HARMONIC ANALYSIS SUMMER 2020, PART B

CHRISTOPH THIELE

15. Lecture: Convergence of Fourier series

Let $L^2(\mathbb{T})$ denote the space of Radon measures m on $\mathbb{T}$ which satisfy

$$\sum_{n\in\mathbb{Z}} |\widehat{m}(n)|^2 < \infty,$$

where

$$\widehat{m}(n) = m(e^{-2\pi in\cdot}).$$

Recall that $L^2(\mathbb{T})$ is a Hilbert space with norm

$$\|m\|_2 = (\sum_{n\in\mathbb{Z}} |\widehat{m}(n)|^2)^{1/2}.$$

In particular, completeness of this space follows from completeness of the space of square summable sequences, since every sequence a_n with $\sum_{n\in\mathbb{Z}} |a_n|^2 < \infty$ is realized by a Radon measure in the sense $a_n = \widehat{m}(n)$. This measure is the limit measure of the harmonic function

$$\sum_n a_n r^{|n|} e^{2\pi in\theta}$$

in the unit disc. Recall also the alternative expressions for the norm,

$$\|m\|_2 = \lim_{r\to 1}(\int_0^1 |u(re^{2\pi i\theta})|^2)^{1/2}$$
$$= \lim_{k\to -\infty}(\sum_{I\in\mathcal{D}_k} |I||F(I))|^2)^{1/2}$$
$$= (|F([0,1))|^2 + \sum_{I\in\mathcal{D}} |I||\Delta F(I))|^2)^{1/2}.$$

For a Radon measure m define the partial sums

$$S_N m(\theta) = \sum_{n=-N}^{N} \widehat{m}(n) e^{2\pi in\theta}.$$

We have the following trivial convergence result in the sense of the norm in $L^2(\mathbb{T})$.

Theorem 65. *For $m \in L^2(\mathbb{T})$ we have*

$$\lim_{N\to\infty} \|S_N f - f\|_2 = 0.$$

Date: July 3, 2020.

Proof. We have

$$\|S_N m - m\|_2^2 = \sum_{|n|>N} |\widehat{m}(n)|^2,$$

and by dominated convergence the right-hand-side tends to 0 as N tends to ∞. □

The following theorem is a much stronger convergence result.

Theorem 66 (Carleson). *Let m be a Radon measure in $l^2(\mathbb{T})$ with martingale extension F. Then for almost every $\theta \in \mathbb{T}$, both limits in the following equation exist and satisfy this equation.*

$$\lim_{N\to\infty} S_N f(\theta) = \lim_{k\to-\infty} F(I_{k,\theta}).$$

The existence of the martingale averages almost everywhere was shown before. This shows the theroem in the special case when only finitely many Fourier coefficients are non-zero.

A more general special case of the theorem is also relatively easy to prove. Assume m is given by integration against a twice continuously differentiable function f. Then the identity again holds for all $\theta \in [0,1)$. The limit on the right hand side is then the value $f(\theta)$. Moreover, for $n \neq 0$

$$\widehat{m}(n) = \int f(x)e^{-2\pi inx}dx = \frac{1}{2\pi in}\int f'(x)e^{-2\pi inx}dx$$
$$= \frac{-1}{4\pi^2 n^2}\int f''(x)e^{-2\pi inx}dx,$$

which implies that

$$|\widehat{m}(n)| \leq \frac{C}{1+n^2}.$$

Hence the Fourier series is absolutely summable on the closed unit disc and the partial sum converge uniformly to the function f on $\mathbb{T}$.

We define the truncated Carleson maximal operator.

$$C_N m(x) = \sup_{0\leq n\leq N} |S_n m(x)|.$$

As a supremum of a collection of continuous functions, $C_N f$ is a continuous function and as such determines an element in $L^2(\mathbb{T})$.

The proof of Theorem 68 will rely on the next theorem.

Theorem 67 (Carleson-Hunt). *There is a constant C independent of N such that, for all $m \in L^2(\mathbb{T})$ and all $N > 0$,*

$$\|C_N m\|_2 \leq C\|m\|_2.$$

As N increases, so does $C_N m$ pointwise as well as its martingale extension. By monotone convergence,

$$C_\infty m(x) = \sup_{0\leq n} |S_n m(x)|.$$

is also in $L^2(\mathbb{T})$, and we have with the same constant as in the last theorem

$$\|C_\infty m\|_2 \le C\|m\|_2.$$

We now prove Theorem 68 using Theorem 69.

Proof. We need to show that for $\delta > 0$ there is some collection $\mathcal{I}$ of daydic intervals with $\sum_{\mathcal{I}} |I| \le \delta$ that for $\theta \notin \bigcup I$

$$\liminf_{N\to\infty} \liminf_{k\to-\infty} |S_N m(\theta) - F(I_{k,\theta})| = 0.$$

Adding and subtracting the $I_{k,\theta}$ martingale average of S_N inside the absolute value sign , and using that the martingale averages of the continuous function S_N converge to $S_N m$, this follows from

$$\liminf_{N\to\infty} \liminf_{k\to-\infty} |I_{k,\theta}|^{-1} |\int_{I_{k,\theta}} S_N m - m| = 0$$

It suffices to show that for $\epsilon > 0$ there exists a collection $\mathcal{I}_\epsilon$ of dyadic intervals with $\sum_{\mathcal{I}_\epsilon} |I| \le \epsilon$ that for $\theta \notin \bigcup I$

$$\liminf_{N\to\infty} \liminf_{k\to-\infty} |I_{k,\theta}|^{-1} |\int_{I_{k,\theta}} S_N m - m| < \epsilon$$

Namely, then we may pick such $\mathcal{I}_\epsilon$ for ϵ rapidly going to zero such that the sum of these ϵ is less than δ and use the union of all collections $\mathcal{I}_\epsilon$ to establish the previous.

Given $\epsilon > 0$, pick N_0 large enough so that

$$\|S_{N_0} m - m\|_2 < c\epsilon^2.$$

for some sufficiently small c to be determined momentarily.

Let $\mathcal{I}_1$ be the set of maximal dyadic intervals such that the martingale average

$$|I|^{-1} \int_I |m - S_{N_0} m| > \epsilon/2$$

Let $\mathcal{I}_2$ be the set of maximal dyadic intervals such that the martingale average

$$|I|^{-1} \int_I C_\infty(m - S_{N_0} m) > \epsilon/2$$

Let $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$. We have by Chebysheff for sufficiently small c

$$\sum_{I\in\mathcal{I}_1} |I| \le \epsilon^{-2} \|m - S_{N_0} m\|_2^2 \le \epsilon/2$$

and by boundedness of the Carleson operator

$$\sum_{I\in\mathcal{I}_1} |I| \le \epsilon^{-2} \|C_\infty(m - S_{N_0} m)\|_2^2 \le epsilon/2.$$

Let $\theta \notin \bigcup\mathcal{I}$. It suffices to show that for all $N > N_0$ and all k we have

$$|I_{k,\theta}|^{-1} |\int_{I_{k,\theta}} S_N m - m| < \epsilon \tag{1}$$

By linearity, we have with $S_N S_{N_0} m = S_{N_0} m$ for $N > N_0$

$$S_N m - m = S_N(m - S_{N_0} m) - (m - S_{N_0} m).$$

Hence we have (1) by the triangle inequality definition of C_∞ for $N > N_0$.

□

We write

$$\tilde{S}_N = \sum_{n<N} \widehat{m}(n)e^{2\pi i n\theta}$$

and note that

$$S_N = \tilde{S}_N - \tilde{S}_{-N-1}$$

Defining a maximal operator

$$\tilde{C}_\infty m(\theta) = \sup_N |\tilde{S}_N m(\theta)|,$$

boundedness of C_∞ follows from boundedness of $\tilde{C}_\infty$.
Define modulation by n as

$$M_n m(f) = m(fe^{2\pi i n \cdot})$$

We have the following invariances

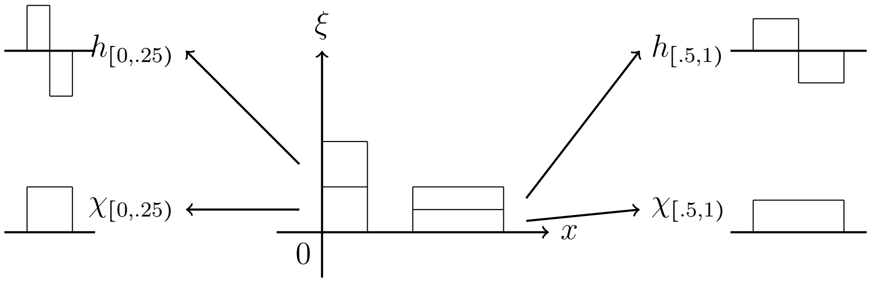
$$\widehat{M_n m}(j) = (M_n m)(e^{-2\pi i j \cdot}) = m(e^{-2\pi i (j-n) \cdot}) = \widehat{m}(j-n)$$

$$\tilde{S}_N(M_n m) = \sum_{j<N} \widehat{m}(j-n)e^{2\pi i j \cdot} = \sum_{j<N-n} \widehat{m}(j)e^{2\pi i (j+n) \cdot} = M_n(\tilde{S}_{N-n} m).$$

Hence

$$\tilde{C}_\infty(M_n m) = \tilde{C}_\infty m$$

The invariance of C_∞ under M_N suggests that to the typical translation and dilation parameter in our outer measure spaces, there should be an additional modulation parameter when specifying embedding maps and outer measure theory.
We shall first discuss a dyadic model of this outer emasure space. Unfortunately, this dyadic model will not directly be applicable to prove the Carleson-Hunt theorem, but a discrete variant of it.One needs a similar but different outer measure space to prove the Carleson-Hunt theorem.
The dyadic model in this case is also called the *Walsh model*. We work with the conditions $(x, \xi) \in [0,1) \times \mathbb{N}$.



We use the functions

$$1_I = 1_{I_l} + 1_{I_r},$$
$$h_I = 1_{I_l} - 1_{I_r}.$$

or

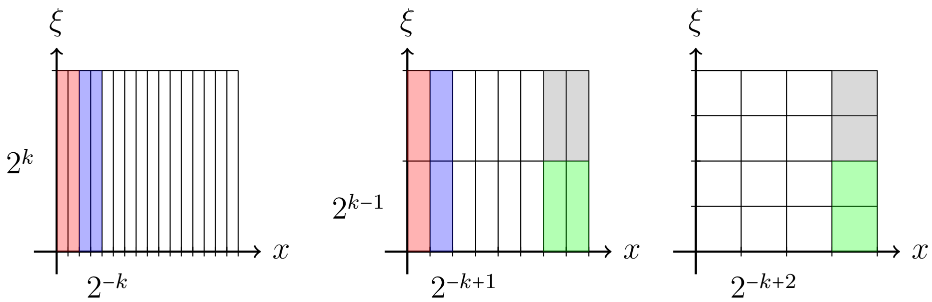
$$\begin{pmatrix} 1_I \\ h_I \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1_{I_l} \\ 1_{I_r} \end{pmatrix}.$$

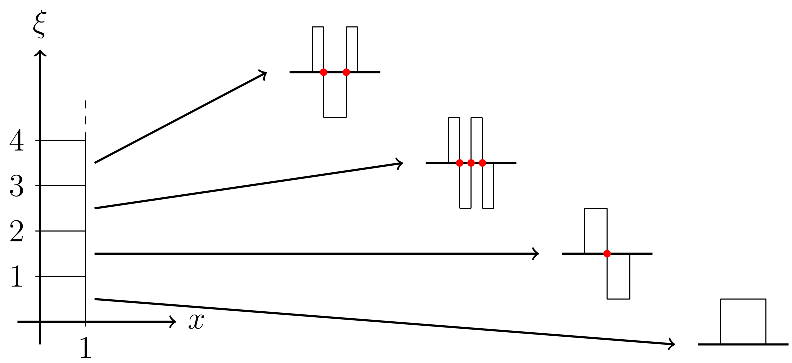
Let $\mathfrak{p}$ be the set of tiles, that is dyadic rectangles $I \times \omega$ (product of dyadic intervals), such that $I \subset [0,1)$ and $\omega \subset [0,\infty)$ and $|I||\omega| = 1$. We define the Walsh wave packet map $w\colon \mathfrak{p} \to L^2(\mathbb{R}_+)$ such that

- $w(I \times [0, |I|^{-1})) = 1_I$;
- if $|I||\omega| = 2$, then

$$\begin{pmatrix} w(I \times \omega_l) \\ w(I \times \omega_r) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} w(I_l \times \omega) \\ w(I_r \times \omega) \end{pmatrix}. \tag{2}$$

To see that this defines a unique wave packet for each tile $I \times \omega$, we call for $\omega = [2^k n, 2^k(n+1))$ the parameter n the height of the interval. We induct on the height of the second interval of a tile. Note that the height of the intervals on the left of the recusrion is always larger or equal to the heights of the tiles on the right, where equality holds precisely for height 0. Tiles with height zero are associated with characteristic functions, and the recursion is consistent for wava packets of height zero on the left hand side. For any height other than zero, there is a unique instance of the recursion involving this tile on the left hand side, and we can use this instance to define the wave packet of the tile.





We write

$$W_n(x) = w([0,1) \times [n, n+1))$$

It turns out (exercise) that for every $n \in \mathbb{N}$ there is a unique $W_{n'}$ with this number of zero crossings. In this sense the W_n resemble real part of $e^{2\pi i n x}$.

The functions W_n are the characters of the gorup $(\mathbb{Z}/2\mathbb{Z})^\infty$, but we will not further elaborate on this now.

We define the Walsh partial sums

$$S_N^W m(x) = \sum_{n=0}^{N-1} m(W_n) W_n(x).$$

Theorem 68 (Billard-Carleson)**.** *Let m be a Radon measure in $l^2(\mathbb{T})$ with martingale extension F. Then for almost every $\theta \in \mathbb{T}$, both limits in the following equation exist and satisfy this equation.*

$$\lim_{N\to\infty} S_N^W f(\theta) = \lim_{k\to-\infty} F(I_{k,\theta}).$$

We define the truncated Walsh-Carleson maximal operator.

$$C_N^W m(x) = \sup_{0\le n\le N} |S_n^W m(x)|.$$

As a supremum of a collection of functions in S^Δ, $C_N^W m$ is a function in S^Δ and as such determines an element in $L^2(\mathbb{T})$.

Theorem 69 (Billard-Carleson-Hunt)**.** *There is a constant C independent of N such that, for all $m \in L^2(\mathbb{T})$ and all $N > 0$,*

$$\|C_N^W m\|_2 \le C\|m\|_2.$$

As in the continuous setting, this theorem implies the Billard-Carleson theorem.

We will prove these theorems in the next few lectures.

16. Walsh analysis

Fix a large K and consider the fine Walsh phase plane

$$\Omega_K = [0,1) \times [0, 2^{-K}).$$

A tile is a dyadic rectangle of area one in the Walsh phase plane,

$$p = I \times \omega = [2^k n, 2^k(n+1)) \times [2^{-k} l, 2^{-k}(l+1)) \subset \Omega_K$$

with integers k, n, l.

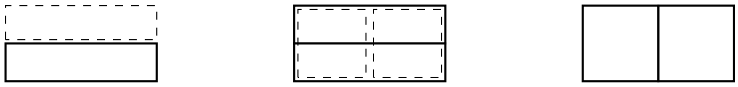
Theorem 70. *Let $\mathfrak{p}_1$ be a collection of pairwise disjoint tiles in Ω_K. Then there is a collection of pairwise disjoint tiles $\mathfrak{p}_2$ contained in Ω_K such that Ω_K is the disjoint union of $\bigcup \mathfrak{p}_1$ and $\bigcup \mathfrak{p}_2$.*

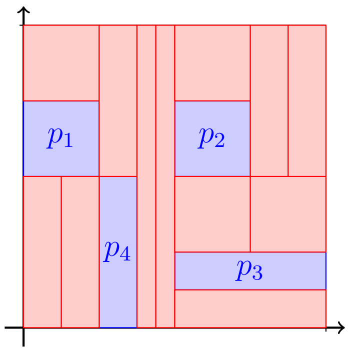
Proof. Call average spatial size A the average of the quantities $|I|$ for the tiles $I \times \omega$ in $\mathfrak{p}_1$. As we have only finitely many possible collections $\mathfrak{p}_1$, this average can only take finitely many values. The maximal possible average is 1, as we have $|I| \le 1$ for all tiles contained in Ω_K. If indeed $A = 1$, then $\mathfrak{p}_1$ consists only of tiles of the form $[0,1) \times [j, j+1)$ for $0 \le j < 2^{-K}$. We let $\mathfrak{p}_2$ be the collection of tiles of this form that are not in $\mathfrak{p}_1$. Then the properties claimed in the theorem are true.
Now assume the theorem is false, and let $\mathfrak{p}_1$ be a counterexample. Assume $\mathfrak{p}_1$ is a maximal counterexample with respect to number of tiles in $\mathfrak{p}_1$. We may also assume that the average A is maximal among all counterexamples with this number of tiles. By the above, $A < 1$. Let p we a tile in $\mathfrak{p}_1$ with minimal $|I|$. Then $|I| < 1$. Let J be the parent of I. As $|I| < 1$ the interval J is also contained in $[0,1)$.
Assume first that both $J_l \times \omega$ and $J_r \times \omega$ are in $\mathfrak{p}_3$ We consider a new collection $\mathfrak{p}_3$, which consists of the tiles of $\mathfrak{p}_1$ except that $J_l \times \omega$ and $J_r \times \omega$ are replaced by $J \times \omega_l$ and $J \times \omega_r$. The collection $\mathfrak{p}_3$ then also consists of pairwise disjoint tiles and has the same union as $\mathfrak{p}_1$ and therefore is also a counterexample. The spatial size average of $\mathfrak{p}_3$ is however larger than that of $\mathfrak{p}_1$, a contradiction to maximality of A.
It remains to consider the case that only one of $J_l \times \omega$ and $J_r \times \omega$ is in $\mathfrak{p}_1$. Assume first that $J \times \omega$ is disjoint from all tiles in $\mathfrak{p}_1$ other than $I \times \omega$. Then we may consider the collection of pairiwise disjoint tiles

$$\mathfrak{p}_4 = \{J_l \times \omega, J_r \times \omega\} \cup \mathfrak{p}_1.$$

It has one more tile than $\mathfrak{p}_1$ and can by maximality not be a counterexample. We can find a collection $\mathfrak{p}_5$ which complements $\mathfrak{p}_4$ in the sense of the theorem. But then $\mathfrak{p}_5$ together with the horizontal sibling of $I \times \omega$ complements $\mathfrak{p}_1$ in the sense of the theorem, a contradiction to $\mathfrak{p}_1$ being a counterexample.
Finally we assume that there is a tile $I' \times \omega'$ in $\mathfrak{p}_1$ other that $J_l \times \omega$ or $J_r \times \omega$ that intersects $J \times \omega$. We have $|I| \ge |I'|$ by minimality of I. As $I' \times \omega'$ is not $J_l \times \omega$ or $J_r \times \omega$, we have $|I| < |I'|$ and hence $J \subset I'$. Hecne $I' \times \omega'$ also intersects $I \times \omega$, a contradcition to disjointness of tiles in $\mathfrak{p}_1$.

Having brought all cases to a contradiction, we see that a counterexample to the statement of the theorem cannot exist. □





Recall the functions $w(I \times \omega)$ from the previous lecture.
Recall that S_K^Δ is the space of functions which are constant on all dyadic intervals of length 2^K.

Theorem 71. *Assume $\mathfrak{p}_1$ is a collection of pairwise disjoint tiles whose union is Ω_K. Then the functions $w(I \times \omega)$ are pairwise orthogonal and span S_K^Δ*

Proof. We again consider the average spatial size A of $\mathfrak{p}_1$ as in the proof of the previous theorem. The minimal possible value of $|I|$ is 2^{-K}. It is attained as average if $\mathfrak{p}_1$ consists only tiles of the form $I \times [0, 2^{-K})$. Since the union of $\mathfrak{p}_1$ is Ω_K, $\mathfrak{p}_1$ then contains all tiles of this form with $I \subset [0, 1)$. The wave packet of such a tile is the characteristic function on I. The span of all these functions clearly is S_K^Δ.
Assume $\mathfrak{p}_1$ is a counterexample to the statement of the theorem, and assume among all counterexamples it has minimal average A. Consider a tile I in $\mathfrak{p}_1$ such that $|I|$ is maximal. We may assume $|I| > 2^K$. Consider the sibling ω' of ω. Since $|\omega| < 2^K$, the sibling ω' is also contained in $[0, 2^{-K})$ and thus $I \times \omega'$ is in Ω_K. By similar arguments as in the proof of the previous theorem, $I \times \omega'$ has to be in $\mathfrak{p}_1$. Now consider a collection $\mathfrak{p}_2$ which arises from $\mathfrak{p}_1$ by replacing $I \times \omega$ and $I \times \omega'$ by $I_l \times \omega''$ and $I_r \times \omega''$, where ω'' is the parent of ω. Then this new collection also is a partition of Ω_K. Since it has smaller spatial size average than the collection $\mathfrak{p}_1$, it is not a counterexample and its wave packets are pairwise orthogonal functions whose span is S_K^Δ. However, since the functions $w(I \times \omega)$ and $w(I \times \omega)$ arise by an orthogonal transformation from the functions $w(I_l \times \omega'')$ and $w(I_r \times \omega'')$, the collection of wave packets of $\mathfrak{p}_1$ has the same span as that of $\mathfrak{p}_2$ and also consists of pairwise orthogonal functions. Hence a counterexample to the statement of the theorem cannot exist. □

Recall that the wave packets $w(I \times \omega)$ have constant modulus 1 on their support and thus satisfy

$$\|w(I \times \omega)\|_{L^2} = |I|$$

Hence the orthogonal bases of the above theorem are not orthonormal. To obtain an orthonormal basis, one needs to consider the L^2 normalized wave packets

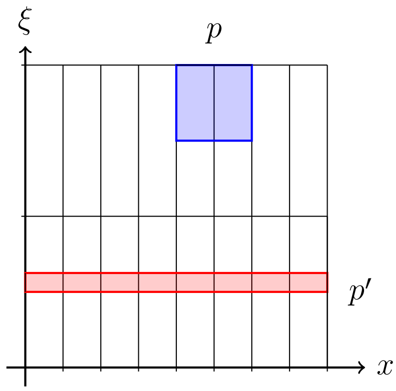
$$w(I \times \omega)|I|^{-1/2}.$$

Combining the previous theorems, we obtain the following consequences.

Theorem 72 (Orthogonality)**.** *Let $p, p' \in \mathfrak{p}$, $p \cap p' = \emptyset$. Then*

(1) $w(p) \perp w(p')$.

Proof. Consider the collection $\mathfrak{p}_1$ consisting of the two tiles p and p'. By the first theorem above we may complete this collection to a tiling of Ω_K with an additonal collection $\mathfrak{p}_2$. The wave packets to tiles of the collection $\mathfrak{p}_1 \cup \mathfrak{p}_1$ by the second theorem are pairwise orthogonal. In particular $w(p)$ and $w(p')$ are pairwise orthogonal. □



Theorem 73. *Let $\mathfrak{p}_1$ be a collection of pairwise disjoint tiles and let p be a tile with*

$$p \subset \bigcup \mathfrak{p}_1.$$

Then $w(p)$ is in the span of the wave packets of tiles in $\mathfrak{p}_1$.

Proof. First consider a collection $\mathfrak{p}_2$ so that $\mathfrak{p}_1 \cup \mathfrak{p}_2$ is a partition of Ω_K as in the first theorem above. Then consider the collection $\mathfrak{p}_2 \cup \{p\}$ of pairwise disjoint tiles and a collection $\mathfrak{p}_3$ so that $\mathfrak{p}_2 \cup \{p\} \cup \mathfrak{p}_3$ is a partition of Ω_K. By the second theorem above, the wave packets of tiles in $\mathfrak{p}_1$ span the orthogonal complement of those of $\mathfrak{p}_3$, and $w(p)$ is in the orthogonal complement of the span of wave packets of tiles of $\mathfrak{p}_3$. Hence $w(p)$ is in the span of the wave packets in $\mathfrak{p}_1$. □

Recall the Walsh functions

$$W_j = w([0,1) \times [j, j+1))$$

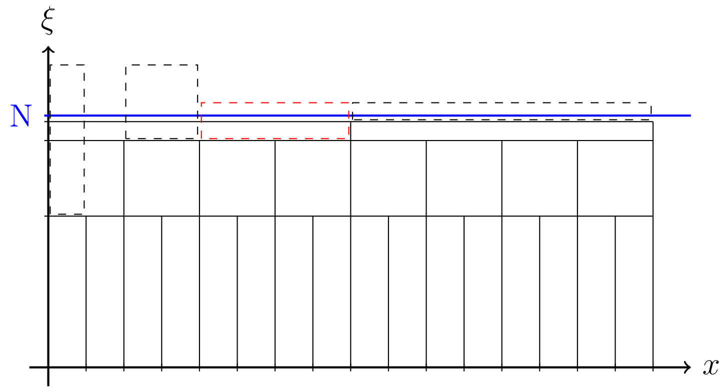
and the Walsh partial sum

$$S_N m(x) = \sum_{0\le j<N} m(W_j)W_j(x)$$

Theorem 74. *We have for every* $0 \le N \le 2^{-K}$

$$S_N m(x) = \sum_{|I||\omega|=2} |I|^{-1} m(w(I\times\omega_l))w(I\times\omega_l)1_{N\in\omega_r}.$$

Proof. Define $\mathfrak{p}_N$ to be the set of tiles $I\times\omega_l \subset \Omega_K$ such that $N\in\omega_r$.



We observe that the elements of $\mathfrak{p}_N$ are pairwise disjoint: suppose $(x,\xi)\in p\cap p'$ for some $p\ne p'$ in $\mathfrak{p}_N$. Since $x\in I\cap I'$, we may assume without loss of generality $I\subsetneq I'$. As the tiles are different, we obtain $|\omega_l|>|\omega_l'|$. But $\xi\in\omega_l\cap\omega_l'\ne\varnothing$ yields $\omega_r\cap\omega_r'=\varnothing$, giving a contradiction with $N\in\omega_r\cap\omega_r'$.
We next observe that the elements of $\mathfrak{p}_N$ cover $[0,1)\times[0,N)$. Namely if $(x,\xi)\in\bigcup\mathfrak{p}_N$, then $\xi<N$ and since $\bigcup\mathfrak{p}_N\subset\Omega_K$ we obtain $(x,\xi)\in[0,1)\times[0,N)$. Assume conversely that $(x,\xi)\in[0,1)\times[0,N)$. Then there is a smallest dyadic interval ω such that $\xi,N\in\omega$. Since $\xi<N$, we obtain $\xi\in\omega_l$ and $N\in\omega_r$. As N is an integer, ω has size at least 2. Moreover, ω_l is contained in $[0,2^{-K})$. There is a dyadic interval I of length $|\omega|^{-1}$ containing x, this interval needs to be contained in $[0,1)$. Then $(x,\xi)\in I\times\omega_l$ and $I\times\omega_l\in\mathfrak{p}_N$.
We have seen that $\mathfrak{p}_N$ consists of pairwise disjoint tiles covering the same set as the tiles $[0,1)\times[j,j+1)$ with $j<N$. Hecne the orthogonal projections onto these spans of wave packets are equal. This proves the theorem.

□

For every measure m we can find a function n constant on dyadic intervals of length 2^K such that

$$\sup_{0\le N\le 2^{-K}} S_N m(x) = S_{n(x)} m(x)$$

17. Walsh embedding theorems

Let $\mathcal{P}$ be the set of all 2^L-multi-tiles in Ω_K, that is all dyadic rectangles $I \times \omega$ of area $|I||\omega| = 2^L$ and $I \times \omega \subset \Omega_K$.
Given a dyadic interval $I \subset [0,1)$ and a number $\xi \in [0, 2^{-K}]$, we define the tree T with top interval $I_T = I$ and top frequency $\xi_T = \xi$ to be the set of all multi-tiles $I' \times \omega' = P' \in \mathcal{P}$ such that $I' \subset I_T$ and $\xi_T \in \omega$. Let $\mathcal{E}$ be the set of all trees.
For a tree T define $\sigma(T) = |I_T|$. Then we may define the corresponding outer measure for any subset A of $\mathcal{P}$

$$\mu(A) = \inf_{\mathcal{E}' \subset \mathcal{E} : A \subset \bigcup \mathcal{E}'} \sum_{T \in \mathcal{E}'} \sigma(T).$$

Consider functions $F : \mathcal{P} \to \mathbb{R}$. For a given tree T, we define the local norms

$$\ell^\infty F(T) = \sup_{P \in T} |F(P)|$$

and for $1 \le q < \infty$

$$\ell^q F(T) = \left(\frac{1}{\sigma(T)} \sum_{P \in T} |F(P)|^q \right)^{1/q}.$$

Further we define the outer quasi norms for $1 \le q \le \infty$

$$L^\infty \ell^q F = \sup_{T \in \mathcal{E}} \ell^q F(T)$$

and for $1 \le p < \infty$

$$L^p \ell^q F = \left(\int_0^\infty \inf\{\mu(A) : L^\infty \ell^q (F 1_{A^c}) \le \lambda^{1/p}\} \, d\lambda \right)^{1/p}.$$

These definitions are analoguous to the previously discussed outer measure theory on the set of dyadic intervals. We do have a number of analoguous results, such as for example

Theorem 75. *There is a constant C (independent of K), such that*

$$\sum_{P \in \mathcal{P}} |I||F(P)| \le C \|F\|_{L^1(\ell^1)}.$$

Similarly, we have quasi-subadditivity, Hölder's inequality, and so on. However, the spaces $L^p \ell^q$ other than $p = q = 1$ are not the center of our interest, instead we will introduce some hybrid spaces.
For each multi-tile $I \times \omega \in \mathcal{P}$ we enumerate the tiles $I \times \omega_i$ contained in $I \times \omega$ from bottom to top with indices $0 \le i < 2^L$. Then we define for each such i the hybrid sizes:

$$S_i(F)(T) = \sup_{P \in T} |F(I \times \omega)| + \sum_{j \ne i} \left(\frac{1}{|I_T|} \sum_{I \times \omega \in T, \xi_T \in \omega_j} |I||F(P)|^2 \right)^{1/2}.$$

We then have the following variant of a local Hölder inequality.

Theorem 76. *Consider a collection $\mathcal{I}$ of at least three different indices $0 \le i < 2^L$ and functions $F_i \mathcal{P} \to \mathbb{R}$ for each $i \in \mathcal{I}$. Then we have*

$$\ell^1(\prod_{i\in\mathcal{I}} F_i)(T) \le C \prod_{i\in\mathcal{I}} S_i(F_i).$$

Proof. We split the tree T into the subtrees

$$T_j = \{P \in T : \xi_T \in \omega_j\}$$

for each $0 \le j < 2^L$ For each such j we choose two indices j_1 and j_2 in $\mathcal{I}$ not equal to i. We estimate all $F_j(P)$ for j not equal to j_1 or j_2 by the supremum over all $P \in T$, and we do Cauchy Schwarz in the indices j_1 and j_2 to obtain

$$\begin{aligned}
\ell^1(\prod_{i\in\mathcal{I}} F_i)(T) &= |I_T|^{-1} \sum_{P\in T} \prod_{i\in\mathcal{I}} F_i(P) \\
&= \sum_{0\le j<2^L} |I_T|^{-1} \sum_{P\in T_j} \prod_{i\in\mathcal{I}} F_i(P) \\
&\le \sum_{0\le j<2^L} |I_T|^{-1} \left(\prod_{i\in\mathcal{I}, i\neq j_1, j_2} S_i F_i(T) \right) \sum_{P\in T_j} F_{j_1} F_{j_2}(P) \\
&\le \sum_{0\le j<2^L} \prod_{i\in\mathcal{I}} S_i F_i \le 2^L \prod_{i\in\mathcal{I}} S_i F_i
\end{aligned}$$

This proves the theorem with the constant $C = 2^L$. □

We obtain an analogue to Hölder's inequality

Theorem 77. *Consider a collection $\mathcal{I}$ of at least three different indices $0 \le i < 2^L$. Consider a collection of indices $1 \le p_i \le \infty$ such that*

$$\sum_{i\in\mathcal{I}} \frac{1}{p_i} = 1.$$

Then

$$L^1 \ell^1(\prod_{i\in\mathcal{I}} F_i) \le C \prod_{i\in\mathcal{I}} L^{p_i} S_i(F_i).$$

Proof. The proof is similar to Hölder's inequality in the dyadic setting, using the previous theorem. □

The usefulness of the hybrid sizes lies in the following embedding theorem.

Theorem 78. *Let $0 \le i < 2^L$. For $2 < p \le \infty$ there exists a constant C such that the following holds. Let m be a measure and define for $P = I \times \omega$ the embedded function*

$$F(P) = |I|^{-1} m(w(I \times \omega_i)).$$

Then

$$L^p S_i F \le C \|m\|_p.$$

Proof. By Marcinkiewicz interpolation, it suffices to show this estimate for $p = \infty$ and a weak endpoint estimate at $p = 2$. We begin with $p = \infty$. We need to show that, for all $T \in \mathcal{E}$

$$S_i F(T) \le C\|m\|_\infty.$$

For the first summand in the definition of S_i, we estimate

$$\sup_{I\times\omega\in T} |F(P)| \le \sup_{P\in\mathcal{P}} |I|^{-1}\|w(I\times\omega_i)\|_1\|m\|_\infty = \|m\|_\infty.$$

For the second sum in the definition of S_i we fix $j \ne i$ and compute

$$\frac{1}{|I_T|}\sum_{I\times\omega\in T,\xi_T\in\omega_j} |I|^{-1}|m(w_{I\times\omega_{I\times\omega_i}})|^2$$

$$= \frac{1}{|I_T|}\sum_{I\times\omega\in T,\xi_T\in\omega_j} |I|^{-1}|m(1_{I_T} w_{I\times\omega_{I\times\omega_i}})|^2 = \frac{1}{|I_T|}\|m1_{I_T}\|_2^2 \le C\|m\|_\infty^2,$$

where we have used orthonormality of the wave packets $|I|^{-1/2}w(I\times\omega_I)$. This follows from the pairwise disjointness of the tiles $I\times\omega_i$, which can be seen as follows. Let $P, P' \in T$ be $P = I\times\omega$ and $P' = I'\times\omega'$, and suppose to get a contradiction that $P \ne P'$ and $I\times\omega_i \cap I'\times\omega_i' \ne \emptyset$. Without loss of generality, we might suppose that $I \subset I'$. As $P \ne P'$, we quickly see $I \ne I'$. Hence $|I| < |I'|$ and by area consideration $|\omega_i'| < |\omega_i|$ and $\omega_i' \subset \omega_i$. As both P and P' are in T_j, we also have that $\omega_j' \subset \omega_j$ Hence the distance between ω_i and ω_j is at most the distance between ω_i' and ω_j'. But the ratio of these distances is $|\omega_i|/|\omega_j|$ as all multi-tiles are affine images of each other. Hence this distance has to be zero. Let η then be the common point of the boundary of ω_i and ω_j and ω_i' and ω_j'. If $i < j$, this point is the left endpoint of ω_j and ω_j' and thus of the form $(l + j/2^L)|\omega|$ for some integer l and of the form $(l' + j/2^L)|\omega'|$ for some possibly different integer l'. Dividing by $|\omega'|$ and multiplying by 2^L gives for some positive M

$$2^{L+M}l + 2^M j = 2^L l' + j$$

This is however impossible because the left hand side is divisibly by a larger power of 2 than the right hand side. A similar argument holds of $j < i$. This is the desired contradiction which proves pairwise orthogonality of the tiles $I\times\omega_i$.

We turn to the weak type endpoint at $p = 2$. We need to prove for all $\lambda > 0$

$$\inf\{\mu(A) : L^\infty S_i(m1_{A^c}) > \lambda\} \le C\lambda^{-2}\|m\|_2^2$$

This will complete the proof of the theorem by Marcinkiewicz interpolation.

Thus we need to construct a collection $\mathcal{E}' \subset \mathcal{E}$ such that

$$\sum_{T\in\mathcal{E}'} |I_T| \le C\frac{\|m\|_2^2}{\lambda^2},$$

and for every $T \in \mathcal{E}$

$$S_i(F1_{(\cup\mathcal{E}')^c})(T) \le C\lambda.$$

We construct $\mathcal{E}'$ as union of E_i and $\bigcup_{j\neq i}\mathcal{E}_j$. First construct $\mathcal{E}_i$. We select a finite sequence of multi-tiles P_n, $n = 0, 1, \ldots$. Assume we have selected P_l for $l < n$, this includes the possibility $n = 0$ and we have not selected any trees yet. If it exists, pick $P_n = I_n \times \omega_n$ such that $|F(P_n)| \ge \frac{\lambda}{2^L 10}$ and P_n is disjoint from P_l with $l < n$. If no such P_n exists, we stop the selction. If such P_n exists, we pick one such that I_n is maximal. Because of finiteness of $\mathcal{P}$, the process will eventually ends for some n.

The rectangles P_l with $l \le n$ are pairwise disjoint. We have by orthogonality

$$\sum_{l=0}^{n} |I_l| \le 2^{2L} 100\lambda^{-2} \sum_{l=0}^{n} |I_l||F(P_l)|^2 \le 2^{2L} 100\lambda^{-2}\|m\|_2^2.$$

Not that this is a much stronger use of orthogonality relation than within a single tree as in the case $p = \infty$, because we have a quite arbitrary collection of disjoint tiles here. For $l \le n$ define T_l by letting I_{T_l} be equal I_l and letting ξ_{T_l} be any integer in ω_l. In particular, $P_l \in T_l$. Let $\mathcal{E}_i$ be the collection of trees T_l with $l \le n$. Let A be the union of $\mathcal{E}_i$. We claim that for $P \notin A$ we have

$$F(P) \le \frac{\lambda}{2^L 10}.$$

Assume to get a contradiction that $F(P) > \frac{\lambda}{2^L 10}$. As P was not selected after P_n, it has to intersect some P_l with $l \le n$. Let l be the smallest such index. As P was not selected in place of P_l, we have $|I| \le |I_l|$. Hence $I \subset I_l$, $\omega_l \subset \omega$, therefore $P \in T_l$. This gives a contradiction to $P \notin A$.

The collection $\mathcal{E}_i$ takes care of the L^∞ part in S_i. Now we want to take care of the L^2 part for $j \neq i$ by selecting a collection $\mathcal{E}_j$. First assume $j < i$. We choose anew collection of trees T_l. Assume we have already picked T_l for $l < n$.

If exists, pick T_n such that

$$\frac{1}{|I_{T_n}|} \sum_{P \in T_n \setminus A, P \notin \bigcup_{l<n} T_l} |I||F(P)|^2 \ge \lambda^2.$$

If no such tree exists, we stop the selcetion. If several such trees exists, pick one such that ξ_{T_n} is an integer and maximal possible. (We can pick an integer as integer intervals are the finest resolution on the vertical axis of the phase plane).

As the collection $\mathcal{P}$ is finite, the selection procedure will terminate. Let $\mathcal{E}_j$ be the colelction of selected trees and let A_j be the union of these trees. As the selection porcess stopped, we have for every tree T

$$|I_T|^{-1} \sum_{P \in T \setminus (A \cup A_j)} |I||F(P)|^2 \le \lambda^2.$$

This will give the desired bound on the second part of the size S_i for given j. It remains to see that the sum of I_{T_l} of the selected trees in $\mathcal{E}_j$ satisfies a suitable bound.
Define the reduced tree

$$\mathcal{R}_n = T_n \setminus (A \cup \bigcup_{l<n} T_l)$$

and the set $\mathcal{M}$ of minimal layers in $\mathcal{R}_n$ in the sense that for $P \in \mathcal{M}_n$ and all $P' \in \mathcal{R}_n$ we have that $I' \cap I = \emptyset$ implies $2^L|I'| \geq |I|$. Then the intervals I for all $P \in \mathcal{M}_n$ have at most 2^L fold overlap and are contained in I_{T_n}, and we have since $\mathcal{M}_n$ is disjoint from A,

$$|I_{T_n}|^{-1} \sum_{P \in \mathcal{M}_n} |I||F(P)|^2 \leq \frac{\lambda}{2^{2L}100} |I_{T_n}|^{-1} \sum_{P \in \mathcal{M}_n} |I| \leq 2^L \frac{\lambda^2}{2^{2L}100}$$

As a consequence, by choice of T_n,

$$|I_{T_n}|^{-1} \sum_{P \in \mathcal{R}_n \setminus \mathcal{M}_n} |I||F(P)|^2 \geq \frac{99\lambda^2}{100}$$

We claim that if

$$P \in \mathcal{R}_n \setminus \mathcal{M}_n$$
$$P' \in \mathcal{R}_{n'} \setminus \mathcal{M}_{n'},$$

with $P \neq P'$, then

$$I \times \omega_i \cap I' \times \omega_i' = \emptyset.$$

In fact, suppose not. Since $I \times \omega_i \neq I' \times \omega_i'$, without loss of generality we can assume $|I| < |I'|$ and $I \subset I'$, hence $\omega_i' \subset \omega_i$, yielding by arguments as above $\omega_j' \cap \omega_j = \emptyset$. We observe that ω_j is below ω_j' since $j < i$ and the distance between ω_i and ω_j is larger than the distance between ω_i' and ω_j'. Hence n' is smaller than n by maximality of the top frequency in the selection process. There is $P'' \in \mathcal{M}_n$ such that $I'' \subset I$ and $2^L|I''| \leq |I|$. Hence $\xi_{T_n} \in \omega_j \subset \omega_j''$ and thus $\omega_i \subset \omega_j''$, implying $\omega_i' \subset \omega_j''$, and $\omega_j' \subset \omega_j''$, and finally $\xi_{T_{n'}} \in \omega_j''$. But we also have $I'' \subset I \subset I'$. Hence $P'' \in T_{n'}$, which yields a contradiction to $P'' \in \mathcal{R}_n$. This proves the above claim, and we have by orthogonality

$$\frac{99}{100}\lambda^2 \sum_n |I_{T_n}| \leq \sum_n \sum_{\mathcal{R}_n \setminus \mathcal{M}_n} |I|F(P)^2 \leq \|m\|_2^2,$$

therefore

$$\sum_n |I_{T_n}| \leq C \frac{\|m\|_2^2}{\lambda^2}.$$

This is the desired estimate. In case $j > i$, we use an analogous argument, but we choose ξ_{T_n} to be a minimal integer. This completes the proof of the weak type estimate and thus the proof of the theorem. □

18. The maximal Walsh embedding

We continue to look at the set $\mathcal{P}$ of 2^L-multitiles. Each sich tile is decomposed int 2^L tiles of the form $I \times \omega_i$ counted from bottom to top. For each such $0 \le i \le 2^{L-1}$ assume we are given a real Radon measure m_i on $[0,1)$. Define $F_i : \mathcal{P} \to \mathbb{R}$ by

$$F_i(P) = |I|^{-1} m_i(w(I \times \omega_i)).$$

For a subset $\mathcal{I}$ of cardinality at least 3 of these indices, define the quartile form

$$\sum_{P \in \mathcal{P}} |I| \prod_{i \in \mathcal{I}} F_i(P).$$

By observations from the last lecture we have the estimate

$$(3) \qquad |\sum_{P \in \mathcal{P}} |I| \prod_{i \in \mathcal{I}} F_i(P)|$$

$$\le C L^1 \ell^1 \prod_{i \in \mathcal{I}} F_i \le C \prod_{i \in \mathcal{I}} L^{p_i} S_i F_i \le C \prod_{i \in \mathcal{I}} \|m_i\|_{p_i}$$

Here $1 < p_i \le \infty$ and $\sum_{i \in \mathcal{I}} 1/p_i = 1$, and

$$S_i(F)(T) = \sup_{P \in T} |F(P)| + \sum_{j \neq i} \left(\frac{1}{|I_T|} \sum_{P \in T_j} |I| |F(P)|^2 \right)^{1/2}.$$

with T_j the set of all tiles in T with $\xi_T \in \omega_j$. The crucial estimate in the line of estimates beginning with (3) is the last inequality, which uses the Walsh embedding theorem of last lecture.

Now recall the Walsh Carleson operator for any function $N \in S_K^{\Delta}$ defined by

$$S_N^W(x) m(x) = \sum_{n=0}^{N(x)-1} m(W_n) W_n(x)$$

$$= \sum_{P \in \P} |I|^{-1} m(w(I \times \omega_0)) w(I \times \omega_0)(x) 1_{N(x) \in \omega_1}$$

where $\mathcal{P}$ is the set of bitiles, that is $L = 1$. Pairing with another measure m' gives

$$m'(S_N^W m) = \sum_{P \in \mathcal{P}} |I| F_0(m) G(m')$$

with

$$G(P) = |I|^{-1} m'(w(I \times \omega_0) 1_{N(.) \in \omega_1})$$

To estimaet this expression, the previous sizes S_i and embedding theorems are only useful for the first factor. To approahc the second factor, define the size

$$S_0^*(G)(T) = \frac{1}{|I_T|} \sum_{P \in T_0} |G(P)| + \left(\frac{1}{|I_T|} \sum_{P \in T_1} |I| |G(P)|^2 \right)^{1/2}.$$

It is somewhat dual to the size S_0, in the sense that we have for every tree T

$$|\sum_{P \in T} |I| F_0(P) G(P)| \le |I_T| S_0 F_0(T) S_0^* G(T)$$

EWith this local duality, one can proceed to prove an outer Hölder type inequality and obtain for $2 < p < \infty$

$$\sum_{P\in\mathcal{P}} |I|F_0(m)G(m') \le C(L^p S_0 F_0)(L^{p'} S_0^* G)$$

To estimate this furtshr, we use both the embedding theorem below and the embedding theorem form the last lecture to obtain

$$|m'(S_N^W m)| \le C\|m\|_p \|m'\|_{p'}.$$

By previosu discussions, this bound for the Walsh Carleson operator will give almost everywhere convergence of Walsh Fourier series of Radon measures in L^p.
The crucial maximal Walsh embedding theorem of this lecture is

Theorem 79. *For $1 < p \le \infty$ there exists a constant C such that the following holds. Let m be a measure and let $N \in S_K^\Delta$ and define for $P = I \times \omega$*

$$G(P) = |I|^{-1} m(1_{N(.)\in\omega_1} w(I \times \omega_0)).$$

Then

$$L^p S_0^* G \le C\|m\|_p.$$

Proof. By Marcinkiewicz interpolation, it suffices to show this estimate for $p = \infty$ and a weak endpoint estimate at $p = 1$. We begin with $p = \infty$. We need to show that, for all $T \in \mathcal{E}$

$$S_0^* G(T) \le C\|m\|_\infty.$$

For the first summand in the definition of S_0^*, we recall from the proof of the previous embedding theorem that all the rectangles $I \times \omega_1$ with $P \in T_0$ are pairwise disjoint. For a bitile P define

$$E_P = \{x \in I : n(x) \in \omega_1\}.$$

Then clearly the sets E_P are pairwise disjoint as $P \in T_0$ and contained in I_T. Hence

$$\sum_{P\in T_0} |E_P| \le |I_T|.$$

On the other hand

$$\frac{1}{|I_T|}\sum_{P\in T_0} |G(P)| \le \frac{1}{|I_T|}\sum_{P\in T_0} |E_P|\|m\|_\infty \le \|m\|_\infty.$$

This proves the desired estimate for the first summand in the definition of S_0^*.
To estimate the second summand in the definition of S_0^*, we compute

$$\sum_{P\in T_1} |I|G(P)^2$$

$$= \sum_{P\in T_1} G(P) m(1_{N(.)\in\omega_1} w(I \times \omega_0))$$

$$= m(\sum_{P\in T_1} G(P) 1_{N(.)\in\omega_1} w(I \times \omega_0))$$

$$\leq \|m 1_{I_T}\|_2 \| \sum_{P \in T_1} G(P) 1_{N(x) \in \omega_1} w(I \times \omega_0)\|_2$$

Fix $x \in I_T$ and let Ω be the nested collection of dyadic intervals ω_1 which contain ξ_T . Let ω_1^x be the largest dyadic interval in Ω which does not contain $N(x)$ and let ω^x be its parent. Then $\omega_0 \subset \omega_x$ whenever $N(x) \notin \omega_1$ and $\omega_0 \cap \omega_x = \emptyset$ whenever $N(x) \in \omega_1$. Hence

$$\sum_{P \in T_1} G(P) 1_{N(x) \in \omega_1} w(I \times \omega_0)(x)$$

$$= (1 - P_x) \sum_{P \in T_1} G(P) w(I \times \omega_0)(x)$$

with the projection operator

$$P_x m = \sum_{|I||\omega_1^x|=1} |I|^{-1} m(w(I \times \omega_1^x)) \omega_1^x$$

This is a projection operator since the tiles of the form $I \times \omega_1^x$ are all pairwise disjoint. The projection operator acts like the identity on all wave packets associated with tiles $I \times \omega_0$ with $\xi_T \in \omega_1$ and $N(x) \notin \omega_1$ because the union of tiles defining the projection operator convers such tiles, and it acts liek the zero operator on tiles associated with tiles $I \times \omega_0$ with $\xi_T \in \omega_1$ and $N(x) \notin \omega_1$, because the union of tiles of the projection operator is disjoint form such tiles.
However, $P_x f(x)$ is bounded by the martingale average at size $|\omega_1^x|^{-1}$ of $|f|$ and thus it is bounded by the Hardy Littlewood maximal operator M of $|f|$. By the L^2 bound on the Hardy Littlewood maximal operator we have

$$\| \sum_{P \in T_1} G(P) 1_{N(.) \in \omega_1} w(I \times \omega_0)\|_2$$

$$\leq 2 \|M(\sum_{P \in T_1} G(P) w(I \times \omega_0))\|_2$$

$$\leq C \| \sum_{P \in T_1} G(P) w(I \times \omega_0)\|_2 \leq C \left(\sum_{P \in T_1} |I| G(P)^2 \right)^{1/2}$$

In the last inequality we have used that the tiles $I \times \omega_0$ are pairwise disjoint for $P \in T_1$. Hence

$$\frac{1}{|I_T|} \sum_{P \in T_1} |I| G(P)^2 \leq C \|m 1_{I_T}\|_2 \left(\sum_{P \in T_1} |I| G(P)^2 \right)^{1/2}$$

and hence

$$\left(\sum \sum_{P \in T_1} |I| G(P)^2 \right)^{1/2} \leq C |I_T|^{1/2} \|m\|_\infty$$

This proves the desired bound on the second summand in the definition of S_0^* and completes the proof of the case $p = \infty$.

We turn to the weak type bound at $p = 1$. For given $\lambda > 0$ we need to construct a collection $\mathcal{E}'$ of trees such that

$$\sum_{T\in\mathcal{E}'} |I_T| \le C\lambda^{-1}\|m\|_1 \tag{4}$$

and for all $T \in \mathcal{E}'$

$$S_0^*(G1_{(\bigcup\mathcal{E}')^c})(T) \le C\lambda. \tag{5}$$

Define the auxiliary function

$$G'(P) = |I|^{-1}\|m1_I 1_{n(.)\in\omega}\|_1$$

We collect a sequence of bitiles similarly to the previous embedding theorem. Assume we have already collected P_l for $l < n$. If it exists, we pick a bitile P_n disjoint from all previously selected P_l such that

$$G'(P_n) \ge \lambda.$$

If no such bitile exists, we stop the collection of tiles. If several such bitiles exists, we choose P_n so that $|I_n|$ is maximal. The construction stops, since there are only finitely many tiles in Ω_K.
Observe that

$$\sum_l |I_l| \le \lambda^{-1}\sum_{l=1}^n \|m1_I 1_{n(.)\in\omega_n}\|_1 \le \lambda^{-1}\|\sum_{l=1}^n m1_I 1_{N(.)\in\omega_n}\|_1 \le \lambda^{-1}\|m\|_1.$$

Here we use in the last inequality that the slected tiles P_l are disjoint and thus for every x, the point $(x, N(x))$ can only be in one of these bitiles.
Define T_n to be the tree with top interval I_n and any integer as top frequency ξ_n so that P_n is in the tree. Let $\mathcal{E}'$ be the collection of selected trees. Then the sum of top intervals of trees in $\mathcal{E}'$ satisfies the desired estimate (4).
To prove (5), we estimate the two summands in the definition of the size S_0^* separately. Let $A = \bigcup\mathcal{E}'$. Then for every tree $T \in \mathcal{E}$ we have

$$\sum_{P\in T_0\setminus A} |I||G(P)| \le \sum_{P\in T_0\setminus A} |m|(1_I 1_{N(.)\in\omega_1})$$

$$\le |m|(\sum_{P\in T_0\setminus A} 1_I 1_{N(.)\in\omega_1})$$

Now let $\mathcal{J}$ be the collection of maximal dyadic intervals contained in I_T such that there exits $P \in T \cap A$ with $P = J \times \omega$ for some ω or such that $|J| = 2^K$ for the finest scale 2^K in our universe. The collection $\mathcal{J}$ partitions I_T and we estimate the last display as

$$\le \sum_{J\in\mathcal{J}} |m|(1_J 1_{N(.)\in\omega_J} \sum_{P\in T_0\setminus A} 1_I 1_{N(.)\in\omega_1}) \le \sum_{J\in\mathcal{J}} |m|(1_J 1_{N(.)\in\omega_J})$$

Here we used that the tiles $I \times \omega_i \in T_0 \setminus A$ are pairwise disjoint for $P \in T_0$ and $\omega_i \subset \omega_J$ if $x \in J$ and $N(x) \in \omega_i$. Now consider the parent J' of J, and consider the bitile $P = J' \times \omega_J$ so that $\xi_T \in \omega_J$ By definition of $\mathcal{J}$, the bitile P is not in A, and by construction of A we have $G(P') \le \lambda$.

Namely, assume $G(P') \geq \lambda$, then by the greedy choice, P' intersects a selected P_l with $J \subset I_l$. But then $J \in A$, a contradiction. hence we estimate the previous display by

$$\leq \sum_{J \in \mathcal{J}} |m|(1_{J'} 1_{N(.) \in \omega_J}) \leq \sum_{J \in \mathcal{J}} \lambda |J'| \leq 2 \sum_{J \in \mathcal{J}} \lambda |J| \leq 2\lambda |I_T|$$

This proves the desired bound for the first summand in the definition of S_0^*.

To estimate the second summand, we compute similarly to before,

$$\sum_{P \in T_1} |I| G(P)^2 = m(\sum_{P \in T_1} G(P) 1_{N(.) \in \omega_1} w(I \times \omega_0))$$

$$= \sum_{J \in \mathcal{J}} m(1_J 1_{N(.) \in \omega_J} \sum_{P \in T_1} G(P) 1_{N(.) \in \omega_1} w(I \times \omega_0))$$

$$\leq \sum_{J \in \mathcal{J}} |m|(1_J 1_{N(.) \in \omega_J} |(1 - P_.)(\sum_{P \in T_1} G(P) w(I \times \omega_0))|)$$

Now we use that $(1 - P_.)(\sum_{P \in T_1} G(P) w(I \times \omega_0))$ has constant modulus on every interval J to estimate the above by

$$\leq \sum_{J \in \mathcal{J}} \lambda |J| \inf_{x \in J} |(1 - P_x)(\sum_{P \in T_1} G(P) w(I \times \omega_0)(x))|)$$

$$\leq \lambda \|(1 - P_.)(\sum_{P \in T_1} G(P) w(I \times \omega_0))\|_1$$

$$\leq 2\lambda |I_T|^{1/2} \|M(\sum_{P \in T_1} G(P) w(I \times \omega_0))\|_2$$

$$\leq 2\lambda |I_T|^{1/2} \|\sum_{P \in T_1} G(P) w(I \times \omega_0)\|_2$$

$$\leq 2\lambda |I_T|^{1/2} (\sum_{P \in T_1} |I| G(P)^2)^2$$

From here we obtain the desired bound on the second summand in the definition of S_0^* as before. This completes the proof of the theorem. $\square$

19. The bilinear Hilbert transform

We introduce the *bilinear Hilbert transform.* For $f, g \in \mathcal{S}(\mathbb{R})$, we define

$$\begin{aligned} B(f,g)(x) :=& \mathrm{pv} \int_{\mathbb{R}} f(x-t)g(x-2t)\frac{dt}{t} \\ =& \lim_{\varepsilon\to 0} \int_{[-\varepsilon,\varepsilon]^c} f(x-t)g(x-2t)\frac{dt}{t} \end{aligned}$$

where in the last passage we gained integrability of the argument by boundedness property of Schwartz functions.
To the transform we associate a trilinear form

$$\begin{aligned} \Lambda(f,g,h) :=& \int_{\mathbb{R}} B(f,g)(x)h(x)\, dx = \\ =& \mathrm{pv} \iint h(x)f(x-t)g(x-2t)\frac{dt}{t}\, dx. \end{aligned}$$

A related integral is given by

$$\int_{\mathbb{R}} \int_0^1 f(x)f(x-t)f(x-2t)\, dt\, dx.$$

If $f = 1_E$ for a set $E \subset \mathbb{R}$ it counts arithmetic progression of length 3 inside E of width at most 1.
More generally, for $\beta(\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3$, we define

$$\Lambda_\beta(f_1, f_2, f_3) = \mathrm{pv} \iint f_1(x-\beta_1 t)f_2(x-\beta_2 t)f_3(x-\beta_3 t)\frac{dt}{t}\, dx.$$

The change of variables $x \mapsto x - \gamma t$ yields

$$\begin{aligned} \Lambda_\beta(f_1, f_2, f_3) =& \mathrm{pv} \iint \prod_{j=1}^3 f_j(x-\beta_j t)\frac{dt}{t}\, dx = \\ =& \mathrm{pv} \iint \prod_{j=1}^3 f_j(x-\gamma t-\beta_j t)\frac{dt}{t}\, dx. \end{aligned}$$

Thus we may add γ to β_j and, without loss of generality, assume $\beta_1 + \beta_2 + \beta_3 = 0$.
The change of variables $t \mapsto \lambda t$ yields

$$\begin{aligned} \Lambda_\beta(f_1, f_2, f_3) =& \mathrm{pv} \iint \prod_{j=1}^3 f_j(x-\beta_j t)\frac{dt}{t}\, dx = \\ =& \mathrm{pv} \iint \prod_{j=1}^3 f_j(x-\beta_j \lambda t)\frac{dt}{t}\, dx. \end{aligned}$$

Thus we may replace β_j by $\lambda\beta_j$ and, without loss of generality, assume $\beta_1^2 + \beta_2^2 + \beta_3^2 = 1$ (unless $\beta_1 = \beta_2 = \beta_3 = 0$, but then $\Lambda = \mathrm{pv}\int_{\mathbb{R}} \frac{dt}{t} = 0$).
Therefore $\beta = (\beta_1, \beta_2, \beta_3)$ is a unit vector perpendicular to the vector $(1,1,1)$. We are down to a 1-parameter family. Moreover,

$$\Lambda_\beta(f_1, f_2, f_3) = \Lambda_{-\beta}(f_1, f_2, f_3),$$

so the parameter belongs to a projective line.

We can't get rid of this parameter dependence. In fact, consider the degenerate cases, i.e. when $\beta_i = \beta_j$ for some $i \neq j$, e.g. $\beta_1 = \beta_2$. The changes of variables described above allow us to assume $\beta_1 = \beta_2 = 0$, $\beta_3 = 1$. Therefore we get

$$\Lambda_{(0,0,1)}(f_1, f_2, f_3) = \text{pv} \iint f_1(x) f_2(x) f_3(x-t) \frac{dt}{t} \, dx = \int f_1 f_2 H f_3,$$

where Hf_3 is the Hilbert transform of f_3. In particular we have the bound

$$\Lambda_{(0,0,1)}(f_1, f_2, f_3) \leq C \|f_1\|_{p_1} \|f_2\|_{p_2} \|f_3\|_{p_3},$$

where $1 < p_1, p_2, p_3 < \infty$, $\frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3} = 1$. The same bound in the non degenerate case can't be proven through a similar simple argument. This should tell us that we can't recover the non degenerate case from the degenerate one.

We look at the symmetries of the trilinear form:

- **Translations.** For $y \in \mathbb{R}$, $T_y f(x) = f(x-y)$. Then
$$\Lambda_\beta(T_y f_1, T_y f_2, T_y f_3) = \Lambda_\beta(f_1, f_2, f_3);$$
- **Dilations.** For $\lambda > 0$, $D_\lambda f(x) = f\left(\frac{x}{\lambda}\right)$. Then
$$\Lambda_\beta(D_\lambda f_1, D_\lambda f_2, D_\lambda f_3) = \lambda \Lambda_\beta(f_1, f_2, f_3);$$
- **Modulations.** For $\eta \in \mathbb{R}$, $M_\eta f(x) = e^{2\pi i \eta x} f(x)$. Then, for $\alpha \in \mathbb{R}^3$,

$$\begin{aligned}\Lambda_\beta(M_{\alpha_1\eta} f_1, M_{\alpha_2\eta} f_2, M_{\alpha_3\eta} f_3) &= \\ = \text{pv} \iint f_1(x-\beta_1 t) f_2(x-\beta_2 t) f_3(x-\beta_3 t) & \\ e^{2\pi i \alpha_1 \eta(x-\beta_1 t) + 2\pi i \alpha_2 \eta(x-\beta_2 t) + 2\pi i \alpha_3 \eta(x-\beta_3 t)} \frac{dt}{t} \, dx &= \Lambda_\beta(f_1, f_2, f_3).\end{aligned}$$

Here the last identity holds provided the vector α is perpenidular to β and $(1,1,1)$. We can define the Hilbert transform of the function f in terms of an integral of $\widehat{f}$ in the following way

$$\text{pv} \int f(x-t) \frac{dt}{t} = c \int \widehat{f}(\eta) \text{sgn}(\eta) \, d\eta.$$

What is the analogous for the Bilinear Hilbert Transform? If $\varphi: \mathbb{R} \to \mathbb{R}$ is an odd Schwartz function such that $\int_0^\infty \varphi(s) \, ds = 1$, then

$$\int_0^\infty \varphi(ts) \, ds = \frac{1}{t} \int_0^\infty \varphi(u) \, du = \frac{1}{t}.$$

By substituting this equality inside the trilinear form we obtain

$$\Lambda_\beta(f, g, h) = \int_0^\infty \left[\iint f(x-\beta_1 t) g(x-\beta_2 t) h(x-\beta_3 t) \varphi(st) \, dt \, dx \right] ds.$$

We want to express the integral in terms of an integral of the Fourier transform of

$$F(y_1, y_2, y_3, y_4) = f(y_1) g(y_2) h(y_3) \varphi(y_4),$$

Let

$$\Gamma = \operatorname{span}\{(1,1,1,0), (-\beta_1, -\beta_2, -\beta_3, s)\},$$

where the vectors are orthogonal and have length $\sqrt{3}$ and $\sqrt{1+s^2}$. We can continue the chain of equality above

$$= c\int_0^\infty \frac{1}{\sqrt{1+s^2}}\left(\iint_\Gamma f\ g\ h\ \varphi\right) ds = c\int_0^\infty \frac{1}{\sqrt{1+s^2}}\left(\iint_{\Gamma^\perp} \widehat{f}\,\widehat{g}\,\widehat{h}\,\widehat{\varphi}\right) ds.$$

In the last equality we used the following result

Claim 80. *Integrating in $F\colon\mathbb{R}^n \to \mathbb{R}$ over a subspace Γ is equivalent to integrating $\widehat{F}\colon\mathbb{R}^n \to \mathbb{R}$ over $\Gamma^\perp$.*

Proof.

$$\widehat{F}(\xi) = \int_{\mathbb{R}^n} F(x)e^{-2\pi i x\cdot\xi}\,dx \qquad \widehat{F}(0) = \int_{\mathbb{R}^n} F(x)\,dx$$

$$F(x) = \int_{\mathbb{R}^n} \widehat{F}(\xi)e^{2\pi i x\cdot\xi}\,d\xi \qquad F(0) = \int_{\mathbb{R}^n} \widehat{F}(\xi)\,d\xi.$$

These are already a first instance of the claim with $\Gamma = \mathbb{R}^n$ and $\Gamma = \{0\}$. For a general subspace Γ, we can assume without loss of generality that Γ is spanned by $x_1, \dots, x_k$, and therefore $\Gamma^\perp$ is spanned by $x_{k+1}, \dots, x_n$. Then

$$\int F(x_1, \dots, x_k, 0, \dots, 0)\,dx_1 \dots dx_n = \widehat{F}^{1,\dots,k}(0,\dots,0) =$$
$$= \int \widehat{F}(0,\dots,0,\xi_{k+1},\dots,\xi_n)\,d\xi_{k+1}\dots d\xi_n,$$

where $\widehat{F}^{1,\dots,k}$ is the Fourier transform only with respect to the first k coordinates. □

In our case, we have

$$\Gamma^\perp = \operatorname{span}\left\{(\alpha_1,\alpha_2,\alpha_3,0), \left(\beta_1,\beta_2,\beta_3,\frac{1}{s}\right)\right\},$$

where $\alpha \perp \beta$, $\alpha \perp (1,1,1)$, $\|\alpha\| = 1$. The two vectors are orthogonal to each other and of length 1 and $\sqrt{1+\frac{1}{s^2}}$. We can continue the chain of equality above

$$= c\int_0^\infty \frac{\sqrt{1+\frac{1}{s^2}}}{\sqrt{1+s^2}} \iint \widehat{f}(\alpha_1\xi+\beta_1\eta)\widehat{g}(\alpha_2\xi+\beta_2\eta)\widehat{h}(\alpha_3\xi+\beta_3\eta)\widehat{\varphi}\left(\frac{1}{s}\eta\right) d\xi\,d\eta\,ds.$$

But

$$\int_0^\infty \frac{1}{s}\widehat{\varphi}\left(\frac{1}{s}\eta\right) ds = \operatorname{sgn}(\eta)\int_0^\infty \widehat{\varphi}(s|\eta|)\frac{ds}{s} =$$
$$= \operatorname{sgn}(\eta)\int_0^\infty \widehat{\varphi}(s)\frac{ds}{s} = \operatorname{sgn}(\eta)\text{const}.$$

Therefore we continue the chain of equalities above

$$= \text{const.} \iint \widehat{f}(\alpha_1\xi + \beta_1\eta)\widehat{g}(\alpha_2\xi + \beta_2\eta)\widehat{h}(\alpha_3\xi + \beta_3\eta)\text{sgn}(\eta)\, d\xi\, d\eta =$$
$$= \text{const.} \iint_{\eta_1+\eta_2+\eta_3=0} \widehat{f}(\eta_1)\widehat{g}(\eta_2)\widehat{h}(\eta_3)\text{sgn}(\eta\cdot\beta)\, d\sigma.$$

In order to prove the wanted bound for the trilinear form we would like to use the Carleson embedding theorem we proved last time. We consider the embedding map into the upper 3-space defined, for $f\colon \mathbb{R} \to \mathbb{R}$, φ Schwartz function, by

$$F(y,\eta,\lambda) = \int f(x)\lambda^{-1}\varphi(\lambda^{-1}(y-x))e^{2\pi i\eta(y-x)}\, dx.$$

In particular, we pick φ such that $\widehat{\varphi}$ has compact support contained in $[-10^{-1}, 10^{-1}]$ and it is nonnegative, and we consider

$$\widehat{\varphi}(\eta_1)\widehat{\varphi}(\eta_2)\widehat{\varphi}(\eta_3).$$

To recover the $\text{sgn}(\eta\cdot\beta)$ we shift the support of the functions $\widehat{\varphi}(\eta_i)$ so that the centre is in β. In particular, for $\eta \in \mathbb{R}^3$ such that

$$\widehat{\varphi}(\eta_1-\beta_1)\widehat{\varphi}(\eta_2-\beta_2)\widehat{\varphi}(\eta_3-\beta_3) > 0,$$

we have $\text{sgn}(\eta\cdot\beta)$. To make value independent on the vector α we integrate the product with variables η_i translated by $s\alpha_i$, obtaining

$$\int \prod_{j=1}^{3} \widehat{\varphi}(\eta_j - \beta_j - s\alpha_j)\, ds.$$

To make value independent on the dilations by factor λ we integrate the product with variables η_i dilated by a factor λ, obtaining

$$\int \prod_{j=1}^{3} \widehat{\varphi}(\lambda\eta_j - \beta_j - s\alpha_j)\, ds \frac{d\lambda}{\lambda} = c\text{sgn}(\eta\cdot\beta),$$

where c is a constant.

As a consequence we can rewrite

$$\widetilde{\Lambda}_\beta(f_1,f_2,f_3) = \iint\limits_{\substack{\eta_1+\eta_2+\eta_3=0\\ \mathrm{sgn}(\eta\cdot\beta)}} \prod_{j=1}^{3} \widehat{f}_j(\eta_j)\mathrm{sgn}(\eta\cdot\beta)\,d\sigma =$$

$$= \int_0^\infty \int_{\mathbb{R}} \iint\limits_{\eta_1+\eta_2+\eta_3=0} \prod_{j=1}^{3} \widehat{f}_j(\eta_j)\widehat{\varphi}(\lambda\eta_j-\beta_j-s\alpha_j)\,d\sigma\,ds\frac{d\lambda}{\lambda} =$$

$$\overset{\text{FT trick}}{=} \int_0^\infty \int_{\mathbb{R}}\int_{\mathbb{R}} \Bigg[\int_{\mathbb{R}^3} \prod_{j=1}^{3} f_j(\eta_j)e^{-2\pi i y\eta_j}$$

$$\widehat{\varphi}(\lambda\eta_j-\beta_j-s\alpha_j)\,d\eta_1\,d\eta_2\,d\eta_3 \Bigg] dy\,ds\frac{d\lambda}{\lambda} =$$

$$= \cdots = \int_0^\infty \int_{\mathbb{R}^2} \prod_{j=1}^{3} F_j(y,\alpha_j s+\beta_j\lambda^{-1},\lambda)\,ds\,dy\,d\lambda.$$

Here in the last few steps we have taken anouthe Foruier transform. The last expression resembles the quartile operator, there is a product of three embedded functions F_i related to wavepackets that have a mutual shift similra to the different tiles within a multi-tile
The following theorem is a continuousanalogue of boudnedness of the quartile operator.

Theorem 81. *For* $2 < p_i < \infty$, $\sum \frac{1}{p_i} = 1$, *there exists* $C_{\beta,p}$ *such that*

$$\Lambda_\beta(f_1,f_2,f_3) \le C_{\beta,p} \prod_{j=1}^{3} \|f_j\|_{p_j}.$$

Note that a priori the constant depends on β, and the dependence make it blow up in a nonintegrable way near the degenerate cases. However, the bound is known to holds also in the degenerate case with a finite constant. This suggests that the estimate given by the theorem above is not optimal, in particular that an uniform bound with a constant independent on β can be proven. This has been done with the same conditions of the statement of the theorem.
The problem in the degenerate case is that, as in the proof of the boundedness of the Carleson operator, we need the translations of a tile to be disjoint. In the degenerate case this fails, with two translated copies overlapping. By taking a tile smaller inverse proportionally to the distance of these translations we can recover the necessary disjointness property. However, in this way the constant blows up morally like the inverse of the distance between these pieces, thus in a nonintegrable way near the degenerate cases.
We conclude the lecture describing an example of an application for the BHT bound, which is historically one of the starting point of the study of the bilinear Hilbert transform.

Consider the Cauchy integral over a Lipschitz curve $y \mapsto y + iA(y)$ given by

$$\int f(x) \frac{1}{y - x + i(A(y) - A(x))} \, dx = \int f(x) \frac{1}{y-x} \left(\frac{1}{1 + i\frac{A(y)-A(x)}{y-x}} \right) dx =$$

$$\overset{\text{Taylor}}{=} c \int f(x) \frac{1}{y-x} \frac{A(y) - A(x)}{y-x} \, dx.$$

This is the so called *Calderon commutator* $\left[* \frac{1}{t^2}, A \right] f$. By expanding the last fraction to an integral we obtain

$$\frac{A(y) - A(x)}{y - x} = \int_0^1 A'(x + (y-x)\alpha) \, d\alpha.$$

By substituting it in the integral above we get

$$\int_0^1 \left[\int f(x) \frac{1}{y-x} A'(x + (y-x)\alpha) \, dx \right] d\alpha,$$

where we recognize the bilinear Hilbert transform (in this case we artificially introduced the parameter α). Therefore the inner integral can be bounded by $C_{\alpha,p} \|f\|_p$. In order to conclude that the Cauchy integral over the Lipschitz curve is bounded by a norm of f, we need $C_{\alpha,p}$ to be integrable near $\alpha = 0$, which corresponds to the degenerate case for the trilinear form. A uniform bound for the bilinear Hilbert transform, i.e. if $C_{\alpha,p} = C_p$ was independent of α, would do the work, but even a weaker result is enough.