Marcinkiewicz interpolation theorem to the extended operator.

■

Remark: The function g in this proof is called the good function at level λ, and the function $b = f - g = \sum_{I\in\mathcal{I}'} b_I$ is called the bad function. The crucial properties of these functions are that g is bounded and thus the L^2 result can be applied, whereas b is built up from mean 0 functions, which makes the operator T_ϵ act locally on the building blocks of b. The decomposition of f into g and b is called the Calderon-Zygmund decomposition at level λ.

3 Marcinkiewicz interpolation lemma

Suppose $f \in L^p(\mathbf{R})$. Then we have Chebyshev's inequality:

$$|\{x : |f(x)| > \lambda\}|\, \lambda^p \leq \|f\|_p^p \quad .$$

On the other hand, we can write by Fubini

$$\int |f(x)|^p\, dx = \int \int_{0<\lambda<|f(x)|} p\lambda^{p-1} 1\, d\lambda\, dx$$

$$= \int_0^\infty \int_{\{x:|f(x)>\lambda\}} p\lambda^{p-1}\, dx\, d\lambda = p\int_0^\infty \lambda^p\, |\{x : |f(x)| > \lambda\}|\, \frac{d\lambda}{\lambda}$$

If the measure space $[0,\infty]$ with measure $d\lambda/\lambda$ had finite measure, then we could conlude the converse of Tschebyscheff's inequality with soem constant C_p. This ofcourse is not the case, but on the range of power functions, λ^{-1} only fails barely to be integrable at 0 and ∞. This is one of the themes behind the following interpolation lemma.

Lemma 4 *Let T be a sublinear operator mapping $\Delta(\mathbf{R})$ to the set of measurable functions on $\mathbf{R}$. (Sublinear means that*

$$|T(f_1 + f_2)(x)| \leq |T(f_1)(x)| + |T(f_2)(x)|$$

for almost every x.) Let $1 \leq p_< < p_> < \infty$ and assume that there is a constnat C such that for $j \in \{<,>\}$ we have

$$|\{x : Tf(x) > \lambda\}| \leq C\lambda^{-p_j} \|f\|_{p_j}^{p_j} \quad .$$

Then T extends to a bounded operator from L^p to L^p for every $p_< < p < p_>$, i.e.,

$$\|Tf\|_p \leq C_p \|f\|_p$$

with a constant C_p depending only on C, $p_<$, $p_>$, and p.

Remark: the constant C_p in general tends to ∞ as p tends to $p_<$ or $p_>$.

Proof

By multiplying T with a constant we can assume $C = 1$ for $j \in \{<,>\}$. Let $f \in \Delta(\mathbf{R})$ and assume $\|f\|_p = 1$.

The idea of the proof is to decompose f into small part and large part and control the small part by the $p_>$- norm and the large part by the $p_<$- norm.

Thus, for each λ define

$$f_> = \max(f, \lambda) \quad ,$$

$$f_< = f - f_> \quad .$$

Then we have

$$p_>^{-1} \|f_>\|_{p_>}^{p_>} = \int_0^\infty \kappa^{p_>} \, |\{|f_>| > \kappa\}| \, \frac{d\kappa}{\kappa} = \int_{\lambda > \kappa} \kappa^{p_>} \, |\{|f| > \kappa\}| \, \frac{d\kappa}{\kappa}$$

$$p_<^{-1} \|f_<\|_{p_<}^{p_<} \leq \int_0^\infty (\kappa + \lambda)^{p_<} \, |\{|f_<| > \kappa\}| \, \frac{d\kappa}{\kappa + \lambda} \leq \int_{\lambda < \kappa} \kappa^{p_<} \, |\{|f| > \kappa\}| \, \frac{d\kappa}{\kappa}$$

We do the decomposition

$$|\{Tf(x) > \lambda\}| \leq \sum_{j \in \{<,>\}} \left| \left\{ x : Tf_j > \frac{\lambda}{2} \right\} \right| \quad .$$

Each term on the right hand side is estimated by (here C_p depends on p, $p_>$, $p_<$)

$$2^{p_j} \lambda^{-p_j} \|f_j\|_{p_j}^{p_j}$$

$$\leq C \int_{\kappa j \lambda} \lambda^{-p_j} \kappa^{p_j} |\{|f| > \kappa\}| \, \frac{d\kappa}{\kappa} \quad .$$

Thus we have

$$\|Tf\|_p^p = p \int \lambda^p |\{|Tf| > \lambda\}| \frac{d\lambda}{\lambda} \quad .$$

$$\leq C_p \sum_j \int \int_{\lambda j \kappa} \lambda^{p-p_j} \kappa^{p_j} |\{|f| > \kappa\}| \frac{d\lambda}{\lambda} \frac{d\kappa}{\kappa} \quad .$$

$$\leq C_p \sum_j \int \kappa^p |\{|f| > \kappa\}| \frac{d\kappa}{\kappa} \quad .$$

$$= C_p \|f\|_p^p \quad .$$

This proves the lemma.
■

4 The Littlewood Paley Square function

The Calderon-Zygmund type Lemma **??** which we proved earlier is closely related to the following lemma. Indeed, if we estimate $\|T_\epsilon f\|_p$ and $\|f\|_p$ as in Lemma **??** from above and below by the expression given by the following lemma, then boundedness of T_ϵ follows immediately. Conversely, we will se in the next section, that Lemma **??** can be deduced from Lemma **??** with the aid of Khintchine's inequality. However, we prove Lemma **??** directly, because we plan to deduce Khintchine's inequality from it.

Lemma 5 *Let $f \in \mathcal{H}(\mathbf{R})$ and $1 < p < \infty$. Then*

(3)
$$\frac{1}{C}\|f\|_p \leq \left\| \left(\sum_I \frac{|\langle f, h_I \rangle|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p \leq C\|p\|$$

In view of this lemma we define the square function Sf of f by

$$Sf = \left(\sum_I \frac{|\langle f, h_I \rangle|^2}{|I|} 1_I \right)^{\frac{1}{2}} \quad .$$

The lemma states that the L^p norms of f and Sf are comparable.
Proof For $p = 2$ we have

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

We observe that it suffices to prove the second inequality in (**??**) for all $1 < p < \infty$, because then we obtain

$$\|f\|_p = \sup_{g \in \mathcal{H}(\mathbf{R}) : \|g\|_{p'} = 1} \langle f, g \rangle = \sup_g \sum_I \langle f, h_I \rangle \langle h_I, g \rangle$$

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

$$\le \sup_g \left\| \left(\sum_I \frac{|\langle f, h_I \rangle|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p \left\| \left(\sum_I \frac{|\langle g, h_I \rangle|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_{p'}$$

$$\le C \|Sf\|_p$$

We prove the second inequality in (**??**). For $1 < p < 2$ the proof is again done by interpolation with a weak type $1,1$ estimate, which itself is proved by a Calderon Zygmund decomposition. The essential property of S that we need for this is that if b is supported on an interval I and has mean zero, then Sb is supported on I. This is obviously the case. We shall omit further details.

In the proof of Lemma **??** we used duality to pass to estimates for $2 < p < \infty$. Since taking the square function is not a linear operation, this idea is not immediately applicable here. One could bypass the nonlinearity by interpreting the square function as the norm of a vector valued linear operator. We shall take a different route and prove the estimate for $2 < p < \infty$ directly.

Let $f \in \Delta(\mathbf{R})$. It is easy to see that the sum defining f converges pointwise absolutely (and in L^2) to a function in L^2, which is constant on small dyadic intervals. Define

$$E_\lambda := \{x : Sf(x) > \lambda\}$$

and let $\mathcal{I}_\lambda$ be the collection of maximal dyadic intervals contained in E_λ. Let γ be a small constant to be determined and call a $I \in \mathcal{I}_\lambda$ exceptional if

$$\|f 1_I\|_2 \le \gamma \lambda |I|^{\frac{1}{2}} \quad .$$

If I is not exceptional, we have with Hölder's inequality:

$$\gamma \lambda |I|^{\frac{1}{p}} \le \|f 1_I\|_p \quad .$$

Assume $I \in \mathcal{I}_\lambda$ is exceptional and let I' be the dyadic interval of length $2|I|$ containing I. Then we have for $x \in I$

$$Sf(x)^2 = \sum_{J \subset I} \frac{|\langle f, h_J \rangle|^2}{|J|} 1_J(x) + \sum_{I' \subset J} \frac{|\langle f, h_J \rangle|^2}{|J|} 1_J(x)$$

$$\le \sum_{J \subset I} \frac{|\langle f 1_I, h_J \rangle|^2}{|J|} 1_J(x) + \lambda^2$$

The last inequality follows because $I' \not\subset E_\lambda$. This gives

$$|\{x \in I : Sf(x) > 2\lambda\}|$$

$$\le \left| \left\{ x \in I : \sum_{J \subset I} \frac{|\langle f, h_J \rangle|^2}{|J|} 1_J(x) > \lambda^2 \right\} \right| \le \frac{\|f 1_I\|_2^2}{\lambda^2} \le \gamma^2 |I|$$

This gives

$$|E_{2\lambda}| \le \sum_{I \in \mathcal{I}_\lambda, I \text{ not exc.}} |I| + \sum_{I \in \mathcal{I}_\lambda, I \text{ exc.}} \gamma^2 |I|$$

(4) $$\leq \frac{\|f\|_p^p}{\gamma^p \lambda^p} + \gamma^2 |E_\lambda|$$

Assume we have for certain λ that

(5) $$|E_\lambda| \leq 8 \cdot 2^p \gamma^p \frac{\|f\|_p^p}{\lambda^p} \quad ,$$

then (**??**) shows that (**??**) also holds for λ replaced by 2λ. However for very small λ inequality (**??**) is satisfied because, for fixed $f \in \Delta(\mathbf{R})$, we have

$$|E_\lambda| \leq \frac{\|f\|_2^2}{\lambda^2} \quad ,$$

which for small λ turns into (**??**).

Inequality (**??**) is a weak type p, p inequality for the sublinear operator S. By Marcinkiewicz interpolation we obtain boundedness of S in L^q for $2 < q < p$. Since $p < \infty$ was arbitrary, this finishes the proof. ■

At this point we return to the question of expanding arbitrary L^p functions into Haar series. Lemma **??** has the following striking consequences.

Let a_I be a collection of numbers and let $\mathcal{I}_n$ be an increasing family of finite subsets of $\mathcal{I}$ with $\bigcup_n \mathcal{I}_n = \mathcal{I}$. Then the sequence

(6) $$\left\| \left(\sum_{I \in \mathcal{I}_n} \frac{|a_I|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p$$

is monotone increasing. Assume the limit is finite. By Lebesgue dominated convergence theorem we find a function $S \in L^p$ with

$$S = \lim_n S_n = \lim_{n \to \infty} \left(\sum_{I \in \mathcal{I}_n} \frac{|a_I|^2}{|I|} 1_I \right)^{\frac{1}{2}} \quad .$$

Define

$$E_{\epsilon,n} := \{x : S_n(x) < (1 - \epsilon) S(x)\}$$

For fixed $0 < \epsilon < \frac{1}{2}$, the sets $E_{\epsilon,n}$ shrink to 0 as $n \to \infty$. Hence

$$\|S 1_{E_{\epsilon,m}}\|_p \leq \epsilon$$

for all m greater than or equal to some appropriate n. Then we have for all $n < m < m'$

$$\left\| \sum_{I \in \mathcal{I}_{m'}} a_I h_I - \sum_{I \in \mathcal{I}_m} a_I h_I \right\|_p = \left\| \sum_{I \in \mathcal{I}_{m'} \setminus \mathcal{I}_m} a_I h_I \right\|_p$$

$$\leq \left\| \left(\sum_{I \in \mathcal{I}_{m'} \setminus \mathcal{I}_m} \frac{|a_I|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p$$

$$\leq \left\| (S^2 - S_n^2)^{\frac{1}{2}} 1_{E^c_{\epsilon,n}} \right\|_p + \|S 1_{E_\epsilon,n}\|_p$$

$$\leq \|(1-(1-\epsilon)^2)^{\frac{1}{2}} S\|_p + \epsilon \leq 2\epsilon^{\frac{1}{2}} \|S\|_p + \epsilon$$

This proves that

$$f_m = \sum_{I \in \mathcal{I}_m} a_I h_I$$

is a Cauchy sequence in L^p. Let f be its limit, then we have for each $I \in \mathcal{I}$:

$$\langle f, h_I \rangle = \lim_m \langle f_m, h_I \rangle = a_I \quad ,$$

hence the limit f is what one expects and does not depend on the choice of the sequence $\mathcal{I}_n$. Vice versa, let $f \in L^p(\mathbf{R})$. We aim to prove that with $a_I := \langle f, h_I \rangle$ the sequence (**??**) has a finite limit. Let f_n be a sequence of functions in $\mathcal{H}(\mathbf{R})$ such that $\|f - f_n\|_p \leq 2^{-n}$. Set $a_{I,n} = \langle f_n, h_I \rangle$. Then we have

$$\left\| \left(\sum_I \frac{|a_{I,n}|^2}{|I|} 1_I(x) \right)^{\frac{1}{2}} \right\|_p$$

$$\leq \left\| \left(\sum_I \frac{|a_{I,0}|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p + \sum_{k=1}^{n} \left\| \left(\sum_I \frac{|a_{I,k} - a_{I,k-1}|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p$$

$$\leq C\|f_0\|_p + C \sum_{k=1}^{n} \|f_k - f_{k-1}\|_p \leq C \sum_{k=0}^{n} 2^{-k} \leq C \quad .$$

Since obviously $\lim_n a_{I,n} = a_I$, we easily see that for every finite subset $\mathcal{I}' of \mathcal{I}$ we have for sufficiently large n that $\mathcal{I}' \subset \mathcal{I}_n$ and

$$\left\| \left(\sum_{I \in \mathcal{I}'} \frac{|a_I|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p$$

$$\leq 1 + \left\| \left(\sum_{I \in \mathcal{I}_n} \frac{|a_I|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p \leq C \quad .$$

This proves (**??**) has a finite limit.

In particular, we see that if $f \in L^p$, then $\sum_I \langle f_I, h_I \rangle h_I$ converges to f in L^p regardless of the order of summation.

5 Khintchine's inequality

Define the k-th Rademacher function on the interval $[0, 1)$ by

$$r_0 = 1_{[0,1)}$$

$$r_k = \sum_{I\in\mathcal{I}:I\subset[0,1):|I|=2^{1-k}} |I|^{\frac{1}{2}} h_I$$

Thus the Rademacher function r_k has modulus 1 on the interval $[0,1)$ and oscillates between 1 and -1, taking each value on intervals of length 2^{-k}.

Lemma 6 *Let $(a_k)_{k\geq 0}$ be a sequence of coefficients with*

$$\sum_k |a_k|^2 < \infty$$

Then $\sum_{k=0}^{\infty} a_k r_k$ converges in L^p for all $0 < p < \infty$ and we have

$$\frac{1}{C_p}\left(\sum_k |a_k|^2\right)^{\frac{1}{2}} \leq \left\|\sum_k a_k r_k\right\|_p \leq C_p \left(\sum_k |a_k|^2\right)^{\frac{1}{2}}$$

for all $0 < p < \infty$.

Remark: Conversely, assume we didnt know that $(a_k)_{k=0}^{\infty}$ is square summable, but $\lim_{n\to\infty}\sum_{k=0}^{n} a_k r_k$ converges to a function in L^p, $0<p<\infty$. Then we have for each n:

$$\left(\sum_{k=0}^{n} |a_k|^2\right)^{\frac{1}{2}} \leq C_p \left\|\sum_{k=0}^{n} a_k r_k\right\|_p \leq C_p \quad ,$$

hence (a_k) is square summable.

Proof Since the Rademacher functions are an orthonormal set of functions in $L^2(\mathbf{R})$ (beware: not complete), the inequality of the lemma holds in the case $p=2$ with $C_2 = 1$. Define

$$f = \sum_k a_k r_k \quad .$$

Assume $2 < p < \infty$. Then we have by Hölders' inequality

$$\left(\sum_k |a_k|^2\right)^{\frac{1}{2}} = \|f\|_2 \leq \|f\|_p \quad .$$

To obtain the reverse inequality, we apply Lemma **??** as follows:

$$\|f\|_p \leq \left\|a_0 1_{[0,1)}\right\|_p + \left\|\sum_{k>0}\left(\sum_{|I|=2^{-k},I\subset\{[0,1)} a_k |I_k|^{\frac{1}{2}} h_I\right)\right\|_p$$

$$\leq |a_0| + C\left\|\left(\sum_{k>0}\left(\sum_{|I|=2^{-k},I\subset\{[0,1)} |a_k|^2 1_I\right)\right)^{\frac{1}{2}}\right\|_p = |a_0| + C\left\|\left(\sum_{k>0}|a_k|^2 1_{[0,1)}\right)^{\frac{1}{2}}\right\|_p$$

$$= |a_0| + C\left(\sum_{k>0}|a_k|^2\right)^{\frac{1}{2}} \leq C\left(\sum_{k\geq 0}|a_k|^2\right)^{\frac{1}{2}} \quad .$$

The finiteness of the square function in this line of arguments also implies the convergence of $\sum_k a_k r_k$ in L^p.

Now let $0 < p < 2$. We have again by Hölder

$$\|f\|_p \le \|f\|_2 \le \left(\sum_k |a_k|^2\right)^{\frac{1}{2}} \quad .$$

Hölder's inequality also implies that $\sum_k a_k r_k$ converges in L^p. It remains to show a lower bound for $\|f\|_p$.

Pick $2 < q < \infty$. Since Khintchine's inequality is already proven for q we have

$$\|f\|_2 \le \left\|f 1_{\{x:|f(x)|\le \frac{1}{2}\|f\|_2\}}\right\|_2 + \left\|f 1_{\{x:|f(x)|\ge \frac{1}{2}\|f\|_2\}}\right\|_2$$

$$\le \frac{1}{2}\|f\|_2 + \|f\|_q \left|\left\{x : |f(x)| \ge \frac{1}{2}\|f\|_2\right\}\right|^{\frac{1}{2}-\frac{1}{q}}$$

$$\le \frac{1}{2}\|f\|_2 + C\|f\|_2 \left|\left\{x : |f(x)| \ge \frac{1}{2}\|f\|_2\right\}\right|^{\frac{1}{2}-\frac{1}{q}}$$

This implies

$$\frac{1}{C} \le \left|\left\{x : |f(x)| \ge \frac{1}{2}\|f\|_2\right\}\right| \le C\frac{\|f\|_p^p}{\|f\|_2^p}$$

And thus

$$\|f\|_2 \le C\|f\|_p \quad .$$

■

Vice versa, we can use Khintchine's inequality to deduce L^p bounds for the square function operator from Lemma **??**.

Namely, let $f \in \mathcal{H}(\mathbf{R})$ and $1 < p < \infty$. Let $I_1, I_2, \ldots$ be an enumeration of $\mathcal{I}$. Then Lemma **??** implies that for each $t \in [0,1)$

$$\left\|\sum_{n=1}^{\infty} r_n(t) \langle f, h_{I_n}\rangle h_{I_n}\right\|_p \le C_p\|f\|_p$$

and hence

$$\int_0^1 \left\|\sum_{n=1}^{\infty} r_n(t) \langle f, h_{I_n}\rangle h_{I_n}\right\|_p^p dt \le C_p^p\|f\|_p^p \quad .$$

On the other hand we have with Khintchine's inequality

$$\int_0^1 \left\|\sum_{n=1}^{\infty} r_n(t) \langle f, h_{I_n}\rangle h_{I_n}\right\|_p^p dt$$

$$= \int_0^1 \int_0^1 \left|\sum_{n=1}^{\infty} r_n(t) \langle f, h_{I_n}\rangle h_{I_n}(x)\right|^p dt\, dx$$

$$\geq C \int_0^1 \left(\sum_{n=1}^{\infty} |\langle f, h_{I_n} \rangle|^2 |h_{I_n}(x)|^2 \right)^{\frac{p}{2}} dx$$

$$= C\|Sf\|_p^p$$

This proves the square function estimate.

Observe that although Lemma **??** does not hold in the case $p = 1$, we can make the same argument as above for $p = 1$. Namely, if $f \in \mathcal{H}(\mathbf{R})$ and $\|T_\epsilon f\|_1 \leq C\|f\|_1$ for all choices $\epsilon_I \in [-1, 1]$, then

$$\|Sf\|_1 \leq C'\|f\|_1$$

for a possibly different constant C'. The L^1 norm of the square function will be further discussed in the following section about the Hardy space $H^1(\mathbf{R})$

Finally, we point out the following probabilistic interpretation of expressions of the type $f = \sum_{k=1}^n a_k r_k$. If for each $1 < k < n$ we flip a coin whether to add or subtract the number a_k, then each possible event of this "Markov process" corresponds to an interval of length 2^{-n} in $[0, 1)$, on which f takes the value of the random sum corresponding to the event.

6 The Hardy space H^1 and BMO

The square function estimates show that we can identify the space $L^p(\mathbf{R})$ with the space H_p of all sequences $(a_I)_{I\in\mathcal{I}}$ which satisfy

$$\|(a_I)_{I\in\mathcal{I}}\|_{H^p} := \left\| \left(\sum_I \frac{|a_I|^2}{|I|} 1_I \right)^{\frac{1}{2}} \right\|_p < \infty \quad . \tag{7}$$

More precisely L^p and H^p are equivalent normed spaces if H^p is equipped with the norm $\|.\|_{H^p}$.

Since the square function estimates break down at $p = 1$ and $p = \infty$, we do no expect such an equivalence to be true for $p = 1$ or $p = \infty$.

However one can study the spaces H^1 and H^∞ of all sequences $(a_I)_{I\in\mathcal{I}}$ which satisfy (**??**) with $p = 1$ or $p = \infty$ respectively. Indeed H^1 and H^∞ are normed spaces with norms $\|.\|_{H^1}$ and $\|.\|_{H^\infty}$. (Ofcourse one would like to think of an element $(a_I) \in H^1$ or H^∞ as a function

$$\sum_I a_I h_I \quad ,$$

and this can be done, but at this time the meaning of such an infinite sum has not been clarified.)

For $1 < p < \infty$, the dual space of H^p is $H^{p'}$, as one can see from the identification with L^p and $L^{p'}$. However, it turns out that H^∞ is not the dual space of H^1. The main purpose of the current section is to identify the dual space of H^1 which will be called BMO.

If λ is a continuous linear functional on $H^1(\mathbf{R})$, denote by b_J the number $\lambda((a_I)_{I\in\mathcal{I}})$ where $a_I = 1$ if $I = J$ and $a_J = 0$ otherwise. If $(a_I)_{I\in\mathcal{I}}$ is any sequence with finitely many nonzero entries, we then have

$$\lambda((a_I)_{I\in\mathcal{I}}) = \sum_I a_I b_I$$

Since the space of sequences with finitely many nonzero entries is dense in H^1 (Lebesgue dominated convergence), the sequence $(b_I)_{I\in\mathcal{I}}$ determines the functional λ. Vice versa, if $(b_I)_{I\in\mathcal{I}}$ is any sequence of real numbers, it extends to a bounded linear functional on H^1 if and only if

$$\left|\sum_I a_I b_I\right| \leq C \, \|(a_I)_{I\in\mathcal{I}}\|_{H^1} \tag{8}$$

for all sequences $(a_I)_{I\in\mathcal{I}}$ with finitely many nonzero entries. Thus we can identify the space $(H^1)^*$ with the space of all sequences $(b_I)_{I\in\mathcal{I}}$ satisfying (**??**). The norm

$$\|b_I\|_{(H^1)^*}$$

of an element in the dual space of H^1 is given by the best constant C in (**??**). In what follows we shall give a more direct characterisation of the elements in the space $(H^1)^*$.

Suppose $\lambda = (b_I)_{I\in\mathcal{I}} \in (H^1)^*$. Let J be any dyadic interval and assume a_I vanishes when $I \not\subset J$. Then we have by Cauchy-Schwarz

$$\left|\sum_I a_I b_I\right| \leq \|\lambda\|_{(H^1)^*} \left\| \left(\sum_I \frac{|a_I|^2}{|I|} 1_I\right)^{\frac{1}{2}} \right\|_1$$

$$\leq \|\lambda\|_{(H^1)^*} |J|^{\frac{1}{2}} \left\| \left(\sum_I \frac{|a_I|^2}{|I|} 1_I\right)^{\frac{1}{2}} \right\|_2 = \|\lambda\|_{(H^1)^*} |J|^{\frac{1}{2}} \left(\sum_I |a_I|^2\right)^{\frac{1}{2}}$$

Let $\mathcal{I}'$ be any finite collection of intervals contained in J and set $a_I = \overline{b_I}$ if $I \in \mathcal{I}'$ and $a_I = 0$ otherwise. Then the previous line of inequalities shows

$$\sum_{I\in\mathcal{I}'} |b_I|^2 \leq \|\lambda\|_{(H^1)^*} |J|^{\frac{1}{2}} \left(\sum_I |b_I|^2\right)^{\frac{1}{2}} \quad .$$

Taking the sup over all sets $\mathcal{I}'$ gives

$$|J|^{-\frac{1}{2}} \left(\sum_{I\in\mathcal{I}'} |b_I|^2\right)^{\frac{1}{2}} \leq \|\lambda\|_{(H^1)^*} \quad . \tag{9}$$

Thus a necessary condition for a sequence $(b_I)_{I\in\mathcal{I}}$ to be in $(H^1)^*$ is that the left hand side of (**??**) is bounded by a constant independently of J. The following lemma states that boundedness of the left hand side of (**??**) independently of J is also sufficient for the sequence $(b_I)_I$ to be in $(H^1)^*$.

Lemma 7 *A sequence* $(b_I)_{I\in\mathcal{I}}$ *is an element of* $(H^1)^*$ *if and only if*

$$\sup_{J\in\mathcal{I}} |J|^{-\frac{1}{2}} \left(\sum_{I\subset J} |b_I|^2\right)^{\frac{1}{2}} < \infty \quad . \tag{10}$$

The quantity on the left hand side of (??) is a norm equivalent to the $(H^1)^*$ *norm.*

We call the quantity on the left hand side of (??) the BMO norm of the sequence $(b_I)_{I\in\mathcal{I}}$ and write

$$\|(b_I)_{I\in\mathcal{I}}\|_{\mathrm{BMO}}$$

for it. The lemma states that the space BMO of alls equences with bounded BMO norm is equivalent to the space $(H^1)^*$ under the identity map.

Proof After the preceding discussion it remains to prove that there is a universal constant C such that for any sequences b_I and any finite sequence a_I we have

$$\left|\sum_I a_I b_I\right| \leq C \left\|a_I\right\|_{H^1} \|(b_I)_I\|_{\mathrm{BMO}} \quad .$$

For each $m \in \mathbf{Z}$ let

$$E_m = \left\{ x : \left(\sum_I \frac{|a_I|^2}{|I|} 1_I(x)\right)^{\frac{1}{2}} \geq 2^m \right\} \quad .$$

Let $\mathcal{I}_m$ be the set of dyadic intervals contained in E_m but not in E_{m+1}, and let $\mathcal{I}_m^{\max}$ be the set of maximal dyadic intervals contained in E_m.

Then we have by Cauchy-Schwarz

$$\left|\sum_I a_I b_I\right| \leq \sum_m \left(\sum_{I\in\mathcal{I}_m} |a_I|^2\right)^{\frac{1}{2}} \left(\sum_I |b_I|^2\right)^{\frac{1}{2}}$$

$$\leq \sum_m \left(\int \sum_{I\in\mathcal{I}_m} \frac{|a_I|^2}{|I|} 1_I(x)\, dx\right)^{\frac{1}{2}} \left(\|(b_J)_J\|_{\mathrm{BMO}}^2 \sum_{I\in\mathcal{I}_m^{\max}} |I_m|\right)^{\frac{1}{2}}$$

$$\leq \|(b_J)_J\|_{\mathrm{BMO}} \sum_m \left(|E_m| 2^{2m}\right)^{\frac{1}{2}} \left(|E_m|\right)^{\frac{1}{2}}$$

Here we have used that

$$\left(\int \sum_{I\in\mathcal{I}_m} \frac{|a_I|^2}{|I|} 1_I(x)\, dx\right)^{\frac{1}{2}} \tag{11}$$

is bounded by 2^{2m+1}. This is clearly true on the complement of E_{m+1}. Let J be a maximal dyadic interval in E_{m+1} and let J' be the dyadic intervals of length $2|J|$ containing J. Then clearly (??) is constant on J', and since J' is not contained in

E_{m+1}, (**??**) is bounded by 2^{m+1} on J'. This shows that (**??**) is bounded by 2^{m+1} on E_{m+1}. Now we easily see that

$$\|(b_J)_J\|_{\mathrm{BMO}} \sum_m \left(|E_m| 2^{2m}\right)^{\frac{1}{2}} (|E_m|)^{\frac{1}{2}} = \|(b_J)_J\|_{\mathrm{BMO}} \sum_m 2^m |E_m|$$

$$\leq 2\|(b_J)_J\|_{\mathrm{BMO}} \int_0^\infty \lambda \left| \left\{ x : \left(\sum_I \frac{|a_I|^2}{|I|} 1_I(x) \right)^{\frac{1}{2}} \geq \lambda \right\} \right| \, d\lambda$$

$$= 2\|(b_J)_J\|_{\mathrm{BMO}} \left\| \left(\sum_I \frac{|a_I|^2}{|I|} 1_I(x) \right)^{\frac{1}{2}} \right\|_1$$

This finishes the proof of the lemma. ■

7 The space BMO as function space

In this section we discuss how to make sense of the expression $\sum_I b_I h_I$ as a function if $(b_I)_I$ is a BMO sequence.

Thus let $(b_I)_I$ be a sequence in BMO. For each $n \geq 0$ define $I_n = [0, 2^n)$ and define a function f_n on I_n by

$$f_n(x) = \sum_{I \subset I_n} b_I h_I(x) - \sum_{I_0 \subset I \subset I_n} b_I |I|^{-\frac{1}{2}}$$

Obviously this sum converges in L^2: the first summand because of the BMO estimate, the second because it is a finite sum. Thus this expression as well as other expressions of this type below are understood to hold for almost exvery x.

If $n < n'$, then f_n and $f_{n'}$ coincide on I_n, because we have on I_n

$$f_{n'}(x)$$

$$= \sum_{I \subset I_n} b_I h_I(x) + \sum_{I_n \subset \neq I \subset I_{n'}} b_I h_I(x) - \sum_{I_0 \subset I \subset I_n} b_I |I|^{-\frac{1}{2}} - \sum_{I_n \subset \neq I \subset I_{n'}} b_I |I|^{-\frac{1}{2}} \quad ,$$

which is equal to $f_n(x)$ because the second and fourth term on the right hand side cancel each other on I_n.

Thus we can define a function f on $[0, \infty)$ by defining unambiguously

$$f(x) = f_n(x)$$

whenever $2^n > x$. Similarly we can define a function on $(-\infty, 0]$, by first defining a sequence of functions on the intervals $[-2^n, 0)$. This defines a function on all of $\mathbf{R}$.

If I is any dyadic interval, define f_I to be the average of f on I. Then we have

$$\|f 1_I - f_I\|_2 \leq |I|^{\frac{1}{2}} \|(b_I)_I\|_{\mathrm{BMO}} \quad . \tag{12}$$

Namely, if $I \subset [0,\infty)$ say, then we have for some sufficiently large n with $I \subset [0,2^n)$

$$f(x) - f_I(x) = f_n(x) - (f_n)_I(x)$$

and

$$f_n(x) = \sum_{J \subset I} b_I h_I + \sum_{I \subset \neq J \subset I_n} b_I h_I - \sum_{I_0 \subset J \subset I_n} b_I \quad .$$

The first term on the right hand side has mean zero on I, whereas the other two terms are constant on I and therefore

$$f_n(x) - (f_n)_I(x) = \sum_{J \subset I} b_I h_I \quad .$$

Thus (**??**) follows form the BMO property of the sequence b_I. The assignment $(b_I)_I \to f$ we have defined is injective, because one easily verifies

$$b_I = \langle f, h_I \rangle \quad ,$$

so if $f = 0$ then $(b_I)_I = 0$.

Vice versa, let g be any function locally in L^2 and satisfying

$$|I|^{-\frac{1}{2}} \| \, \|f - f_I\|_{L^2(I)} \leq C \tag{13}$$

Then we can define a sequence (b_I) by

$$b_I = \langle g, h_I \rangle \quad .$$

Then (**??**) implies that b_I is a BMO sequence. This assignment $g \to (b_I)_I$ is not quite injective, because if $b_I = 0$ for all I, then all we can conclud is that if J is any interval, then

$$\sum_{I \subset J} \langle g, h_I \rangle \, h_I = 0$$

in L^2 sense, and since $\{|J|^{-\frac{1}{2}} 1_J\} \cup \{h_I : I \subset J\}$ is an orthonormal basis of $L^2(J)$ we conclude g is a constant on I in L^2 sense. By letting J being $[0,2^n)$ and $[-2^n,0)$ for larger and larger n we conclude that g coincides with a function

$$c_+ 1_{[0,\infty)} + c_- 1_{(-\infty,0]}$$

almost everywhere.

By the above identifications, we can therfore identify BMO with the space of all locally L^2 functions satisfying (**??**) modulo the space $c_+ 1_{[0,\infty)} + c_- 1_{(-\infty,0]}$. We will be somewhat sloppy and call these equivalence classes of functions themselves BMO-functions.

Obviously every function in L^∞ defines a BMO- function because it is locally in L^2 and satisfies

$$|I|^{-\frac{1}{2}} \| \, \|f - f_I\|_{L^2(I)} \leq |I|^{-\frac{1}{2}} (\|f 1_I\|_2 + \|f_I 1_I\|_2) \leq 2 \|f\|_\infty \quad .$$

Moreover, if $f \in L^\infty$, then also

$$f + c_+ 1_{[0,\infty)} + c_- 1_{(-\infty,0]} \in L^\infty \quad ,$$

for all c_+ and c_-, so $L^\infty/\{c_+ 1_{[0,\infty)} + c_- 1_{(-\infty,0]}\}$ is a continuously imbedded subspace of BMO.

However, a function in BMO does not necessarily come from a function in L^∞, an example is given by the function

$$f(x) = \log|x|$$

Let $I = [a, b)$ be a dyadic interval contained in $[0, \infty)$. We shall prove

$$\|f - f_I\|_{L^2(I)} \leq \|f - f(b)\|_{L^2(I)} \leq C \quad .$$

The first inequality is clear by an orthogonality argument: the constant function is orthogonal to $f - f_I$ in $L^2(I)$ and thus

$$\|f - f_I\|^2_{L^2(I)} = \|f - f(b)\|^2_{L^2(I)} + \|f_I - f(b)\|^2_{L^2(I)} \quad .$$

For the second inequality we shall decompose $I = \bigcup_n I_n$ where I_1 is the right half of I, I_2 is the right half of $I \setminus I_1$ and so on. Then I_n sits between $(1/2)^n b$ and b, hence

$$\log|x| - \log|b| \leq n|\log(1/2)|$$

on I_n. Thus

$$\|f - f(b)\|_{L^2(I_n)} \leq Cn|I_n|^{\frac{1}{2}} = Cn2^{-\frac{n}{2}}|I|^{\frac{1}{2}}$$

Summing over n gives

$$\|f - f(b)\|_{L^2(I)} \leq C|I| \quad .$$

A similar argument holds for intervals contained in $(-\infty, 0]$. This proves that $\log|x|$ is a BMO function. However, $\log|x|$ is not in L^∞ because it is unbounded near 0 and ∞.

To make the point that H^∞ is not the dual of H^1, we present a function which is in L^∞ and thus in BMO, but whose square function is not in L^∞, therefore its sequence of coefficients $\langle f, h_I \rangle$ is not in H^∞. Define

$$f = \sum_{k \in \mathbf{Z}} 1_{[2^{2k}, 2^{2k+1}]} - 1_{[2^{2k+1}, 2^{2k+2})}$$

Then f is a bounded function. Obviously we have

$$f(x) = -f(2x)$$

Therefore an easy scaling argument shows that

$$\left\langle f, h_{[0,2^k)} \right\rangle = (-1)^k c 2^{-\frac{k}{2}} \quad .$$

for a certain constant c. Since f changes sign on the left half of $[0, 2^k)$ but not on the right half of $[0, 2^k)$ it follows that $c \neq 0$. Thus we have for the square function:

$$\left(\sum_I \frac{|\langle f, h_I \rangle|^2}{|I|} 1_I\right) \geq \left(\sum_k c^2 1_{[0,2^k)}\right)^{\frac{1}{2}}$$

(It is actually not hard to see that we have equality here. The right hand side however is not in L^∞, since its values tend to ∞ as x tends to $+0$.

The following definition shows that in the condition (**??**) for BMO functions we could have used any $0 < p < \infty$ instead of $p = 2$. As before we denote by f_I the average of f on I.

Lemma 8 *Let* $f \in \mathrm{BMO}$ *and* I *and interval. Then*

$$\frac{1}{C_p} \|f - f_I\|_{L^p(I)} |I|^{-\frac{1}{p}} \leq \|f - f_I\|_{L^2(I)} |I|^{-\frac{1}{2}} \leq C_p \|f - f_I\|_{L^p(I)} |I|^{-\frac{1}{p}} \tag{14}$$

and

$$E_n = |\{x \in I : |f(x) - f_I| \geq \lambda\}| \leq c_1 e^{\overline{c_2 \|f\|_{\mathrm{BMO}}}^{\lambda}} |I| \tag{15}$$

Proof We first consider the case $p > 2$. Then the second inequality in (**??**) follows by Hölder, therefore it suffices to prove the first inequality.

Fix an interval I. Let Sg denote the square function of g. For any function $g \in L^p(I)$ with mean 0 on I we have

$$\left|\sum_I \langle f, h_I \rangle \langle g, h_I \rangle\right|$$

$$\leq C \|f\|_{\mathrm{BMO}} \|Sg\|_1 \leq C \|f\|_{\mathrm{BMO}} \|Sg\|_p |I|^{1-\frac{1}{p}} \leq C \|f\|_{\mathrm{BMO}} \|g\|_p |I|^{1-\frac{1}{p}}$$

Thus there is an element f' in $L^{p'}(I)$ (which we can assume to have mean zero) such that

$$\|f'\|_{p'} \leq \|f\|_{\mathrm{BMO}} |I|^{\frac{1}{p'}}$$

and

$$\sum_I \langle f, h_I \rangle \langle g, h_I \rangle = \int_I f'(x) g(x)\, dx$$

for all $g \in L^p(\mathcal{I})$. Since for all $g \in L^2(x)$ we have

$$\sum_I \langle f, h_I \rangle \langle g, h_I \rangle = \int_I (f(x) - f_I) g(x)\, dx$$

we conclude that $f - f_I$ coincides with f' on I. This proves the first inequality in (**??**).

For $0 < p < 2$ the first inequality follows by Hölder, whereas the second inequality follows by a short calculation with Chebysheff's inequality as we have done for Khinchine's inequality. We omit further details.

It remains to prove (**??**). Using (**??**) we have for $p > 2$

$$|\{x \in I : |f(x) - f_I| > \lambda\}| \leq C\lambda^{-p}\|f'\|_p^p \leq C_p\lambda^{-p}\|f\|_{\mathrm{BMO}}^p|I|$$

For small λ, (**??**) simply follows from $p = 2$. So far we have not discussed the constant C_p, but tracking it down we obtain the bound $C_p \leq (Ap)^p$ for some A independent of p. This we leave as an exercise. For large λ we pick p so that $2Ap\|f\|_{\mathrm{BMO}} = \lambda$. Then (**??**) follows again.
■

Inequality (**??**) is called the John-Nirenberg inequality. It clearly implies inequality (**??**).

8 Maximal truncated singular integrals

We consider again the discrete singular integral operators

$$T_\epsilon f = \sum_{I\in\mathcal{I}} \epsilon_I \langle f, h_I\rangle h_I \quad ,$$

which are well defined on L^p for $1 < p < \infty$. Assume we wanted to calculate the value of $f(x)$, then we could write the following series:

$$\lim_{k\to\infty} \sum_{2^{-k}<|I|<2^k} \epsilon_I \langle f, h_I\rangle h_I(x) \quad . \tag{16}$$

Since $h_I(x) \neq 0$ only if $x \in I$, the sum in this expression is really a finite sum. The expression

$$\sum_{2^{-k}<|I|<2^k} \epsilon_I \langle f, h_I\rangle h_I$$

is called a (discrete) truncated singular integral.

Thus we are naturally interested in convergence of the truncated singular integrals. If $1 < p < \infty$ and $f \in L^p$, then all we know so far is $T_\epsilon f \in L^p$. Thus point evaluation makes only sense almost everywhere, and we can only expect this limit to exist almost everywhere.

The general approach to such almost everywhere convergence problems is to consider the corresponding maximal function

$$T_\epsilon^{\max} f(x) := \sup_k \left| \sum_{|I|<2^k} \epsilon_I \langle f, h_I\rangle h_I(x)\right|$$

This is a measurable function which takes values in $\mathbf{R}^+ \cup \{\infty\}$, it is also called the maximal truncated singular integral.

Suppose we knew the following lemma

Lemma 9 *For all $1 < p < \infty$ there is a constant C_p such that for all $f \in L^p(\mathbf{R})$ and all sequences $(\epsilon_I)_I$ with $|\epsilon_I| \le 1$ for all I we have*

$$\|T_\epsilon^{\max} f\|_p \le C_p \|f\|_p \quad .$$

Then we could prove almost everywhere convergence to $f(x)$ of the series (**??**) as follows. We need to prove, that for each δ the series (**??**) converges to $f(x)$ outside a set of measure δ. Thus we have to prove that for each n there is a k_0 such that for all $k > k_0$ we have

$$\left|\left\{x : \left|\sum_{2^{-k}<|I|<2^k} \epsilon_I \langle f, h_I\rangle h_I(x) - f(x)\right| > 2^{-n}\right\}\right| \le 2^{-n}\delta \quad ,$$

because then the series converges outside a set of measure $\sum_n \delta 2^{-n} = C\delta$. Fix δ and n. Let $g \in \mathcal{H}(\mathbf{R})$ such that $\|f - g\|_p$ is smaller than a certain number depending on p, δ, and n, and let k_0 be large enough so that

$$\sum_{2^{-k_0}<|I|<2^{k_0}} \epsilon_I \langle g, h_I\rangle h_I = g \quad .$$

Then the above set can be estimated by

$$\begin{aligned}
&\left|\left\{x : \left|\sum_{2^{-k}<|I|<2^k} \epsilon_I \langle f-g, h_I\rangle h_I(x)\right| > 2^{-n-2}\right\}\right| + \\
&\left|\left\{x : \left|\sum_{2^{-k}<|I|<2^k} \epsilon_I \langle g, h_I\rangle h_I(x) - g(x)\right| > 2^{-n-2}\right\}\right| + \\
&\left|\left\{x : |f(x) - g(x)| > 2^{-n-2}\right\}\right| \\
&\le C_p 2^{p(n+2)} \|T_\epsilon^{\max}(f-g)\|_p^p + 0 + C_p 2^{p(n+2)} \|f-g\|_p^p \\
&\le C(p,n,\delta)\|f-g\|_p^p \le \delta 2^{-n}
\end{aligned}$$

where the last inequality follows by appropriate choice of g. Thus we have almost everywhere convergence of the truncated singular integrals for all $f \in L^p$.

It remains to prove the Lemma **??**.

Proof Let $1 < p < \infty$ and $f \in L^p$, then

$$\tilde{f} = \sum_I \epsilon_I \langle f, h_I\rangle h_I$$

is in L^p with $\|\tilde{f}\|_p \le C\|f\|_p$, and we have

$$\epsilon_I \langle f, h_I\rangle = \left\langle \tilde{f}, h_I\right\rangle \quad .$$

By replacing f by $\tilde{f}$ it therfore suffices to prove the lemma with $\epsilon_I = 1$ for all I.

If J is an interval, then integrating over J is a continuous operation in L^p and we have

$$|J|^{-1} \int_J f(y)\,dy$$

$$= |J|^{-1} \int_J \sum_{|I| \geq |J|} \langle f, h_I \rangle \, h_I(y)\,dy$$

$$= \sum_{|I| \geq |J|} \langle f, h_I \rangle \, h_I(x)$$

for each $x \in J$. Thus

$$T_1^{\max} f(x) = \sup_{J : x \in J} \frac{1}{|J|} \left| \int_J f(y)\,dy \right|$$

This is bounded by the (dyadic) Hardy Littlewood maximal operator

$$Mf(x) = \sup_{J : x \in J} \frac{1}{|J|} \int_J |f|(y)\,dy$$

By Marcinkiewicz interpolation it suffices to prove a weak type p, p estimate for the Hardy Littlewood maximal operator for all $1 \leq p < \infty$. (observe that trivially M is bounded in L^∞.)

Let E_λ be the set where $Mf(x) > \lambda$. By definition of f, for each x in this set there is a dyadic interval I containing x such that

$$\lambda < |I|^{-1} \int_I |f|(y)\,dy \leq |I|^{-\frac{1}{p}} \|f 1_I\|_p$$

Let $\mathcal{I}'$ be the collection of maximal dyadic intervals satisfying

$$\|f 1_I\|_p^p \geq \lambda^{-p} |I| \quad .$$

Then we have

$$|E_\lambda| \leq \sum_{I \in \mathcal{I}'} |I| \leq \lambda^{-p} \sum_{I \in \mathcal{I}'} \|f 1_I\|_p^p \leq \lambda^{-p} \|f\|_p^p$$

This proves the weak type p, p estimate for the sublinear operator M and thus finishes the proof of Lemma **??**.

■

9 Paraproducts and a discrete $T1$ theorem

So far we have studied discrete singular integral operators of the type

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

They are bounded in L^p if and only if ϵ_I are bounded. The if part follows from square function estimates, whereas the only if part is easily seen from

$$T_\epsilon h_I = \epsilon_I h_I \quad .$$

To complete the picture, we need to study another type of discrete singular integral operators. For a sequence $(\epsilon_I)_I$ define for $f \in \mathcal{H}(\mathbf{R})$

$$P_\epsilon f = \sum_I \langle f, \chi_I \rangle \, h_I \quad , \tag{17}$$

where $\chi_I = |I|^{-\frac{1}{2}} 1_I$, and the infinite sequence is easily seen to converge pointwise absolutely if $f \in \mathcal{H}(\mathbf{R})$. The operator P_ϵ is called a discrete paraproduct and obviously has similar localization properties as T_ϵ. Again we would like to know for which sequences ϵ_I the paraproduct (**??**) defines a bounded operator from L^p to L^p. As a first observation, since the χ_I do not form an orthonormal set in L^2, we do not expect P_ϵ to be bounded in L^2 for all bounded sequences ϵ_I.

The answer to this question is given by

Lemma 10 *If $|I|^{\frac{1}{2}}\epsilon_I$ is a* BMO *sequence, then the operator P_ϵ satisfies*

$$\|P_\epsilon f\|_p \leq C_p \|(|I|^{\frac{1}{2}}\epsilon_I)_I\|_{\mathrm{BMO}} \|f\|_p$$

for $1 < p < \infty$. Vice versa, if P_ϵ is bounded in L^p for any $1 < p < \infty$, then $|I|^{\frac{1}{2}}\epsilon_I$ is a BMO sequence.

Proof Assume $|I|^{\frac{1}{2}}\epsilon_I$ is a BMO sequence. Observe that

$$|I|^{-\frac{1}{2}} \, |\langle f, \chi_I \rangle| \, 1_I \leq Mf \quad .$$

where M denotes the Hardy Littlewood maximal function. Thus we have, denoting by Sf the square function of f,

$$\begin{aligned}
\|P_\epsilon f\|_p &= \sup_{\|g\|_{p'}=1} \langle P_\epsilon f, g \rangle \\
&= \sum_I \epsilon_I \, \langle f, \chi_I \rangle \, \langle h_I, g \rangle \\
&\leq \|(\epsilon_I)_I\|_{\mathrm{BMO}} \left\| \left(\sum_I |\langle f, \chi_I \rangle \, \langle h_I, g \rangle|^2 \, |I|^{-1} 1_I \right)^{\frac{1}{2}} \right\|_1 \\
&\leq \|(\epsilon_I)_I\|_{\mathrm{BMO}} \left\| \left(\sum_I |\langle h_I, g \rangle|^2 \, |I|^{-1} 1_I \right)^{\frac{1}{2}} Mf \right\|_1 \\
&\leq \|(\epsilon_I)_I\|_{\mathrm{BMO}} \, \|Mf\|_p \|Sg\|_{p'} \leq C_p \, \|(\epsilon_I)_I\|_{\mathrm{BMO}} \, \|f\|_p \quad .
\end{aligned}$$

Here we have used boundedness of the Hardy Littlewood maximal operator in L^p and boundedness of the square function in $L^{p'}$. The same calculation shows that the expression (**??**) converges in L^p regardless of the order of summation.

Vice versa, assume P_ϵ is bounded in L^p. Let J be any dyadic interval and let $f = 1_J$. Then we have

$$\left\| \sum_{I \subset J} |I|^{\frac{1}{2}} \epsilon_I h_I \right\|_p = \left\| \sum_{I \subset J} \epsilon_I \, \langle f, \chi_I \rangle \, h_I \right\|_p$$

$$\leq C_P \left\| \sum_I \epsilon_I \langle f, \chi_I \rangle h_I \right\|_{L^p} \leq C_p \|f\|_p \leq C_p |J|^{\frac{1}{p}}$$

Here the passage from the first to the second line is easily seen from passing to the norm of the square function, which is equivalent to the norm of a function. This calculation and Lemma **??** show that $(\epsilon_I)_I$ is a BMO sequence. ■

If we formally apply P_ϵ to the constant function 1, then we have

$$P_\epsilon(1) = \sum_I |I|^{\frac{1}{2}} \epsilon_I h_I \quad .$$

Thus the condition in the previous Lemma can be described as $P_\epsilon(1) \in \mathrm{BMO}$. A variant of this condition gives the $T(1)$-theorem below its name.

By duality, we have that the adjoint operator P_ϵ^* of P_ϵ is bounded in L^p for $1 < p < \infty$ if and only if ϵ_I is a BMO sequence. Since we have

$$\langle P_\epsilon f, g \rangle = \sum_I \epsilon_I \langle f, \chi_I \rangle \langle h_I, g \rangle \quad ,$$

we obtain for the adjoint operator P_ϵ^* of P_ϵ formally the expression

$$P_\epsilon^* f = \sum_I \epsilon_I \langle f, h_I \rangle \chi_I \quad , \tag{18}$$

which is a finite sum if $f \in \mathcal{H}(\mathbf{R})$ and which converges weakly in L^p if $(\epsilon_I)_I$ is a BMO sequence and $f \in L^p$. To see that (**??**) actually converges strongly in L^p regardless of the order of summation, observe that by Lebesgue's dominated convergence theorem (**??**) converges strongly in some order of summation if

$$P_\epsilon^* f = \sum_I |\epsilon_I| \, |\langle f, h_I \rangle| \, \chi_I \quad ,$$

converges in L^p in that order of summation. But by Lebesgue's monotone convergence theorem the latter converges strongly in L^p in any order of summation.

More generally, we have the following discrete $T(1)$ theorem

Theorem 1 *Let $(\epsilon_I^{(j)})_{I\in\mathcal{I}}$ be bounded sequences of real numbers for $j = 1, 2, 3$. Then the operator T, originally defined on $\mathcal{H}(\mathbf{R})$ by*

$$Tf = \sum_I \epsilon_I^{(1)} \langle f, h_I \rangle h_I + \sum_I \epsilon_I^{(2)} \langle f, \chi_I \rangle h_I + \sum_I \epsilon_I^{(3)} \langle f, h_I \rangle \chi_I$$

is bounded in L^p for $1 < p < \infty$ if $T(1) \in \mathrm{BMO}$ and $T^*(1) \in \mathrm{BMO}$.

Here ofcourse the conditions $T(1)$ and $T^*(1)$ in BMO are to be interpreted in the sense that $(\epsilon_I^{(2)})_{I\in\mathcal{I}}$ and $(\epsilon_I^{(3)})_{I\in\mathcal{I}}$ are BMO sequences.

10 The continuous $T(1)$ theorem

10.1 The Theorem

Each topic in the discrete theory of Part I translates into a continuos counterpart more closely related to Fourier analysis. Rather than discussing the whole theory in the continuous case again, we shall focus on the continuous $T(1)$ theorem. This theorem was originally proved in 1984 and is one of the highlights of the theory of singular integral operators. We formulate the theorem in $\mathbf{R}^n$, but since we would like to focus on the ideas, we shall for simplicity prove the case $n = 1$ only.

A Calderon Zygmund kernel (with parameters $A > 0$ and $\gamma > 0$) K is a measurable function on $\mathbf{R}^n \times \mathbf{R}^n$ which satisfies

$$|K(x,y)| \leq A \frac{1}{|x-y|^n} \tag{19}$$

for all $x \neq y$,

$$|K(x,y) - K(x',y)| \leq A \frac{|x-x'|^\gamma}{|x-y|^{n+\gamma}}$$

for all x, x', y such that $|x - x'| < |x - y|/2$, and

$$|K(x,y) - K(x,y')| \leq A \frac{|y-y'|^\gamma}{|x-y|^{n+\gamma}}$$

for all x, y', y such that $|y - y'| < |x - y|/2$.

Observe that these conditions imply that K is continuos away from the diagonal $x = y$, in fact it is Hölder continuous of degree γ. The Calderon Zygmund conditions become weaker as γ approaches 0, with $\gamma = 0$ the conditions reduce to (**??**), which turns out to weak by itself to give a reasonable boundedness theory.

To compare this with the discrete $T(1)$ theorem observe that the kernel

$$\sum_I \epsilon_I^{(1)} h_I(x) h_I(y) + \epsilon_I^{(2)} \chi_I(x) h_I(y) + \epsilon_I^{(3)} h_I(x) \chi_I(y) \tag{20}$$

satisfies the conditions

$$|K(x,y)| \leq A |I_{x,y}|^{-1}$$

for all positive $x \neq y$, where $I(x,y)$ is the smallest dyadic interval containing x and y,

$$|K(x,y) - K(x',y)| = 0$$

for all pairwise distinct positive x, x', y such that the smallest dyadic interval containing x and x' has smaller length than the smallest dyadic interval containing x and y, and

$$|K(x,y) - K(x,y')| = 0$$

for all pairwise distinct positive x, y', y such that the smallest dyadic interval containing y and y' has smaller length than the smallest dyadic interval containing x and

y; and similarly on the neagtive axis. These properties of (**??**) are easily seen, using that $\epsilon_I^{(j)}$ are bounded.

Given a Calderon Zygmund kernel, we would like to define an operator T mapping $\mathcal{S}(\mathbf{R}^n)$ to $\mathcal{S}'(\mathbf{R}^n)$ by

$$(Tf,g) = \int_{\mathbf{R}^n}\int_{\mathbf{R}^n} K(x,y)f(y)g(x)\,dy\,dx \quad . \tag{21}$$

However, it is not clear how to make sense of this expression, because K is not an integrable function. We shall therefore not attempt to define T by (**??**). However, we observe that K is locally integrable away from the diagonal, therefore we make the following definition:

Definition 1 *An operator $T : \mathcal{S}(\mathbf{R}^n) \to \mathcal{S}'(\mathbf{R}^n)$ is said to have a Calderon Zygmund kernel K if (**??**) holds for all compactly supported Schwartz functions f and g which have disjoint supports.*

It is easy to see that if T has Calderon Zygmund kernel K, then the adjoint operator T^* defined by

$$\langle T^*f, g\rangle = \langle f, Tg\rangle$$

has the Calderon Zygmund kernel

$$K^*(x,y) = \overline{K(y,x)} \quad ,$$

the Calderon Zygmund conditions for K^* being easily verified.

Observe that an operator can have at most one Calderon Zygmund kernel, because we can obtain the value $K(x,y)$ as a limit of (Tf,g) where the supports of f and g shrink to x and y respectively. However, several operators can have the same Calderon Zygmund kernel, e.g. for each $x \in \mathbf{R}^n$ and each $c \in \mathbf{R}$ the operator defined by $(Tf,g) = cf(x)g(x)$ has the kernel $K = 0$.

The last example also shows, that an operator T with a Calderon Zygmund kernel K is not necessarily bounded in $L^2(\mathbf{R}^n)$. The $T(1)$ theorem, which we are about to state, characterizes the operators with Calderon Zygmund kernels which are bounded in $L^2(\mathbf{R}^n)$.

Recall the following notation:

$$\mathrm{D}^p_\lambda f(x) = \lambda^{-\frac{1}{p}} f(\lambda^{-1}x)$$

$$T_y f(x) = f(x-y)$$

$$D^\alpha f = (2\pi i)^{-|\alpha|}\partial_1^{\alpha_1}\dots\partial_n^{\alpha_n} f$$

$$X^\beta f(x) = x_1^{\beta_1}\dots x_n^{\beta_n} f(x)$$

$$\|f\|_{[N]} = \sup_{|\alpha|,|\beta|\le N} \left\| D^\alpha X^\beta f\right\|_2$$

Theorem 2 *Let T be a continuos operator from $\mathcal{S}(\mathbf{R}^n)$ to $\mathcal{S}'(\mathbf{R}^n)$ and assume T has a Calderon Zygmund kernel K. Then T extends to a bounded operator from $L^2(\mathbf{R}^n)$ to itself if and only if there exist an $N > 0$ and an $A' > 0$ such that*

$$\|T(T_x \mathrm{D}^2_\lambda f)\|_2 \leq A' \|f\|_{[N]} \tag{22}$$

$$\|T^*(T_x \mathrm{D}^2_\lambda f)\|_2 \leq A' \|f\|_{[N]} \tag{23}$$

for all $x \in \mathbf{R}^n$, $\lambda \in \mathbf{R}^+$, and $f \in \mathcal{S}(\mathbf{R}^n)$. More precisely, we have

$$\|Tf\|_2 \leq C_N \max(A, A') \|f\|$$

for all $f \in \mathcal{S}(\mathbf{R})$, where C_N depends only on N, and A is the constant in the conditions on the Calderon-Zygmund kernel.

The necessity of the condition is immediately clear, because if T is bounded then it follows that

$$\|T(T_x \mathrm{D}^2_\lambda f)\|_2 \leq C\|f\|_2 \leq C\|f\|_{[N]}$$

and likewise for T^*. The proof of sufficiency of this condition will be done in the next few sections.

10.2 Identifying the part corresponding to $\sum_I \epsilon_I h_I \otimes h_I$

We begin to prove boundedness of T as in Theorem **??**. For simplicity we discuss the one dimensional case only. Moreover, for simplicity of notation we shall not discuss the dependence of the constants on A, A', and N.

First we observe that if f is a compactly supported Schwartz function with mean zero, then Tf is actually in L^1. By hypothesis of the theorem it is clear that $Tf \in L^2$. Let B be a ball with radius larger than 1 around the origin of $\mathbf{R}$ such that the support of f is contained in B. Then it suffices to show that f restricted to the complement of $3B$ is a finite measure, i.e.,

$$|(Tf, g)| \leq C\|g\|_\infty$$

for some constant C and all $g \in \mathcal{S}(\mathbf{R})$ supported outside $2B$. Since f and g have disjoint support, we can write

$$|(Tf, g)| = \left| \int \int K(x, y) f(y) g(x)\, dx dy \right|$$

Using the mean 0 property of f this is the same as

$$\left| \int \int (K(x, y) - K(x, 0)) f(y) g(x)\, dx dy \right|$$

Since $|y - 0| < \frac{1}{2}|x - 0|$ on the support of $f(y)g(x)$, this is bounded by

$$\int \int \frac{|y|^\gamma}{|x|^{1+\gamma}} |f(y)||g(x)|\, dx dy$$

$$\leq C \int \frac{1}{|x|^{1+\gamma}} |g(x)|\, dx$$

$$\leq C\|g\|_\infty \int_{|B|}^{\infty} |x|^{-1-\gamma}\, dx \leq C\|g\|_\infty \quad .$$

This proves that $Tf \in L^1$ if f is compactly supported and has mean zero. By symmetry the same holds for T^*. Hence we can take the mean of Tf and T^*f for such functions f.

Our plan is to first prove boundedness of T under the additional assumption that all these means are zero, i.e.,

$$\int Tf(x)\, dx = 0$$

and

$$\int T^* f(x)\, dx = 0$$

for all compactly supported $f \in \mathcal{S}(\mathbf{R})$ with $\int f(x)\, dx = 0$. To motivate this assumption we discuss the meaning of this assumption in the discrete model. Assume we have T given by the kernel

$$\sum_I \sum_I \epsilon_I^{(1)} h_I(x) h_I(y) + \epsilon_I^{(2)} \chi_I(x) h_I(y) + \epsilon_I^{(3)} h_I(x) \chi_I(y) \tag{24}$$

In the discrete setting the natural ("smooth") function with compact support is a Haar function h_I. Observe that

$$\int Th_I(x)\, dx = \epsilon_I^{(3)} |I|^{\frac{1}{2}}$$

$$\int T^* h_I(x)\, dx = \epsilon_I^{(2)} |I|^{\frac{1}{2}}$$

Thus the additional assumption in the discrete case is the same as saying that all $\epsilon_I^{(2)}$ and $\epsilon_I^{(3)}$ are 0. In this sense we are dealing with the $\sum_I \epsilon_I^{(1)} h_I \otimes h_I$-part only.

The next step is to decompose the operator T similarly to the decomposition in the discrete case. Pick a Schwartz function ϕ supported in the unit ball $B_1(0)$ with $\widehat{\phi}(0) = 1$. Define ϕ_n and ψ_n by

$$\widehat{\phi_j} = \mathrm{D}^{\infty}_{2^n} \widehat{\phi} \quad ,$$

$$\psi_j = \phi_j - \phi_{j-1} \quad .$$

Obviously $\widehat{\phi}_n(0) = 1$ and $\widehat{\psi}_j(0) = 0$ for all j.

It is easy to see that for each $f \in \mathcal{S}(\mathbf{R})$

$$\lim_{j\to\infty} f * \phi_j = f$$

$$\lim_{j\to-\infty} f * \phi_j = 0$$

$$\lim_{j\to\infty} T(f * \phi_j) * \phi_j = Tf$$

$$\lim_{j\to-\infty} T(f*\phi_j)*\phi_j = 0$$

Thus we can write T as a telescoping series

$$T = \sum_{j=-\infty}^{\infty} T_j$$

where T_j is defined by

$$T_j f = T(f*\phi_j)*\phi_j - T(f*\phi_{j-1})*\phi_{j-1}$$

Convolution with ϕ_j is a smoothing operation at scale 2^{-j}. In the discrete model this corresponds to replacing a function f on each dyadic interval of length 2^{-j} by its average:

$$f \to \sum_{|I|>2^{-j}} \langle f, h_I\rangle\, h_I \quad .$$

Thus the operatior $f \to T(f*\phi_j)*\phi_j$ in the case of (**??**) with vanishing paraproducts is given by

$$f \to \sum_{|I|>2^{-j}} \epsilon^{(1)}_{(I)}\, \langle f, h_I\rangle\, h_I \quad .$$

The analogue of the operator T_j is then given by

$$f \to \sum_{|I|=2^{1-j}} \epsilon^{(1)}_{(I)}\, \langle f, h_I\rangle\, h_I \quad ,$$

which is taking a single scale of T.

We study the operators T_j. First observe that we can write for $T_j f$

$$T(f*\phi_j)*\phi_j - T(f*\phi_j)*\phi_{j-1} + T(f*\phi_j)*\phi_{j-1} - T(f*\phi_{j-1})*\phi_{j-1}$$

$$= T(f*\phi_j)*\psi_j + T(f*\psi_j)*\phi_{j-1}$$

Since both pieces are similar we will only consider

$$T_j^{(1)} = T(f*\phi_j)*\psi_j$$

Our strategy now is as follows. We will prove in the next section that the operators T_j are almost orthogonal, i.e.,

$$\|T_i T_j^*\| \le C 2^{-\gamma'|i-j|}$$

$$\|T_i^* T_j\| \le C 2^{-\gamma'|i-j|}$$

for some $\gamma' < \gamma$. (In the discrete model they actually are pairwise orthogonal). Then we will prove a lemma by Cotlar and Stein which says that almost orthogonality of the T_j implies that $\sum_j T_j$ is a bounded operator in L^2.

10.3 Estimates on $T_iT_j^*$ and $T_i^*T_j$

First we prove an easy general lemma, which will be applied to the operators $T_i^*T_j$ and $T_iT_j^*$.

Lemma 11 *Let $A > 0$ and let K be a bounded continuous (weaker assumptions would do) function on $\mathbf{R}^2$ such that for all $x, y \in \mathbf{R}$ we have*

$$\int |K(t,y)|\, dt \leq A \quad ,$$

$$\int |K(x,t)|\, dt \leq A \quad .$$

Then the operator $S : \mathcal{S}(\mathbf{R}) \to \mathcal{S}'(\mathbf{R})$ defined by

$$(Sf, g) = \int\int K(x,y) f(y) g(x)\, dy dx$$

extends to a bounded operator from L^2 to L^2. More precisely it satisfies

$$|(Sf, g)| \leq 2A \|f\|_2 \|g\|_2 \quad .$$

Proof By linearity it suffices to prove this for $\|f\|_2 = \|g\|_2 = 1$. We observe that $|f(y)g(x)| \leq |f(y)|^2 + |g(x)|^2$. Hence

$$|(Sf, g)| = \left| \int\int K(x,y) f(y) g(x)\, dy dx \right|$$

$$\leq \int\int |K(x,y)| |f(y)|^2\, dx dy + \int\int |K(x,y)| |g(x)|^2\, dx dy$$

$$\leq A \int |f(y)|^2\, dy + A \int |g(x)|^2\, dx \leq 2A$$

This completes the proof. ■

We return to the discussion of the operators T_j.

Recall the formula

$$\Lambda * f(x) = (\Lambda, \check{f})$$

where $\Lambda \in \mathcal{S}'(\mathbf{R})$, $f \in \mathcal{S}(\mathbf{R})$, and $\check{f}(x) = f(-x)$. Also, for $f, g \in \mathcal{S}(\mathbf{R})$ we have the formula

$$f * g = \int f(t) T_t g\, dt$$

where the right hand side is a vector valued integral, which can be interpreted as limit of the Riemann sums

$$\frac{1}{K} \sum_{k=-K^2}^{K^2} f(k/K) T_{k/K} g$$

as $K \to \infty$, with convergence in $\mathcal{S}(\mathbf{R})$. Since T is continuous form $\mathcal{S}(\mathbf{R})$ to $\mathcal{S}(\mathbf{R}')$, we obtain

$$\begin{aligned} T(f * \phi_j) * \psi_j(x) &= \left(T(f * \phi_j), T_x \check{\psi} \right) \\ &= \left(T \left(\int f(t) T_t \phi_j \, dt \right), T_x \check{\psi} \right) \\ &= \int f(t) \left(T(T_t \phi_j, T_x \check{\psi}_j \right) \, dt \end{aligned}$$

Thus the operator T_j is given by a smooth kernel

$$K_j(x,t) = \left(T(T_t \phi_j, T_x \check{\psi}_j \right) \, dt$$

By hypothesis (**??**) of the theorem we have

$$|K_j(x,t)| \leq 2^j \|\phi\|_{[N]} \|\psi\|_2 \leq C 2^j \quad .$$

If $|x - t| > 2^{-j+2}$, then $T_t \phi_j$ and $T_x \check{\psi}_j$ have disjoint support and we have

$$|K_j(x,t)| \leq \left| \int \int K(x', t') \phi_j(t' - t) \check{\psi}_j(x' - x) \, dt' dx' \right|$$

$$\begin{aligned} &= \left| \int \int (K(x', t') - K(x, t')) \phi_j(t' - t) \check{\psi}_j(x' - x) \, dt' dx' \right| \\ &\leq C \int \int \frac{|x - x'|^\gamma}{|x' - t'|^{1+\gamma}} |\phi_j(t' - t)| |\check{\psi}_j(x' - x)| \, dt' dx' \\ &\leq C \|\phi\|_1 |x - t|^{-(1+\gamma)} \int \int |x - x'|^\gamma |\check{\psi}_j(x' - x)| \, dt' dx' \\ &\leq C \|\phi\|_1 \|\psi\|_1 |x - t|^{-(1+\gamma)} (2^{-j})^\gamma \end{aligned}$$

Putting both estimates together we obtain

$$|K(x,t)| \leq C 2^j \min \left(1, \left(\frac{|x - t|}{2^{-j}} \right)^{-1-\gamma} \right) \quad .$$

Similarly, we obtain

$$\begin{aligned} \left| D_1^\alpha D_2^\beta K(x,t) \right| &= \left| \left(T(D_1^\alpha T_t \phi_j), D_2^\beta T_t \check{\psi}_j \right) \right| \\ &\leq C 2^{j(1+|\alpha|+|\beta|} \min \left(1, \left(\frac{|x - t|}{2^{-j}} \right)^{-1-\gamma} \right) \quad . \end{aligned}$$

The additional mean zero assumptions on T and T^* yield

$$\int K_j(x,y) \, dx = 0 \tag{25}$$

$$\int K_j(x,y) \, dy = 0 \tag{26}$$

Consider first (**??**). It is equivalent to

$$\int\int K_j(x,y)f(y)\,dxdy = 0$$

which is the same as

$$\int T_j f(x)\,dx = 0$$

However,

$$T_j f = T(f * \phi_j) * \psi_j$$

If $T_j f(x)$ was integrable, this would already imply (**??**) because ψ_j has mean zero. Let B be a large ball centered at the origin such that $f * \phi_j$ is supported in B. We consider $T(f * \phi)$ separately on $2B$ and the complement of $2B$. On $2B$ it is in L^1 because it is locally in L^2. Therfore we can apply the above argument. For $x \in 2B^c$ we have that $T(f * \phi_j)(x)$ is given by

$$\int K(x,y)T(f * \phi_j)(y)\,dy 1_{2B^c}(x) \quad .$$

Convolving this with ψ_j and taking the mean gives

$$\int\int\int K(x,y)T(f * \phi_j)(y)1_{2B^c}(x)\phi(t-x)\,dydtdx \quad ,$$

$$= \int\int [K(x,y) - K(x,0)]T(f * \phi_j)(y)1_{2B^c}(x)\phi(t-x)\,dydtdx \quad .$$

Now the integrand is absolutely integrable, and we can use Fubini to integrate first in x to see that this is equal to 0. (This is the point of throwing in $K(x,0)$: before that we couldn't apply Fubini.) This proves (**??**).

Equation (**??**) is equivalent to

$$\int T_j^* f(x)\,dx = 0$$

We have

$$T_j^* f = T^*(f * \overline{\check{\psi}_j}) * \overline{\check{\phi}_j} \quad ,$$

which has mean zero by the special assumption on T^* because $f * \overline{\check{\psi}_j}$ has mean zero.

Now we have to calculate $\|T_i^* T_j\|$ and $\|T_i T_j^*\|$. By symmetry it suffices to calculate the first norm only and assume $i \geq j$.

The kernel of $T_i^* T_j$ is given by

$$K_{i,j}(x,y) = \int \overline{K_i(t,x)} K_j(t,y)\,dt$$

By the cancellation condition (**??**) this is equal to

$$K_{i,j}(x,y) = \int \overline{K_i(t,x)}[K_j(t,y) - K_j(x,y)]\,dt$$

First we consider this integral on the region E of all t such that

$$|t-x| \le 2^{-i}2^{\gamma'(i-j)} \quad .$$

We have

$$\begin{aligned}
&\left|\int_E \overline{K_i(t,x)}[K_j(t,y)-K_j(x,y)]\,dt\right| \\
&\le \left\|\overline{K_i(x,.)}\right\|_1 |E| \sup_{t\in E} |\partial_1 K_j(t,y)| \\
&\le C2^{-i}2^{\gamma'(i-j)}2^{2j}\min\left(1,\left(\frac{|x-y|}{2^{-j}}\right)^{1+\gamma}\right)
\end{aligned}$$

In the last line we could replace the supremum over t by evaluation at x (an loose at most a factor of 2) because $|x-t| \le 2^{-j}$. Now we consider the integral on the complement of E intersected with the set F of all t such that $2|x-t| < |x-y|$. There we have

$$\begin{aligned}
&\left|\int_{E^c\cap F} \overline{K_i(t,x)}[K_j(t,y)-K_j(x,y)]\,dt\right| \\
&\le \left\|\overline{K_i(.,y)}\right\|_{L^1(E^c)} \sup_{t\in F^c} |K_j(t,x)| \\
&\le C2^{-\gamma\gamma'(i-j)}\min\left(1,\left(\frac{|x-y|}{2^{-j}}\right)^{1+\gamma}\right)
\end{aligned}$$

In the last line we have used that $|x-y|$ and $|t-y|$ are of comparable size within a factor of 2. Finally, we consider the integral on $E^c\cap F^c$. There we have

$$\begin{aligned}
&\left|\int_{E^c\cap F^c} \overline{K_i(t,x)}[K_j(t,y)-K_j(x,y)]\,dt\right| \\
&\le \left\|\overline{K_j(.,y)}\right\|_1 \sup_{t\in E^c\cap F^c} \|K_i(.,x)\|_1 \\
&\le C2^{i} \sup_{t\in E^c\cap F^c} \min\left(1,\left(\frac{|x-t|}{2^{-i}}\right)^{1+\gamma}\right) \\
&\le C2^{j}2^{-\gamma(i-j)}\min\left(1,\left(\frac{|x-y|}{2^{-j}}\right)^{1+\gamma}\right)
\end{aligned}$$

In the last line we have used that for each $t\in E^c\cap F^c$ the minimum in the previous line is attained by the element

$$\left(\frac{|x-t|}{2^{-i}}\right)^{1+\gamma} \le 2^{-(1+\gamma)(i-j)}\left(\frac{|x-y|}{2^{-j}}\right)^{1+\gamma}$$

These three estimates together give

$$|K_{i,j}(x,y)| \le C2^{j}2^{-\gamma''(i-j)}\min\left(1,\left(\frac{|x-y|}{2^{-j}}\right)^{1+\gamma}\right)$$

for some $\gamma''>0$.

By Lemma (**??**) this easily implies $\|T_i^*T_j\| \le 2^{-\gamma''|i-j|}$, which we had aimed to prove.

10.4 The Cotlar-Stein Lemma

The Cotlar-Stein lemma is a lemma on bounded operators on a Hilbert space. We will write f for an element of this Hilbert space and $\|f\|_2$ for its norm. The norm of a bounded operator T form the Hilbert space to itself will be denoted by $\|T\|$.

Lemma 12 *Let $(T_j)_{j=1}^J$ be a finite set of bounded linear operators on a Hilbert space. Assume that for each j we have*

$$\sum_i \|T_j T_i^*\|^{\frac{1}{2}} \leq C$$

$$\sum_i \left\|T_j^* T_i\right\|^{\frac{1}{2}} \leq C$$

Then

$$\left\|\sum_j T_j\right\| \leq C \quad .$$

Proof

First we recall that the operator norm of an operator T on a Hilbert space satisfies

$$\|T\| = \|T^*\| \tag{27}$$

and

$$\|T\|^2 = \|T^*T\| \quad . \tag{28}$$

The first statement follows as

$$\|T\|^2 = \sup_{\|f\|_2=1} \|Tf\|_2 = \sup_{\|f\|_2, \|g\|_2=1} |\langle Tf, g\rangle|$$

$$= \sup_{\|f\|_2, \|g\|_2=1} |\langle f, T^*g\rangle| = \ldots = \|T^*\|$$

For the second statement we observe

$$\|T\|^2 = \sup_{\|f\|_2=1} \|Tf\|_2^2 = \sup_f \langle Tf, Tf\rangle = \sup_f \langle T^*Tf, f\rangle \leq \sup_f \|T^*Tf\|_2 = \|T^*T\|$$

and on the other hand

$$\|T^*T\| \leq \|T\|\|T^*\| \leq \|T\|^2 \quad .$$

We make a few more observations to motivate the proof of Cotlar-Stein. We trivially have

$$\|\sum_j T_j\| \leq \sum_j \|T_j\| \quad . \tag{29}$$

if we assume $T_i^*T_j = 0$ for all $i \neq j$ we have

$$\|\sum_j T_j\|^2 = \|\sum_j \sum_i T_j^* T_i\| = \|\sum_j T_j^* T_j\| = \sum_J \|T_j\|^2 \quad .$$

thus we can replace the sum in (**??**) by a square sum, which is clearly better. We observe that the same would be true if we had assumed $T_j T_i^* = 0$ for all $j \neq i$, because then we could run this argument on $\sum_j T_j^*$ and the invoke (**??**).

If we assume both $T_j^* T_i = 0$ and $T_j T_i^* = 0$ for all $i \neq j$, then we can continue to estimate higher powers: For self adjoint operators T equation (**??**) gives $\|H^2\| = \|H\|^2$, which can be iterated to give

$$\|H^n\| = \|H\|^n$$

for n some power of 2. Since T^*T is self adjoint for any operator T, we obtain

$$\left\|\sum_j T_j^* T_j\right\|^n = \left\|\sum_{j_1, j_2, \ldots, j_n} T_{j_1}^* T_{j_1} T_{j_2}^* \ldots T_{j_n}^* T_{j_n}\right\|$$

$$= \| \sum_j (T_j^* T_j)^n \| \leq \sum_j \|T_j^* T_j\|^n$$

Thus we can replace the sum in (**??**) by an $2n$- sum, which in the limit becomes a supremum:

$$\| \sum_j T_j \| \leq \sup_j \|T_j\| \quad ,$$

which is clearly the best one can do.

We return to the proof of Cotlar Stein. We have for n any power of 2:

$$\| \sum_j T_j \|^{2n} = \| \sum_{j_1, j_2, \ldots j_{2n}} T_{j_1}^* T_{j_2} T_{j_3}^* T_{j_4} \ldots T_{j_{2n-1}}^* T_{j_{2n}} \|$$

This is bounded by either one of the two following quantities

$$\sum_{j_1, j_2, \ldots j_{2n}} \|T_{j_1}^* T_{j_2}\| \|T_{j_3}^* T_{j_4}\| \ldots \|T_{j_{2n-1}}^* T_{j_{2n}}\|$$

$$\sum_{j_1, j_2, \ldots j_{2n}} \|T_{j_1}^*\| \|T_{j_2} T_{j_3}^*\| \|T_{j_4} T_{j_5}\| \ldots \|T_{j_{2n}}\|$$

For each tuple $j_1, \ldots j_{2n}$ we can take the geometric mean of the two terms, which gives

$$\| \sum_j T_j \|^{2n} \leq \sum_{j_1, j_2, \ldots j_{2n}} \left(\|T_{j_1}^*\| \|T_{j_1}^* T_{j_2}\| \|T_{j_2} T_{j_3}^*\| \ldots \|T_{j_{2n-1}}^* T_{j_{2n}}\| \|T_{j_{2n}}\| \right)^{\frac{1}{2}}$$

Now we unravel the sum and use the assumptions of the lemma:

$$\leq C^{\frac{1}{2}} \sum_{j_2, \ldots j_{2n}} \sum_{j_1} \left(\|T_{j_1}^* T_{j_2}\| \|T_{j_2} T_{j_3}^*\| \ldots \|T_{j_{2n-1}}^* T_{j_{2n}}\| \|T_{j_{2n}}\| \right)^{\frac{1}{2}}$$

$$\leq C^{\frac{3}{2}} \sum_{j_3, \ldots j_{2n}} \sum_{j_2} \left(\|T_{j_2} T_{j_3}^*\| \ldots \|T_{j_{2n-1}}^* T_{j_{2n}}\| \|T_{j_{2n}}\| \right)^{\frac{1}{2}}$$

$$\leq \ldots \leq C^{\frac{4n-1}{2}} \sum_{j_{2n}} \|T_{j_{2n}}\|^{\frac{1}{2}}$$
$$\leq JC^{2n}$$

Taking roots gives

$$\|\sum_j T_j\| \leq C(J)^{\frac{1}{2n}} \quad ,$$

which proves the lemma as n tends to ∞. ■

10.5 The paraproducts

We have proved the $T(1)$ theorem under the additional assumptions

$$\int Tf(x)\,dx = 0 \tag{30}$$

$$\int T^*f(x)\,dx = 0 \tag{31}$$

for all compactly supported f with $\int f(x)\,dx = 0$. Now we remove these additional assumptions. Let T be as in the $T(1)$ theorem. It suffices to construct two operators P_1 and P_2 which are bounded in L^2, have Calderon Zygmund kernels, and satisfy

$$\int Pf(x)\,dx = \int Tf(x)\,dx \tag{32}$$

$$\int P^*f(x)\,dx = 0 \tag{33}$$

$$\int Qf(x)\,dx = 0$$

$$\int Q^*f(x)\,dx = \int T^*f(x)\,dx$$

Namely, then $T - P - Q$ has a Calderon Zygmund kernel, satisfies the assumptions of the $T(1)$ theorem, and the additional cancellation assumptions (**??**) and (**??**). Thus $T - P - Q$ is bounded in L^2 by what we have seen before. Since P and Q are bounded, so is T, which had to be shown.

There are many ways to construct such operators P and Q. Given the discrete theory, this becomes particularly easy if one assumes the existence of a compactly supported sufficiently regular wavelet. We shall do so and prove the existence of such wavelets later. (It turns out that the existence of such wavelets takes some work, which in total makes our apporach not as economic as it migt appear at first glance. However, since studying wavelets is of its own interest, we choose this approach.)

Thus we assume that there is a compactly supported, N times continuously differenctiable function (the N from the $T(1)$ theorem, which we can assume to be larger than 1) ψ with mean 0, such that, if we define for each dyadic interval $I = [2^k n, 2^k(n+1))$ the function ψ_I to be

$$\psi_I = \mathrm{D}^2_{2^k} T_n \psi \quad ,$$

we have

(34)
$$\sum_I \langle f, \psi_I \rangle \, \psi_I = f$$

for all $f \in L^2(\mathbf{R})$. Observe that the collection ψ_I was an orthonormal basis of $L^2(\mathbf{R})$ if $\|\psi\| = 1$; however, to simplify the later task of constructiong ψ we shall not make this assumption.

Let ϕ be a C^∞ function supported in $[0,1]$ with mean 1, and define ϕ_I in analogy to ψ_I.

By symmetry we shall only dicuss P. Define for $f \in \mathcal{S}(\mathbf{R})$:

$$Pf = \sum_I |I|^{-\frac{1}{2}} \langle f, \psi_I \rangle \, \epsilon_I \phi_I$$

where

$$\epsilon_I = \int T\phi_I(x)\, dx$$

Assume for now that the sequence $(\epsilon_I)_I$ is a BMO sequence, we will prove this fact later. Then we have for every $g \in \mathcal{S}(\mathbf{R})$:

$$\begin{aligned}
|(Pf, g)| &= \left| \sum_I \langle f, \psi_I \rangle \, |I|^{-\frac{1}{2}} \epsilon_I \, \langle \phi_I, g \rangle \right| \\
&\leq \|\epsilon_I\|_{\mathrm{BMO}} \left\| \left(\sum_I \left| \sum_I \langle f, \psi_I \rangle \, \langle \phi_I, g \rangle \right|^2 |I|^{-1} 1_I \right)^{\frac{1}{2}} \right\|_1 \\
&\leq C \|\epsilon_I\| \left(\sum_I \langle f, \psi_I \rangle \, |^2 \right)^{\frac{1}{2}} \|Mg\|_2 \\
&\leq C \|\epsilon_I\| \|f\|_2 \|g\|_2 \quad .
\end{aligned}$$

Hence P extends to a bounded operator on L^2. In this calculation we have used that ϕ_I is dominated by a multiple of the characteristic function of I and therfore we can estimate $\langle g, \phi_I \rangle$ by the Hardy Littlewood maximal function Mg of g. The above calculation also justifies the infinite sum in the definition of P.

We continue to assume that ϵ_I is a BMO sequence and therefore $|I|^{-\frac{1}{2}} \epsilon_I \leq C$ for all I. We have for the kernel of P away form the diagonal $x = y$:

(35)
$$K(x,y) = \sum_I |I|^{-\frac{1}{2}} \epsilon_I \overline{\psi_I(y)} \phi_I(x)$$

Let B be a large number so that the support of ψ is contained in $-B/2, B/2$. Then it is easy to see that the the support of ψ_I is contained in BI, where BI denotes the interval with the same center as I and length $B|I|$. For each I with $2B|I| < |x-y|$ the corresponding term in the sum (**??**) vanishes because either x is not in the support of

ϕ_I or y is not in the support of ψ_I. Moreover, for each k there is at most 1 intervals I of length 2^k such that $\phi_I(x)$ is nonzero. Hence we easily see

$$|K(x,y)| \le C \sum_{k:2^k>(2B)^{-1}|x-y|} 2^{-k} \le C|x-y|^{-1} \quad .$$

Moreover the same argument gives

$$|\partial_x K(x,y)| = \left|\sum_I |I|^{-\frac{1}{2}} \epsilon_I \overline{\psi_I(y)} {\phi_I}'(x)\right|$$

$$\le C|B| \sum_{k:2^k>(2B)^{-1}|x-y|} 2^{-2k} \le C|x-y|^{-2} \quad .$$

Likewise we obtain an estimate for $\partial_y K$. By the fundamental theorem of calculus this proves that K is a Calderon Zygmund kernel.

We now prove that ϵ_I is a BMO sequence.

Let φ be a compactly supported function which is constant equal to 1 on some interval around the origin. Let J be any dyadic interval and consider x_0, λ_0 so that $\lambda_0 < C|J|$ and $\mathrm{D}^\infty_{\lambda_0} T_{x_0}\varphi$ is constant on $4BJ$.

Then we have

$$\sum_{I\subset J} \left| \int T\psi_I(x)\, dx\right|^2 \le C\sum_I \left|\left\langle T\psi_I(x), \mathrm{D}^\infty_{\lambda_0} T_{x_0}\varphi\right\rangle\right|^2 + C\sum_{I\subset J}\left|\int_{4BJ^c} T\psi_I(x)\, dx\right|^2$$

By the reproducing formula (**??**) the first term on the right hand side is bounded by

$$C\left\|T^*\mathrm{D}^\infty_{\lambda_0} T_{x_0}\varphi\right\|_2^2 \le C\lambda \le C|J|$$

Here we have used the full strength of the assumption in the $T(1)$ theorem. For the second term we consider each I separately:

(36)
$$\left|\int_{4BJ^c} T\psi_I(x)\, dx\right|$$

$$\le \left|\int_{4BJ^c}\int_y \left(K(x,y) - K(x,c(I))\right)\psi_I(y)\, dy\, dx\right|$$

$$\le \left|\int_{4BJ^c}\int_y |I|^\gamma |x-c(I)|^{-1-\gamma}\psi_I(y)\, dy\, dx\right|$$

$$\le |I|^{\frac{1}{2}}\left|\int_{4BJ^c} |I|^\gamma |x-c(I)|^{-1-\gamma} dx\right|$$

$$\le |I|^{\frac{1}{2}}(|I|/|J|)^\gamma$$

For the square sum of these terms over all $I\subset J$ we obtain

$$\sum_{I\subset J}\left|\int_{2BJ^c} T\psi_I(x)\, dx\right|^2$$

$$\leq C \sum_{k>0} \sum_{I \subset J : |I| = 2^{-k}|J|} |I|(|I|/|J|)^{2\gamma}$$

$$\leq C \sum_{k>0} |J| 2^{-2\gamma k} \leq C|J| \quad .$$

Thus we have proved that

$$\sum_{I \subset J} |\epsilon_I|^2 \leq C|J|$$

for each dyadic interval J. This proves that ϵ_I is a BMO sequence.

It remains to prove that P satisfies (**??**) and (**??**). Let f be a compactly supported smooth functiona with mean zero. Since P satisfies the assumptions of the $T(1)$ theorem, the function Pf is integrable.

We observe that

$$\int Pf(x) = \int \sum_I |I|^{-\frac{1}{2}} \epsilon_I \langle f, \psi_I \rangle \, \phi(x) \, dx$$

$$= \sum_I \epsilon_I \langle f, \psi_I \rangle \quad .$$

To justify this calculation, we need to show that the right hand side converges absolutely, which follows from the easy observation that $\langle f, \psi_I \rangle$ is an H^1 sequence because ϵ_I is a BMO sequence. Let φ be as above. We claim that

$$\sum_I \epsilon_I \langle f, \psi_I \rangle = \lim_{\lambda \to \infty} \sum_I \langle f, \psi_I \rangle \langle T\psi_I, D_\lambda^\infty \varphi \rangle \quad .$$

To prove this claim let $J = [-2^k_o, 2^{k_0})$ be an interval around the origin that contains the support of f. Let c be a small constant so that φ is constant 1 on $[-c, c]$ and assume λ satisfies $c\lambda > 2^{k_0+1}$. Then we have

$$\left| \sum_I \epsilon_I \langle f, \psi_I \rangle - \sum_I \langle f, \psi_I \rangle \langle T\psi_I, D_\lambda^\infty \varphi \rangle \right|$$

$$\leq \left| \sum_I \langle f, \psi_I \rangle \int_{[-c\lambda, c\lambda]^c} T\psi_I(x) \, dx \right|$$

We split the sum into those I with $I \subset J$ and those I with $|I| \geq |J|$ and $I \cap J \neq \emptyset$. The first part can be estimated by the calculations done in (**??**) by

$$\leq C \left| \sum_{I \subset J} \langle f, \psi_I \rangle |I|^{\frac{1}{2}} (|I|/\lambda)^\gamma \right|$$

$$\leq C \sum_{I \subset J} \|f\|_\infty |I| (|I|/\lambda)^\gamma$$

$$\leq C \|f\|_\infty |J| (|J|/\lambda)^\gamma$$

The second part can be estimated by

$$\sum_{I:|I|\geq|J|,I\cap J\neq\emptyset} |\langle f,\psi_I\rangle|\,\|T\psi_I\|_1 \tag{37}$$

Observe that if F is a compactly supported function with $F'=f$ then

$$\langle f,\psi_I\rangle = -\langle F,\psi_I'\rangle \leq |I|^{-\frac{3}{2}}\|F\|_1 \quad .$$

Moreover,

$$\begin{aligned}\|T\psi_I\|_1 &\leq \|T\psi_I\|_{L^1(4BI)} + \|T\psi_I\|_{L^1(4BI)'}\\ &\leq C|I|^{\frac{1}{2}}\|T\psi_I\|_2 + C|I|^{\frac{1}{2}} \leq C|I|^{\frac{1}{2}} \quad .\end{aligned}$$

Hence (**??**) can be estimated by

$$\sum_{I:|I|\geq|J|,I\cap J\neq\emptyset} C|I|^{-1} \leq C|J|^{-1}$$

Letting $|J|$ and $\lambda/|J|$ tend to ∞ proves the claim. Thus we have

$$\int Pf(x)\,dx = \lim_{\lambda\to\infty}\sum_I \langle f,\psi_I\rangle\,\langle T\psi_I, D_\lambda^\infty\varphi\rangle \quad .$$

$$= \lim_{\lambda\to\infty}\sum_I \langle f,\psi_I\rangle\,\langle \psi_I, T^*D_\lambda^\infty\varphi\rangle \quad .$$

$$= \lim_{\lambda\to\infty}\left\langle\sum_I \langle f,\psi_I\rangle\,\psi_I, T^*D_\lambda^\infty\varphi\right\rangle \quad .$$

$$= \lim_{\lambda\to\infty}\langle f, T^*D_\lambda^\infty\varphi\rangle \quad .$$

$$= \int Tf(x)\,dx \quad .$$

This proves (**??**).

It remains to dscuss P^*. Let again $J=[-2_0^k,2^{k_0}]$ contain the support of f. Let λ satisfy $c\lambda > B2^{k_0+1}$. Then

$$\left\langle\int P^*f, D_\lambda^\infty\varphi\right\rangle = \sum_I |I|^{-\frac{1}{2}}\epsilon_I\,\langle f,\phi_I\rangle\,\langle\psi_I, D_\lambda^\infty\varphi\rangle$$

$$\left\langle\int P^*f, D_\lambda^\infty\varphi\right\rangle = \sum_{I:|I|>|J|,I\cap J\neq\emptyset} |I|^{-\frac{1}{2}}\epsilon_I\,\langle f,\phi_I\rangle\,\langle\psi_I, D_\lambda^\infty\varphi\rangle$$

The last reduction follows because if $I\cap J=\emptyset$ then $\langle f,\phi_I\rangle = 0$ and if $|I|<|J|$ and $I\subset J$ then BI is contained in $[-c\lambda,c\lambda]$ and therfore $\langle\psi_I, D_\lambda^\infty\varphi\rangle = 0$. However, similarly to before,

$$\langle f,\phi_I\rangle \leq C|I|^{-\frac{3}{2}}$$

$$\langle\psi_I, D_\lambda^\infty\varphi\rangle \leq |I|^{\frac{3}{2}}\lambda^{-1}$$

and therfore

$$\sum_{I:|I|>|J|, I\cap J\neq\emptyset} |I|^{-\frac{1}{2}}\epsilon_I \langle f, \phi_I\rangle \langle \psi_I, D_\lambda^\infty \varphi\rangle$$

$$\leq \sum_{I:|I|>|J|, I\cap J\neq\emptyset} |I|^{-\frac{1}{2}}\lambda^{-1} \leq |J|^{-\frac{1}{2}}\lambda^{-1}$$

This tends to 0 as λ tends to ∞, and thus proves

$$\int P^* f(x)\, dx = 0 \quad .$$

This completes the discussion of the paraproducts and therfore the proof of the $T(1)$ theorem.

11 Wavelets

11.1 The scaling function

Let ψ be a smooth compactly supported function with mean zero. We fix the notation that for every dyadic interval $I = [2^k n, 2^k(n+1))$ we have

$$\psi_I = T_n D_{2^k}^2 \psi \quad .$$

Then an easy consequence of the $T(1)$ theorem is that the operator

$$Tf = \sum_I \langle f, \psi_I\rangle \psi_I$$

is bounded in L^2. One immediately observes that one can realax the assumptions on ψ quite a bit, for example $\psi \in \mathcal{S}(\mathbf{R})$ or ψ compactly supported and continuously differentiable would suffice.

Ofcourse if one wants that T is the identity operator, as we did in the proof of the $T(1)$ theorem, one has to impose stricter conditions on ψ. Such a ψ is called a wavelet. In the following sections we will study the existence and construction of wavelets. Naturally this theory has a bit more of an algebraic flavor, because one is interested in the algebraic identity that T as above is the identity operator.

The function ψ and its siblings ψ_I we are looking for should resemble the Haar functions h_I. It turns out, that the theory becomes easier and somewhat more natural if on constructs parallel a function ϕ and siblings ϕ_I which take the role of the characteristic functions χ_I. In fact we shall look at ϕ_I first.

We shall work entirely in the Hilbert space $L^2(\mathbf{R})$. A function ϕ is called scaling function, if

1. For all dyadic intervals I and J of the same length with $I \neq J$ we have

$$\langle \phi_I, \phi_J\rangle = 0$$

2. The function ϕ is in the closed linear span of all ϕ_I with $|I| = 1/2$. (Then an analog statement holds for all ϕ_J.)

3. The functions ϕ_I are a complete set in $L^2(\mathbf{R})$. Define V_k to be the closed linear span of $\{\phi_I : |I| \geq 2^k\}$, then the intersection of all V_k is empty.

The spaces V_k in this definition are sometimes said to form a multiresolution analysis.

At this stage of our discussion it is still not completely clear how to obtain scaling functions other than the Haar function. We postpone this point and first discuss how the scaling function leads to a wavelet.

From the analogy to the Haar system, it is plausible to look for wavelets in the orthogonal complement W_0 of V_0 inside V_{-1}. We shall first find a nice characterisation of such functions.

Every function in V_{-1} is given as a linear combination

$$f = \sum_{|I|=1/2} \epsilon_I \phi_I$$

for some square summable sequence ϵ_I. It will be useful to take the Fourier transform as follows

$$\begin{aligned} \widehat{f} &= \mathcal{F}\left(\sum_n \epsilon_{[n/2,(n+1)/2)} T_{n/2} \mathrm{D}^2_{1/2} \phi\right) \\ &= \sum_n \epsilon_{[n/2,(n+1)/2)} M_{n/2} \mathrm{D}^2_2 \widehat{\phi} = m_f(\xi) \mathrm{D}^2_2 \widehat{\phi} \quad , \end{aligned}$$

where m_f is the Fourier transform of ϵ_I in the sense

$$m_f(\xi) = \sum_n \epsilon_{[n/2,(n+1)/2)} e^{2\pi i (n/2)\xi}$$

Observe that m_f is periodic with period 2. The map $F : f \to m_f$ is obviously (up to a scalar factor of $\sqrt{2}$) an isometry $F : V_{-1} \to L^2([0,2])$ of Hilbert spaces.

The orthonormal basis $\{\phi_I : |I| = 1/2\}$ is mapped to the exponential basis. In particular we have

$$\widehat{\phi} = m\widehat{\phi_I} = m\mathrm{D}^2_2 \widehat{\phi} \quad .$$

We can read this equation as a constraint on ϕ to be a scaling function: $\widehat{\phi}/\mathrm{D}^2_2\widehat{\phi}$, if suitably defined, should be a 2-periodic, locally L^2 function.

Observe that

$$\begin{aligned} \langle Tn\phi, \phi \rangle = \left\langle M_n \widehat{\phi} \widehat{\phi} \right\rangle &= \int e^{2\pi i \xi n} |\widehat{\phi}(\xi)|^2 \, d\xi \\ &= \sum_l \int_0^1 e^{2\pi i \eta n} |\widehat{\phi}(\eta + l)|^2 \, d\eta \end{aligned}$$

This gives

$$\sum_l |\widehat{\phi}(\xi + l)|^2 \, d\xi = \|\phi\|_2^2 = 1$$

This gives the following identity for the function m_ϕ:

$$1 = \sum_n \left| m_\phi(\xi + n) D_2^2 \widehat{\phi}(\xi + n) \right|^2$$

$$= \sum_{n \ \text{even}} \left| m_\phi(\xi) D_2^2 \widehat{\phi}(\xi + n) \right|^2 + \sum_{n \ \text{odd}} \left| m_\phi(\xi) D_2^2 \widehat{\phi}(\xi + n) \right|^2$$

$$= \frac{1}{2} \left(|m_\phi(\xi)|^2 + |m_\phi(\xi + 1)|^2 \right) \tag{38}$$

Clearly the image $F(V_0)$ is the closed linear span of all functions $m_\phi(\xi)e^{2\pi i \xi n}$. We know that these functions form (up to scalar) an orthonormal basis of $F(V_0)$. We could have derived (**??**) from this information without reference to the scaling function.

We can identify $L^2([0,2])$ with the space $L^2([0,1], \mathbf{R}^2)$ (all square integrable functions on $[0,1]$ with values in the Hilbert space $\mathbf{R}^2$) via the map G defined by

$$Gm(\xi) = (m(\xi), m(\xi + 1))$$

The space $GF(V_0)$ is then the space of all $\mu G(m_\phi)$ with $\mu \in L^2[0,1]$, and the identification of $f \in V_0$ with $\mu_f \in L^2([0,1])$ is (up to scalar) an isometry.

Define $\psi \in V_{-1}$ by

$$m_\psi(\xi) = e^{\pi i \xi} m_\phi(\xi + 1) \quad .$$

Then for all $\xi \in [0,1]$ we easily see that $Gm_\psi(\xi)$ is orthogonal to $Gm_\phi(\xi)$. Thus the orthogonal complement of $GF(V_0)$ is easily seen to be the set of all functions $\mu G(m_\psi)$ with $\mu \in L^2([0,1])$.

Since we have

$$1 = \frac{1}{2} \left(|m_\psi(\xi)|^2 + |m_\psi(\xi + 1)|^2 \right)$$

we see that the set of all $T_n \psi$ forms an orthonormal basis of its span. Moreover, this span is clearly equal to the complement W_0 of V_0 in V_{-1}.

Thus the set of ψ_I with I dyadic forms an orthonormal basis of $L^2(\mathbf{R})$

11.2 Regularity of wavelets

So far we have only discussed the L^2 theory of wavelets. In general one is certainly interested in having smoothness and decay of wavelets. However, just as in the Balian Low theorem for Gabor bases, it turns out that one cannot have Schwartz functions which are wavelets.

Lemma 13 *Let $\psi \in L^2(\mathbf{R})$ be a generating wavelet, i.e., the set ψ_I is an orthonomral basis of $L^2(\mathbf{R})$. If ϕ has n bounded continuous derivatives and*

$$|\phi(x)| \leq C(1 + |x|^{-n-1}) \quad ,$$

then $\widehat{\phi}(0)$ has a zero of order at least n.

Observe that the conclusion of the lemma can be read that ϕ has vanishing moments, i.e., its integral against all polynomials of degree at most $n-1$ vanishes. The idea of the proof is that we will pair ϕ with ϕ_I with $|I|$ very small, the scalar product being 0. Then The decay of ϕ_I allows to pretend ϕ_I is supported only near I. The differentiability assumption on ϕ allows us to pretend it equals its Taylor polynomial of degree n near the small interval I. Thus ϕ_I integrated against this polynomial vanishes. By varying I we can vary the polynomial, in fact we can obtain a basis of the set of polynomials of degree n this way. This will prove the lemma.

Proof

The proof goes by induction by n. The lemma is trivial for $n=0$. Let $n \geq 0$ and assume the lemma has been proved for all $n' \leq n$.

If n is large, we can easily see that we find a primitive $D^{-1}\phi$ (abuse of notation) of ϕ which satisfies

$$\left|D^{-1}\phi(x)\right| \leq C(1+|x|^{-n-1}) \quad ,$$

Namely, we just consider the integral $\int_{-\infty} x\phi(t)\,dt$, which gives the correct decay for $x \to -\infty$, and the integral $-\int_x^{-\infty} \phi(t)\,dt$, which gives the correct decay for $x \to +\infty$, and we observe that these two functions coincide because by induction hypothesis ϕ has mean 0. Inductively we see that there is a n-th primitive of ϕ which satisfies

$$\left|D^{-n}\phi\right| \leq C(1+|x|)^{-2} \quad .$$

An easy argument using the full induction hypothesis shows that $\mathcal{F}(D^{-1}\phi)$ vanishes of order $n-1$.

We do n partial integrations to obtain

$$0 = \langle \phi_I, \phi \rangle = \left\langle D^{-n}\phi_I, D^n\phi \right\rangle$$

The right hand side is given by an abslotely integrable integral.

Now let J be any interval and I be a dyadic intervals such that $AI \subset J$ for some large A. Let M be the mean of $D^n\phi$ on J,and let M' be the mean of $D^{-n}\phi$. If we assume that $M \neq 0$, then

$$|M'| = |I|^{-\frac{1}{2}} \left| \int D^{-n}\phi_I(x)\,dx \right|$$

$$\leq |I|^{-\frac{1}{2}} \left| \int_J D^{-n}\phi_I(x)\,dx \right| + |I|^{-\frac{1}{2}} \|D^{-n}\phi_I\|_{L^1(J^c)}$$

$$\leq M^{-1}|I|^{-\frac{1}{2}} \left| \int_J D^{-n}\phi_I(x) D^n\phi(x)\,dx \right| + |I|^{-\frac{1}{2}} \left\|D^{-n}\phi_I\right\|_1 |J| \|D^{n+1}\phi\|_\infty + CA^{-1}$$

$$\leq M^{-1}|I|^{-\frac{1}{2}} \left| \int D^{-n}\phi_I(x) D^n\phi(x)\,dx \right| + M^{-1}|I|^{-\frac{1}{2}} \|D^{-n}\phi_I\|_{L^1(J^c)} \|D^n\phi\|_\infty + C|J| + CA^{-1}$$

$$\leq CA^{-1} + C|J| + CA^{-1}$$

Choosing A large enough gives $M' \leq CJ$. If M' was not zero, this would imply that the averages of ϕ over all sufficiently small intervals where 0, which would imply that

ϕ was 0. This is a contradiction. Hence $M' = 0$ and $D^{-n}\phi$ has mean 0. This proves that $\widehat{\phi}$ vanishes of order $n+1$.
■

Assume we have a scaling function which does not have mean 0 (which is natural to assume). Then the conclusion of the above lemma says that

$$\widehat{\psi} = m_\psi D_2^2 \widehat{\phi}$$

vanishes of order n at 0. Hence m_ψ vanishes of order n at 0. This implies that m_ϕ vanishes of order n at 1.

11.3 Construction of the scaling function

We would like to see examples of scaling functions. The characteristic function of $[0,1]$ is a scaling function. It is interesting to observe by an easy calculation (exercise) that teh above framework gives the Haar function as corresponding wavelet ψ.

Next, we try for V_0 being the set of all coninuous, piecewise linear functions with discontinuities in the derivative at most at the integer points. Clearly the scaled copies of V_0 are nested as required, and qualify for a multiresolution analysis.

The point is to find an orthogonal basis of translates in V_0. We can make the ansatz

$$f(n) = \begin{array}{ll} (-1)^n \gamma^n & \text{if } n \le 0 \\ a\gamma^{-n} & \text{if } n > 0 \end{array}$$

The parameters γ and a can be chosen so that f is orthogonal to $T_1 f$ and $T_2 f$. Then it follows easily (exercise) that f is orthogonal to all $T_l f$ with $l > 0$, which is what we needed. The corresponding set of wavelets is called the Franklin system, it consists again of linear splines.

One can do such constructions for splines of any order. These spline constructions however do not give compactly supported wavelets.

It turns out that one obtains compactly supported wavelets if one starts with a compactly supported scaling function. Then only finitely many terms in

$$\phi = \sum_{|I|=1/2} \epsilon_I \phi_I$$

are nonzero, because $\langle \phi, \phi_I \rangle = 0$ for all I sufficiently far away from the origin. Thus m_ϕ is a trigonometric polynomial. By definition, m_ψ is also a trigonometric polynomial. Thus ψ is also a finite linear combination of ϕ_I with $|I| = 1/2$, and hence it is also compactly supported.

Before w econstruct ψ, we will try to find candidates for the function m_ϕ. In fact we start with $M = 2|m_\phi|^2$. Then M is also a trigonometric polynomial of period 2, is positive, satisfies

(39) $$M(\xi) + M(\xi + 1) = 1$$

and, since we would like differentiability of the wavelet, vanishes of order $2n$ at 1. We make the ansatz

$$M(\xi) = P(1 - \cos^2(\pi\xi/2)) \cos^2(\pi\xi/2) \quad ,$$

where P is some polynomial. This guarantees the vanishing of the right order at 1.

Setting $\cos^2(\xi/2) = y$ then (**??**) will be satisfied if we have for P the identity

$$y^n P(1-y) + (1-y)^n P(y) = 1 \quad . \tag{40}$$

The polynomials y^n and $(1-y)^n$ have no common divisors, hence abstract algebra (Euclid's algorithm) tells us that the equation

$$y^n g_1(1-y) + (1-y)^n g_2(y) = 1$$

has exactly one solution g_1, g_2 with two polynomials g_1 and g_2 of order at most $n-1$. By symmetry and uniqueness we observe $g_1(y) = g_2(1-y)$. Hence (**??**) has exactly one solution P of degree at most $n-1$ (there are more solutions of higher degree ofcourse). We try to calculate the Taylor series of P around $y = 0$. We observe for y close to 0:

$$P(y) = (1-y)^{-n} \left[1 - y^n P(1-y)\right]$$

This gives

$$P(y) = (1-y)^{-n} + O(y^n)$$

By taking derivatives we can easily calculate the first $n-1$ coefficients of P and since we know P has order at most $n-1$ we obtain

$$P(y) = \sum_{k=0}^{n-1} \binom{n-1+k}{k} y^k$$

Observe that P is nonnegative for $y \geq 0$, hence M is nonnegative. Thus we can find a function m such that $|m|^2 = M$. The point however is to choose m so that it is again a trigonometric polynomial.

Obviously we can write

$$M(\xi) = \sum_{k=0}^{n-1} a_k \cos^k(\pi\xi)$$

with real numbers a_k. We can find and α and roots c_j such that

$$M(\xi) = \alpha \prod_{j=1}^{n-1} (\cos(\pi\xi) - c_j) \quad .$$

Set $z = e^{i\pi\xi}$. Then we have

$$M(\xi) = \alpha \prod_{j=1}^{n-1} \left(\frac{z + \overline{z}}{2} - c_j\right)$$

Since $|z| = 1$ we have $\overline{z} = z^{-1}$ and we can write

$$|M(\xi)| = \left| z^n \alpha \prod_{j=1}^{n-1} \left(\frac{z + \overline{z}}{2} - c_j \right) \right|$$

$$= \left| 2^{-n} \alpha \prod_{j=1}^{n} \left(1 - 2c_j z + z^2 \right) \right|$$

On the right hand side we have a polynomial in z. We plan to split this polynomial into linear factors. The factor

$$1 - 2c_j z + z^2$$

has two roots whose product is 1, hence they aree of the form r_j and r_j^{-1}. If r_j is complex, then since $r_j + r_j^{-1}$ is real we have that $|r_j| = 1$. Thus $r_j = \overline{r_j^{-1}}$ corresponds to a zero of M and therfore must be a double (or higher, but even) zero because M is positive. If c_j is not real, then the complex conjugate $\overline{c_j}$ also appears and we have a quadruple of roots r_j, r_j^{-1}, $\overline{r_j}$ and $\overline{r_j{}^{-1}}$.

Thus we see that we can pair a root r_j always with $\overline{r_j{}^{-1}}$, and therfore we can write

$$|M(\xi)| = \left| 2^{-n} \alpha \prod_{j=0}^{n-1} (z - r_j)(z - \overline{r_j{}^{-1}}) \right| \quad .$$

For $|z| = 1$ we have

$$|z - \overline{r_j{}^{-1}}| = \left| \overline{z^{-1} - r_j{}^{-1}} \right| = |r_j|^{-2} \, |z - r_j| \quad .$$

Therefore

$$|M(\xi)| = |\alpha'| \prod_{j=0}^{n-1} |z - r_j|^2 \quad .$$

Defining

$$m(\xi) = \sqrt{|\alpha'|/2} \prod_{j=0}^{n-1} \left(e^{i\pi\xi} - r_j \right)$$

We see that m is a trigonometric polynomial with $|m|^2 = M$.

From the above discussion we also see that each r_j is either real or can be paired with its complex conjugate. Therefore we can assume that

$$m(\xi) = \sum_{k=0}^{n-1} c_k e^{i\pi k\xi} \tag{41}$$

with real coefficients c_k.

Now that we have a candidate for $m = m_\phi$, we try to find ϕ. To avoid factors of $\sqrt{2}$ we define $\tilde{m} = \sqrt{2}m$

The identity we need for ϕ is

$$\phi(\xi) = \tilde{m}(\xi)\phi(2^{-1}\xi)$$

Assuming that $\widehat{\phi}(0) = 1$ and $\widehat{\phi}$ is continuous at 0 we obtain by formal iteration

$$\phi(\xi) = \prod_{j=0}^{\infty} \tilde{m}(2^{-j}\xi) \quad . \tag{42}$$

We shall study the existence of the limit on the right hand side.

We know that $\tilde{m}$ is a trigonometric polynomial, hence it is an entire function, and it satisfies satisfies $\tilde{m}(0) = 1$. Thus we have the estimate

$$|m(z) - 1| \leq C|z|$$

for $|z| \leq 1$ and, using (**??**) again,

$$m(z) \leq Ce^{C|Im(z)|}$$

for all z.

First we study (**??**) on the ball $|z| \leq 1$. Since

$$\sum_{j=0}^{\infty} |M(z) - 1| \leq \sum_{j=0}^{\infty} C2^{-j} \leq 2C$$

is finite uniformly in $|z| \leq 1$, by elementary calculus the product (**??**) converges uniformly for $|z| \leq 1$. Since all factors are holomorphic, it converges to a holomorphic function.

Let B_k the ball around the origin of radius 2^k. We have

$$\prod_{j=0}^{\infty} \tilde{m}(2^{-j}\xi) = \prod_{j=0}^{k} \tilde{m}(2^{-j}\xi) \prod_{j=0}^{\infty} \tilde{m}(2^{-j}2^{-k-1}\xi) \quad .$$

We see from the previous result that the second factor on the right hand side converges, hecne the whole product converges. To estimate this function, weonly have to estimate the first factor on the right hand side (the second one being bounded by a constant)

$$\begin{aligned} \left| \prod_{j=0}^{k} \tilde{m}(2^{-j}z) \right| &\leq \prod_{j=0}^{k} Ce^{C2^{-j}|Im(z)|} \\ &\leq C^k e^{2C|Im(z)|} \end{aligned}$$

For all z in the annulus $B_k \setminus B_{k-1}$ we therefore have

$$\left| \prod_{j=0}^{\infty} \tilde{m}(2^{-j}) \right| \leq C|z|^C e^{C|Im(z)|}$$

By the Paley Wiener theorem we find a compactly supported distribution ϕ such that

$$\widehat{\phi}(\xi) = \prod_{j=0}^{\infty} \tilde{m}(2^{-j}\xi) \quad .$$

We would like to see that ϕ or equivalently $\widehat{\phi}$ is in L^2. To this end define on $\mathbf{R}$ for $k > 0$

$$f_k = \prod_{j=0}^{k} \tilde{m}(2^{-j}\xi) 1_{[-2^k, 2^k]}(\xi)$$

Then we have, using that $\tilde{m}$ has period 2,

$$\|f_k\|_2^2 = \int_{-2^k}^{2^k} \prod_{j=0}^{k} |\tilde{m}(2^{-j}\xi)|^2 \, d\xi$$

$$\|f_k\|_2^2 = \int_0^{2^{k+1}} \prod_{j=0}^{k} |\tilde{m}(2^{-j}\xi)|^2 \, d\xi$$

$$\|f_k\|_2^2 = \int_0^{2^k} \prod_{j=0}^{k-1} |\tilde{m}(2^{-j}\xi)|^2 \left[|\tilde{m}(2^{-k}\xi)|^2 + |\tilde{m}(2^{-k}\xi + 1)|^2 \right] \, d\xi$$

$$\|f_k\|_2^2 = \int_0^{2^k} \prod_{j=0}^{k-1} |\tilde{m}(2^{-j}\xi)|^2 \, d\xi$$

$$= \|f_{k-1}\|_2^2$$

By induction

$$\|f_k\|_2^2 = \|f_0\|_2^2 = \int_{-1}^{1} \left| \tilde{m}(\xi)^2 \right| \, d\xi = \int_0^1 1 \, d\xi = 1 \quad .$$

Since f_k converges pointwise to $\widehat{\phi}$, we have that $\widehat{\phi}$ is in L^2 by Fatou's lemma.

Thus we have our compactly supported $\phi \in L^2$ as a candidate for the scaling function.

Given ϕ we define ψ as before. We claim that for each $f \in L^2$ we have

$$\|f\|_2^2 = \sum_I |\langle f, \psi_I \rangle|^2 \quad .$$

To this end we first observe that

$$\sum_{|I|=1/2} |\langle f, \phi_I \rangle|^2 = \sum_{|I|=1} |\langle f, \phi_I \rangle|^2 + \sum_{|I|=1} |\langle f, \psi_I \rangle|^2$$

Namely, the left hand side is equal to

$$\sum_{|I|=1/2} \left| \left\langle \widehat{f}, \widehat{\phi}_I \right\rangle \right|^2$$

$$= \sum_n 2^{-1} \left| \int \widehat{f}(\xi) \widehat{\phi}(\xi/2) e^{\pi i n \xi} \, d\xi \right|^2$$

By Plancherel, this is equal to

$$= \sum_n 2^{-\frac{1}{2}} \int_0^2 \left| \sum_k \widehat{f}(\xi + 2k) \widehat{\phi}(\xi/2 + k) \right|^2 \, d\xi$$

$$= \sum_n \int_0^1 \left| \sum_k \widehat{f}(\xi+2k)\widehat{\phi}((\xi+2k)/2) \right|^2 + \left| \sum_k \widehat{f}(\xi+2k+1)\widehat{\phi}((\xi+2k+1)/2) \right|^2 \, d\xi$$

$$= \sum_n \int_0^1 \left| \sum_k \widehat{f}(\xi+2k)\widehat{\phi}((\xi+2k)/2) m_\phi(\xi+2k)/2 + \right.$$

$$\left. \sum_k \widehat{f}(\xi+2k+1)\widehat{\phi}((\xi+2k+1)/2) m_\phi(\xi+2k+1) \right|^2 \, d\xi$$

$$= \sum_n \int_0^1 \left| \sum_k \widehat{f}(\xi+2k)\widehat{\phi}((\xi+2k)/2) m_\psi(\xi+2k)/2 + \right.$$

$$\left. \sum_k \widehat{f}(\xi+2k+1)\widehat{\phi}((\xi+2k+1)/2) m_\psi(\xi+2k+1) \right|^2 \, d\xi$$

$$= \sum_n \int_0^1 \left| \sum_k \widehat{f}(\xi+k)\widehat{\phi}(\xi+k) \right|^2 \, d\xi + \int_0^1 \left| \sum_k \widehat{f}(\xi+k)\widehat{\psi}(\xi+k) \right|^2 \, d\xi$$