Rev. Mat. Iberoamericana **24** (2008), no. 3, 721–744

The Walsh model for M_2^* Carleson

Ciprian Demeter, Michael Lacey, Terence Tao and Christoph Thiele

Abstract

We study the Walsh model of a certain maximal truncation of Carleson's operator related to the Return Times Theorem.

1. Introduction

Let $\mathcal{D}$ denote the collection of all the dyadic intervals of the form $[2^i m, 2^i(m+1))$, $i, m \in \mathbb{Z}$ and let $C^{\mathcal{D}}(\mathbb{R}_+)$ be the set of all the functions $f : \mathbb{R}_+ \to \mathbb{R}$ that are finite linear combinations of characteristic functions of dyadic intervals.

For $l \geq 0$ we recall that the $l-$th Walsh function W_l is defined recursively by the formula

$$W_0 = 1_{[0,1)}$$

$$W_{2l} = W_l(2x) + W_l(2x-1)$$

$$W_{2l+1} = W_l(2x) - W_l(2x-1).$$

We recognize that W_1 is the Haar function also denoted by h.

Definition 1.1 *A tile P is a rectangle $I_P \times \omega_P$ of area one, such that I_P and ω_P are dyadic intervals. If $P = [2^i n, 2^i(n+1)) \times [2^{-i} l, 2^{-i}(l+1))$ is such a tile, we define the corresponding Walsh wave packet w_P by*

$$w_P(x) = 2^{-i/2} W_l(2^{-i}x - n).$$

The intervals I_P and w_P will be referred to as the time and frequency intervals of the tile P.

2000 Mathematics Subject Classification: 42B25, 42B20.
Keywords: Carleson's operator, multiplier norm.

Definition 1.2 *A bitile P is a rectangle $I_P \times \omega_P$ of area two, such that I_P and ω_P are dyadic intervals. For any bitile*

$$P = [2^i n, 2^i(n+1)) \times [2^{-i+1}l, 2^{-i+1}(l+1))$$

we define the lower tile

$$P_1 = [2^i n, 2^i(n+1)) \times [2l2^{-i}, (2l+1)2^{-i})$$

and the upper tile

$$P_2 = [2^i n, 2^i(n+1)) \times [(2l+1)2^{-i}, (2l+2)2^{-i}).$$

If ω_P is the frequency interval of the bitile P then we will use the notations $\omega_{P,1}$ and $\omega_{P,2}$ for the the frequency intervals of the sub-tiles P_1 and P_2.

We next recall the definition of the Walsh-Fourier transform. Except on a set of measure 0 (which we shall always ignore), every $x \in \mathbb{R}_+$ can be identified with a doubly-infinite set of binary digits

$$x = ...a_2 a_1 a_0.a_{-1}a_{-2}...$$

where $a_n \in \mathbb{Z}_2$ and a_n is eventually zero as $n \to \infty$. We define two operations on $\mathbb{R}_+$ by

$$a_n(x \oplus y) := a_n(x) + a_n(y)$$
$$a_n(x \otimes y) := \sum_{m \in \mathbb{Z}} a_m(x) a_{n-m}(y),$$

where the addition and multiplication in the right hand terms are considered modulo 2. We next define the function $e : \mathbb{R}_+ \to \{-1, 1\}$ to be 1 when $a_{-1} = 0$ and -1 when $a_{-1} = 1$. Using this we can introduce the Walsh-Fourier transform of a function $f \in C^{\mathcal{D}}(\mathbb{R}_+)$ to be

$$\widehat{f}(\xi) := \int e(x \otimes \xi) f(x) dx.$$

We also note that the inverse Walsh-Fourier transform $f^{\check{}}$ and the Walsh-Fourier transform coincide in this context.

In the following we will denote with $\mathbf{S}^{\text{univ}}$ the collection of all the bitiles. It is known, see [6], that the almost everywhere convergence of the Walsh series for $f \in L^p$

$$\sum_{l \geq 0} \langle f, W_l \rangle W_l(x)$$

is a consequence of the estimate

$$\|\mathbf{W} f\|_p \lesssim \|f\|_p,$$

where

$$\mathbf{W} f(x) = \| \sum_{P \in \mathbf{S}^{\text{univ}}} \langle f, w_{P_1} \rangle w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \|_{L^\infty_\theta}.$$

Define the M_2^* norm of a family of Walsh multipliers m_k as

$$\|(m_k(\theta))_{k\in\mathbb{Z}}\|_{M_2^*(\theta)} = \sup_{\|g\|_2=1} \|\sup_k |(\widehat{g}m_k)\check{}(x)|\|_{L^2_x}.$$

In this paper we will be concerned with getting estimates for the operator

$$\mathbf{W}^{\max} f(x) = \|(\sum_{P\in\mathbf{S}^{\mathrm{univ}}:|I_P|<2^k} \langle f, w_{P_1}\rangle\, w_{P_1}(x) 1_{\omega_{P,2}}(\theta))_{k\in\mathbb{Z}}\|_{M_2^*(\theta)}.$$

Theorem 1.3 *For each* $1 < p < \infty$ *we have*

$$\|\mathbf{W}^{\max} f\|_p \lesssim_p \|f\|_p. \tag{1.1}$$

It has been acknowledged, see for example [4], [7], that the Walsh models provide a lot of the intuition that lies behind their Fourier analog. In our case, the interest in proving Theorem 1.3 is motivated by its connections with the following Return Times Theorem due to Bourgain [2].

Theorem 1.4 *Let* $\mathbf{X} = (X, \Sigma, \mu, \tau)$ *be a dynamical system and let* $1 \le p, q \le \infty$ *satisfy* $\frac{1}{p} + \frac{1}{q} \le 1$. *For each function* $f \in L^p(X)$ *there is a universal set* $X_0 \subseteq X$ *with* $\mu(X_0) = 1$, *such that for each second dynamical system* $\mathbf{Y} = (Y, \mathcal{F}, \nu, \sigma)$, *each* $g \in L^q(Y)$ *and each* $x \in X_0$, *the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} f(\tau^n x) g(\sigma^n y)$$

converge ν*- almost everywhere.*

In [3] we extend Bourgain's theorem to a larger range of p and q. Our argument there relies on estimates like the one in Theorem 1.4 for a model operator which is the Fourier counterpart of $\mathbf{W}^{\max}$. We hope that our presentation here for the simpler Walsh model will ease the understanding of the the proof in [3].

We note that in order to prove Theorem 1.3 it suffices to assume that the summation in the definition of the operator $\mathbf{W}^{\max}$ runs over a finite collection $\mathbf{S} \subset \mathbf{S}^{\mathrm{univ}}$ of bitiles, and to prove inequality (1.1) with bounds independent on $\mathbf{S}$. We fix the collection $\mathbf{S}$ for the remaining part of the paper.

The argument relies on first splitting the collection of bitiles into structured collections called trees. The bitiles in each tree give rise to a modulated Littlewood-Paley decomposition. The model operator $\mathbf{W}^{\max}$ restricted to each such a tree is estimated in Section 3, by using the Caldéron-Zygmund-type estimates from Section 2.

In Section 5 the operator $\mathbf{W}^{\max} f(x)$ is estimated pointwise, and it is shown that for each x the contribution to $\mathbf{W}^{\max} f(x)$ comes from one stack of trees. Crucial to estimating this contribution is a weighted version of a maximal multiplier result due to Bourgain. This is proved in Section 4. The different pieces of the proof are put together in the last section of this paper.

2. Variational norm estimates for averages

Let H be a separable Hilbert space equipped with a norm $|\cdot|_H$ and denote by $L^q(\mathbb{R}, H)$ the measurable functions on $\mathbb{R}$ with values in H whose q-th power are integrable. Let $\mathbb{E}(f|\mathcal{D}_k)$ denote the conditional expectation with respect to the σ-algebra on $\mathbb{R}$ generated by the dyadic intervals of length 2^k. We include the case $k = \infty$ by setting $\mathbb{E}(f|\mathcal{D}_\infty) = 0$. From now on we will use the notation

$$g_I(x) = \frac{1}{|I|^{1/2}} g\left(\frac{x - l(I)}{|I|}\right)$$

for each dyadic interval $I = [l(I), r(I))$.

Lemma 2.1 (Jump inequality) *Consider* $1 < q < \infty$ *and* $f \in L^q(\mathbb{R}, H)$. *For each* x *and* $\lambda > 0$ *define the entropy number* $M_\lambda(x)$ *be the maximal length of a chain* $\infty = k_0 > k_1 > k_2 > \cdots > k_{M_\lambda(x)}$ *such that for each* $1 \le m \le M_\lambda(x)$

$$|\mathbb{E}(f|\mathcal{D}_{k_m})(x) - \mathbb{E}(f|\mathcal{D}_{k_{m-1}})(x)|_H \ge \lambda.$$

Then

$$\|\lambda M_\lambda^{1/2}(x)\|_{L^q_x(\mathbb{R},H)} \le C_q \|f\|_{L^q(\mathbb{R},H)}$$

where the constant C_q *remains bounded for* q *in any compact subinterval of* $(1, \infty)$.

Proof. This result is well known, we briefly sketch the proof for completeness. First we establish that the number $M_\lambda(x)$ of λ-jumps can be estimated by counting the $\lambda/2$-jumps in a greedy algorithmic way. Let $k_0(x) = \infty$ and for $m \ge 1$ let $k_m(x)$ be the minimal number, if it exists, such that $|\mathbb{E}(f|\mathcal{D}_{k_m(x)})(x) - \mathbb{E}(f|\mathcal{D}_{k_{m-1}(x)})(x)|_H \ge \lambda/2$. Let $\tilde{M}_\lambda(x)$ be the maximal index for which $k_{\tilde{M}_\lambda(x)}$ exists. Define

$$A_x = \{k_0(x), k_1(x), \ldots, k_{\tilde{M}_\lambda(x)}\}$$
$$\mathcal{I}_x = \{J \in \mathcal{D} : x \in J,\ |J| = 2^k \text{ for some } k \in A_x\}.$$

Then one easily checks that $M_\lambda(x) \le \tilde{M}_\lambda(x)$. The crucial additional property of this greedy selection is that the initial parts of the sequence k_m coincide for two nearby values of x until the value of 2^{k_m} gets smaller than the length of the smallest dyadic interval containing both values.

For each x and each selected interval $J \in \mathcal{I}_x$, let $\mathcal{I}_J$ be the collection of dyadic intervals contained in J but not contained in any interval from $\mathcal{I}_x$ of length smaller than $|J|$. By vector valued Caldéron-Zygmund theory we have

$$\Big\| \Big(\sum_{J \in \mathcal{I}_x} \Big| \sum_{I \in \mathcal{I}_J} \epsilon_I \langle f, h_I \rangle h_I(x) \Big|_H^2 \Big)^{1/2} \Big\|_{L^q(\mathbb{R})} \le C_q \|f\|_{L^q(\mathbb{R},H)} \tag{2.1}$$

uniformly in all choices of signs $\epsilon_I \in \{-1, 1\}$. For $q = 2$ this is an easy Hilbert space argument using orthogonality of the functions h_J. For $q < 2$ we use a Caldéron-Zygmund decomposition of $|f|$ to obtain a weak endpoint at $q = 1$ and then interpolate. For $q > 2$ we use BMO techniques, i.e., we estimate the sharp maximal function

$$g^{\#}(x) = \sup_{x \in I} \Big(\frac{1}{|I|} \int_I (g - g_I)^2 \Big)^{1/2} = \sup_{x \in I} \Big(\frac{1}{|I|} \int_I g^2 - g_I^2 \Big)^{1/2}$$

of the function g on the left hand side of (2.1) by the maximal function of $|f|$, and then use standard L^q bounds for the sharp function and the maximal function.

Inequality (2.1) implies

$$\Big\| \Big(\sum_{1 \le m \le \tilde{M}_\lambda(x)} |\mathbb{E}(f|\mathcal{D}_{k_m(x)})(x) - \mathbb{E}(f|\mathcal{D}_{k_{m-1}(x)})(x)|_H^2 \Big)^{1/2} \Big\|_q \le C_q \|f\|_{L^q(\mathbb{R},H)}$$

and using that all jumps are at least $\lambda/2$ proves the lemma. ∎

Define the r- variational norm of a sequence g_k of elements in H to be

$$\|g_k\|_{V^r(k)} := \sup_k |g_k|_H + \sup_{M, k_0, k_1, \dots, k_M} \Big(\sum_{m=1}^{M} |g_{k_m} - g_{k_{m-1}}|_H^r \Big)^{1/r}$$

One may also define some "weak" variational norm

$$\|g_k\|_{V^{r,\infty}(k)} := \sup_k |g_k|_H + \sup_{\lambda > 0} \lambda M_\lambda^{1/r}.$$

where M_λ is the maximal number of indices $k_0, k_1, \dots, k_M$ such that $|g_{k_m} - g_{k_{m-1}}|_H \ge \lambda$ for all $1 \le m \le M$. We have the usual estimate for the $V^r(k)$ norm in terms of M_λ

$$\|g_k\|_{V^r(k)} \le \|g_k\|_\infty + C \int_0^\infty \lambda^r M_\lambda \frac{d\lambda}{\lambda}$$

The jump inequality in Lemma 2.1 is almost a $V^{2,\infty}$ inequality, with the difference that in that inequality λ is independent of x, while in an honest

$V^{2,\infty}$ inequality the parameter λ may be maximized at every x individually. Hence the jump inequality is somewhat weaker than a $V^{2,\infty}$ inequality. By integrating over all λ and using Fubini one can abandon this disadvantage of λ being constant in x and prove honest $V^r(k)$ norm estimates with $r > 2$.

Lemma 2.2 (Variational estimate) *Let $1 < q < \infty$ and $f \in L^q(\mathbb{R}, H)$. Then for $2 < r < \infty$ we have*

$$\Big\| \|\mathbb{E}(f|\mathcal{D}_k)(x)\|_{V^r(k)} \Big\|_{L^q_x} \le C_q(1 + (r-2)^{-1})\|f\|_{L^q(\mathbb{R},H)}$$

where C_q remains bounded on any compact interval of $(1, \infty)$.

Proof. For each x and $\lambda > 0$ we denote by $M_\lambda(x)$ the entropy number of the collection $\{\mathbb{E}(f|\mathcal{D}_k)(x) : k \in \mathbb{Z}\}$. We first consider this inequality for $|f|$ being the characteristic function of a set A. Then $M_\lambda = 0$ for $\lambda > 1$. Hence we can write for $2 < r < \infty$

$$\|\mathbb{E}(f|\mathcal{D}_k)(x)\|_{V^r(k)} \le C \left(\int_0^1 \lambda^2 M_\lambda(x) \ \lambda^{r-2} \frac{d\lambda}{\lambda} \right)^{1/r}$$

The right hand term is an $L^r_\lambda(d\mu)$ norm of $(\lambda^2 M_\lambda)^{1/r}(x)$ with respect to an appropriate measure space of total mass $\|\mu\| = \int_0^1 \lambda^{r-3} d\lambda = (r-2)^{-1}$.

In the case $q = r$ we get

$$\begin{aligned} \|\|\mathbb{E}(f|\mathcal{D}_k)(x)\|_{V^r(k)}\|_{L^r_x} &\le C\|\|(\lambda^2 M_\lambda(x))^{1/r}\|_{L^r_\lambda(d\mu)}\|_{L^r_x} \\ &= C\|\|(\lambda^2 M_\lambda(x))^{1/r}\|_{L^r_x}\|_{L^r_\lambda(d\mu)} \\ &\le C\||A|^{1/r}\|_{L^r_\lambda(d\mu)} \\ &\le C(r-2)^{-1/r}|A|^{1/r}. \end{aligned}$$

Here we have used that

$$\int \lambda^2 M_\lambda(x)\, dx \le C|A|$$

from the jump inequality in Lemma 2.1 applied with $q = 2$. We remark that $(r-2)^{-1/r}$ is bounded by $1 + (r-2)^{-1}$.

If $q > r$, then we invoke Hölder's inequality

$$\|(\lambda^2 M_\lambda(x))^{1/r}\|_{L^r_\lambda(d\mu)} \le (r-2)^{1/q-1/r}\|(\lambda^2 M_\lambda(x))^{1/r}\|_{L^q(d\mu)}$$

and

$$\int (\lambda^2 M_\lambda(x))^{q/r}\, dx \le C_{2q/r}|A|$$

and then proceed as above to obtain

$$\Big\| \big\| \mathbb{E}(f|\mathcal{D}_k)(x) \big\|_{V^r(k)} \Big\|_{L^q_x} \leq C_{2q/r}(r-2)^{-1/r}|A|^{1/q}$$

Observe that $2 < 2q/r < q$, so we can write C_q instead of $C_{2q/r}$.

If $q < r$, we will prove a weak type inequality

$$m\big\{x : \big\|\mathbb{E}(f|\mathcal{D}_k)(x)\big\|_{V^r(k)} \geq \nu\big\} \leq C_q(1+(r-2)^{-1})\nu^{-q}\|f\|^q_{L^q(\mathbb{R},H)}.$$

Define

$$E = \Big\{x : \sup_{x\in I\in\mathcal{D}} \frac{1}{|I|}\int_I |f|(y)dy \geq \nu\Big\}.$$

Outside E, we may replace f by the good part g of the Calderón-Zygmund decomposition of f in order to calculate the value of $\mathbb{E}(f|\mathcal{D}_k)$. As usual we have

$$\|g\|_{L^r(\mathbb{R},H)} \leq C\nu^{1-q/r}\|f\|^{q/r}_{L^q(\mathbb{R},H)}.$$

Hence we have

$$\begin{aligned}
m\{x : \|\mathbb{E}(f|\mathcal{D}_k)(x)\|_{V^r(k)} \geq \nu\} \\
&\leq |E| + m\{x\in E^c : \|\mathbb{E}(f|\mathcal{D}_k)(x)\|_{V^r(k)} \geq \nu\} \\
&\leq C_q\nu^{-q}\|f\|^q_{L^q(\mathbb{R},H)} + C\nu^{-r}\|\|\mathbb{E}(f|\mathcal{D}_k)(x)\|_{V^r(k)}\|^r_{L^r_x(E^c)} \\
&\leq C_q(1+(r-2)^{-1})\nu^{-q}\|f\|^q_{L^q(\mathbb{R},H)}
\end{aligned}$$

The Lemma now follows by Marcinkiewicz interpolation, passing from restricted weak type to strong type inequalities. ∎

3. General facts about Walsh time-frequency analysis

The endpoints of the dyadic intervals will be called *dyadic points*. For each dyadic interval $\omega = [a, b]$, the subintervals $\omega_1 := [a, \frac{a+b}{2}]$ and $\omega_2 := [\frac{a+b}{2}, b]$ will be referred to as the left and right children of ω, respectively.

Definition 3.1 *For two tiles (or bitiles) P and P' we write $P \leq P'$ if $I_P \subseteq I_{P'}$ and $\omega_{P'} \subseteq \omega_P$.*

Definition 3.2 *A tree with top $(I_{\mathbf{T}}, \xi_{\mathbf{T}})$ is a collection of bitiles $\mathbf{T} \subseteq \mathbf{S}$ such that $I_P \subseteq I_{\mathbf{T}}$ and $\xi_{\mathbf{T}} \in \omega_P$ for each $P \in \mathbf{T}$. An i-tree is a tree $\mathbf{T}$ such that $\xi_{\mathbf{T}} \in \omega_{P,i}$ for each $P \in \mathbf{T}$.*

Definition 3.3 *Fix some $f : \mathbb{R}_+ \to \mathbb{R}$. For a finite subset of bitiles $\mathbf{S}' \subseteq \mathbf{S}$ define its size relative to f as*

$$\mathrm{size}(\mathbf{S}') := \sup_{\mathbf{T}} \left(\frac{1}{|I_{\mathbf{T}}|} \sum_{P \in \mathbf{T}} |\langle f, w_{P_1} \rangle|^2 \right)^{\frac{1}{2}}$$

where the supremum is taken over all the 2-trees $\mathbf{T} \subset \mathbf{S}'$.

We recall a few important results regarding the size.

Proposition 3.4 *For each $1 < s < \infty$, each 2-tree $\mathbf{T}$ and each $f \in L^s(\mathbb{R}_+)$ we have*

$$\left(\frac{1}{|I_{\mathbf{T}}|} \sum_{P \in \mathbf{T}} |\langle f, w_{P_1} \rangle|^2 \right)^{1/2} \lesssim \inf_{x \in I_{\mathbf{T}}} M_s f(x).$$

Proof. See for example Lemma 1.8.1 in [5]. ■

The following Bessel type inequality, see for example [4], will be used to organize collections of bitiles into trees.

Proposition 3.5 *Let $\mathbf{S}' \subseteq \mathbf{S}$ be a collection of tiles and define*

$$\Delta := [-\log_2(\mathrm{size}(\mathbf{S}'))],$$

where the size is understood with respect to some function $f \in L^2(\mathbb{R}_+)$. Then $\mathbf{S}'$ can be written as a disjoint union $\mathbf{S}' = \bigcup_{n \geq \Delta} \mathbf{P}_n$, where $\mathrm{size}(\mathbf{P}_n) \leq 2^{-n}$ and each $\mathbf{P}_n$ consists of a family $\mathcal{F}_{\mathbf{P}_n}$ of pairwise disjoint trees satisfying

$$\sum_{\mathbf{T} \in \mathcal{F}_{\mathbf{P}_n}} |I_{\mathbf{T}}| \lesssim 2^{2n} \|f\|_2^2, \tag{3.1}$$

with bounds independent of $\mathbf{S}'$, n and f.

Elementary computations show that for each tile $P = [2^i n, 2^i(n+1)) \times [2^{-i} l, 2^{-i}(l+1))$, each $l' \geq 0$ and each $\xi \in [2^{-i} l', 2^{-i}(l'+1))$ we have

$$w_P(x) = 1_{I_P}(x) e(2^{-i} l \otimes x)$$
$$w_P(x) e(\xi \otimes x) = \epsilon(P, \xi) 1_{I_P}(x) e(2^{-i}|l' - l| \otimes x)$$

where $\epsilon(P, \xi) \in \{-1, 1\}$ depends on P and ξ but not on x. In particular, if $\mathbf{T}$ is a 2-tree and $P \in \mathbf{T}$ then

$$w_{P_1}(x) e(\xi_{\mathbf{T}} \otimes x) = \epsilon(P, \xi_{\mathbf{T}}) w_{P'}(x)$$

where $P' = [2^i n, 2^i(n+1)) \times [2^{-i}, 2^{-i+1})$, and thus $w_{P_1}(x) e(\xi_{\mathbf{T}} \otimes x)$ is constant on both the left half and the right half of I_P.

An immediate consequence is that for each $k \in \mathbb{Z}$ and each $a_P \in \mathbb{R}$

$$e(\xi_{\mathbf{T}} \otimes x) \sum_{P \in \mathbf{T}: |I_P| \geq 2^k} a_P w_{P_1}(x) = \mathbb{E}(e(\xi_{\mathbf{T}} \otimes \cdot) \sum_{P \in \mathbf{T}} a_P w_{P_1} | \mathcal{D}_{k-1})(x).$$

Since

$$\begin{aligned} w_{P_1}(x) e(\xi_{\mathbf{T}} \otimes x) &= \epsilon(P, \xi_{\mathbf{T}}) 2^{-i/2} W_1(2^{-i}x - n) \\ &= \epsilon(P, \xi_{\mathbf{T}}) h_{I_P}(x) \end{aligned}$$

where h is the Haar function, the classical theory of wavelets and John-Nirenberg's inequality imply the following.

Theorem 3.6 *Let* $\mathbf{T}$ *be a 2-tree and assume* $(a_P)_{P \in \mathbf{T}} \in \mathbb{R}$ *satisfy*

$$\Big(\frac{1}{|I|} \sum_{\substack{P \in \mathbf{T} \\ I_P \subseteq I}} |a_{I_P}|^2 \Big)^{1/2} \leq B,$$

for each dyadic interval I. *Then for each* $1 < s < \infty$

$$\Big\| e(\xi_{\mathbf{T}} \otimes \cdot) \sum_{P \in \mathbf{T}} a_P w_{P_1} \Big\|_{BMO} \lesssim B$$

and

$$\Big\| e(\xi_{\mathbf{T}} \otimes \cdot) \sum_{P \in \mathbf{T}} a_P w_{P_1} \Big\|_s \lesssim_s B |I_{\mathbf{T}}|^{1/s}.$$

As an immediate consequence of Theorem 3.6 and of Lemma 2.2 we obtain the following.

Theorem 3.7 *Let* $\mathbf{T}$ *be a 2-tree,* $f : \mathbb{R}_+ \to \mathbb{R}$ *and let* $\operatorname{size}(\mathbf{T})$ *denote the size of* $\mathbf{T}$ *with respect to the function* f. *Then for each* $1 < s < \infty$

$$\Big\| \| \sum_{\substack{P \in \mathbf{T} \\ |I_P| \leq 2^k}} a_P w_{P_1}(x) \|_{V^r(k)} \Big\|_{L^s_x} \lesssim_s \operatorname{size}(\mathbf{T}) |I_{\mathbf{T}}|^{1/s}.$$

4. A generalization of a Lemma of Bourgain

In this section we generalize a maximal multiplier result due to Bourgain [1] We begin with the following easy consequence of Minkowski's inequality.

Lemma 4.1 *Let* Ξ *be a finite set. Consider also two sequences* a_k *and* b_k *in the Hilbert space* $l^2(\Xi)$ *and define* $a_k \star b_k \in l^2(\Xi)$ *by* $(a_k \star b_k)_\xi = (a_k)_\xi (b_k)_\xi$. *Then*

$$\|a_k \star b_k\|_{V^r(k)} \lesssim (\sum_{\xi \in \Xi} \|(a_k)_\xi\|^2_{V^r(k)} \|(b_k)_\xi\|^2_{V^r(k)})^{1/2}.$$

Proposition 4.2 *Let H be a Hilbert space. Assume we are given a set A of linear functionals $f \to f^{(\alpha)} = \langle f, e^{(\alpha)} \rangle$, $e^{(\alpha)} \in H$, of norm less than ϵ such that*

$$\sum_{\alpha \in A} |f^{(\alpha)}|^2 \leq |f|^2$$

for each $f \in H$. Set $N = \epsilon^2 |A|$. Let f_k be a sequence of H-valued functions on $\mathbb{R}$ such that we have the variational inequality

$$\big\| \|f_k(x)\|_{V^r(k)} \big\|_{L^2_x} \leq F.$$

Then we have

$$\big\| \| \sup_k |f_k^{(\alpha)}(x)| \|_{l^2(A)} \big\|_{L^2_x} \leq C N^{r/4-1/2} F.$$

A special example of a collection of linear functionals as in the Lemma can be obtained by choosing the $e^{(\alpha)}$ to be an orthonormal family of vectors and $\epsilon = 1$. Our main application will involve a more general set of linear functionals. We remark that the difficulty in this proposition comes from the fact that we take the supremum in k before we take the square sum of the components.

Proof. Fix x and define $C_x = \{f_k(x)\}$ and $d(x) = \operatorname{diam}(C_x)$. It suffices to prove the Proposition in the case C_x is finite and then to invoke the Monotone Convergence Theorem. Also, we can assume with no loss of generality that C_x contains the origin $\mathbf{0}$. For each $\lambda > 0$ denote by $N_\lambda(x)$ the minimum number of balls with radius λ and centered at elements of C_x, whose union covers C_x. It is an easy exercise to prove that

$$\sup_{\lambda > 0} \lambda N_\lambda^{1/r}(x) \lesssim_r \|f_k(x)\|_{V^r(k)}, \tag{4.1}$$

with the implicit constant depending only on r. For each $n \geq -\log_2(d(x))$, let $C_{n,x}$ be a collection of elements of $(C_x - C_x)$ such that

$$|c|_H \leq 2^{-n+2} \text{ for each } c \in C_{n,x},$$
$$\sharp C_n \leq N_{2^{-n}}(x) + 1$$

and each $c \in C_{n,x}$ can be written as

$$c = \sum_{n \geq -\log_2(d(x))} c_n \text{ with } c_n \in C_{n,x}. \tag{4.2}$$

Here is how $C_{n,x}$ is constructed. For each $n \geq -\log_2(d(x))$ define $B_{n,x}$ to be a collection of $N_{2^{-n}}(x)$ elements of C_x such that the balls with centers in $B_{n,x}$

and radius 2^{-n} cover C_x. If $n = [-\log_2(d(x))] - 1$ define $B_{n,x} = \{\mathbf{0}\}$. For each $n \geq -\log_2(d(x))$ and each $c \in B_{n,x}$, choose an element $c' \in B_{n-1,x}$ such that the ball centered at c and with radius 2^{-n} intersects the ball centered at c' and with radius 2^{-n+1}. Define

$$C_{n,x} := \{c - c' : c \in B_{n,x}\} \cup \{\mathbf{0}\}.$$

Since C_x is finite, for each $c \in C_x$ there is n such that $c \in B_{n,x}$. To verify the representation (4.2) for an arbitrary $c \in C_x$, denote as above by c' the element from $B_{n-1,x}$ associated with c, by c'' the element from $B_{n-2,x}$ associated with c' and so on, and note that this sequence will eventually terminate with $\mathbf{0}$. Hence we can write

$$c = (c - c') + (c' - c'') + \dots .$$

Note also that by construction, each element of $C_{n,x}$ has norm at most 2^{-n+2}.

This together with inequality (4.1) further allows us to write for each x and α

$$\begin{aligned}\sup_k |f_k^{(\alpha)}(x)| &\leq \sum_{n \geq -\log_2(d(x))} \sup_{c_n \in C_{n,x}} |c_n^{(\alpha)}| \\ &\lesssim \sum_{n \geq -\log_2(d(x))} \min\left(2^{-n}\epsilon, \Big(\sum_{c_n \in C_{n,x}} |c_n^{(\alpha)}|^2\Big)^{1/2}\right).\end{aligned}$$

Summing over α we get

$$\begin{aligned}\sum_\alpha (\sup_k |f_k^{(\alpha)}(x)|)^2 &\lesssim \sum_{n \geq -\log_2(d(x))} \min\left(2^{-2n}N, \sum_{c_n \in C_{n,x}} |c_n|_H^2\right) \\ &\lesssim 2^{-2n} \sum_{n \geq -\log_2(d(x))} \min(N, N_{2^{-n}}(x)).\end{aligned}$$

Taking finally the L^2 norm in x gives

$$\begin{aligned}&\Big\| \| \sup_k |f_k^{(\alpha)}(x)| \|_{l^2(A)} \Big\|_{L_x^2}^2 \\ &\lesssim \int \sum_{2^{-n} < d(x)/N^{1/2}} 2^{-2n} N \, dx + \int \sum_{d(x)/N^{1/2} \leq 2^{-n} \leq d(x)} 2^{-2n} N_{2^{-n}}(x) \, dx \\ &\lesssim \int d^2(x) \, dx + \int \sum_{d(x)/N^{1/2} \leq 2^{-n}} 2^{-(2-r)n} 2^{-rn} N_{2^{-n}}(x) \, dx \\ &\lesssim \int d^2(x) \, dx + N^{r/2-1} \int d(x)^{2-r} \sum_n 2^{-rn} N_{2^{-n}}(x) \, dx \\ &\lesssim N^{r/2-1} \int \|f_k(x)\|_{V^r(k)}^2 \, dx \lesssim N^{r/2-1} F^2.\end{aligned}$$

This finishes the proof. ∎

Corollary 4.3 *Let $2 < r < \infty$. Assume we are given a set $\Xi \subset \mathbb{R}_+$ of cardinality $N > 1$ and assume that there is no dyadic interval of length 1 which contains more than one point in Ξ. For every $k \geq 0$ define Ω_k to be the union of all dyadic intervals of length 2^{-k} which have nonempty intersection with Ξ. For each $\omega \in \Omega_k$ let ϵ_ω be a number so that for every nested sequence of intervals $\omega_k \in \Omega_k$ we have*

$$\|\epsilon_{\omega_k}\|_{V^r(k)} \leq \sigma. \tag{4.3}$$

Define

$$\Delta_k f(x) = \Big(\sum_{\omega \in \Omega_k} \epsilon_\omega 1_\omega \widehat{f}\Big)\check{}(x).$$

Then

$$\big\|\sup_{k\geq 0} |\Delta_k f|\big\|_2 \lesssim_r \sigma N^{r/4-1/2}\|f\|_2.$$

Proof. Fix $f \in L^2(\mathbb{R}_+)$. For each $k \geq 0$ we will denote by $\omega_{\xi,k}$ the unique dyadic interval in Ω_k such that $\xi \in \omega_{\xi,k}$, and by $w_\xi(x) = e(x \otimes \xi)$. Let H be the N dimensional Hilbert space $l^2(\Xi)$. Define the sequence of functions $f_k : \mathbb{R} \to H$, $k \geq 0$, by

$$\begin{aligned}(f_k(x))_\xi &= \epsilon_{\omega_{\xi,k}}(\widehat{f}1_{\omega_{\xi,k}})\check{}(x)\\ &= \epsilon_{\omega_{\xi,k}} w_\xi(x)\mathbb{E}(f w_\xi|\mathcal{D}_k)(x),\end{aligned}$$

and note that

$$(f_k(x))_\xi = w_\xi(y)(f_k(x \oplus y))_\xi \tag{4.4}$$

for all $y \in [0,1)$.

To construct the vectors $e^{(\alpha)}$, choose some small negative integer m so that all w_ξ are constant on dyadic subintervals of $[0,1)$ of length 2^m. We write $w_\xi(J)$ for this constant value on such an interval J. For each such interval J_α, $\alpha \in A := \{1, 2, \ldots, 2^{-m}\}$, define

$$e^{(\alpha)} = (2^{m/2} w_\xi(J_\alpha))_{\xi\in\Xi}.$$

The corresponding linear functionals are of norm $\epsilon = 2^{m/2}|\Xi|^{1/2}$. We also have

$$\sum_\alpha |g^{(\alpha)}|^2 \leq \int_0^1 |\sum_{\xi\in\Xi} g_\xi w_\xi(x)|^2\, dx \leq \sum_\xi |g_\xi|^2,$$

for each $g \in H$. In the last inequality we have used that the functions w_ξ are orthogonal on $[0,1)$. Hence the functionals satisfy the assumption of

Proposition 4.2 with $N = \epsilon^2|A| = |\Xi|$. We observe the following

$$\begin{aligned}
\|\|\sup_k |f_k^{(\alpha)}(x)|\|_{l^2(A)}\|_{L^2_x} &= \int_{\mathbb{R}_+}\int_0^1 \sup_k |\sum_{\xi\in\Xi} \epsilon_{\omega_{\xi,k}} w_\xi(y)(\widehat{f}1_{\omega_{\xi,k}})\check{}(x)|^2 dy\, dx \\
&= \int_{\mathbb{R}_+}\int_0^1 \sup_k |\sum_{\xi\in\Xi} \epsilon_{\omega_{\xi,k}} w_\xi(y)(\widehat{f}1_{\omega_{\xi,k}})\check{}(x\oplus y)|^2 dy\, dx \\
&= \int_{\mathbb{R}_+} \sup_k |\sum_{\xi\in\Xi} \epsilon_{\omega_{\xi,k}} (\widehat{f}1_{\omega_{\xi,k}})\check{}(x)|^2\, dx \\
&= \|\sup_{k\geq 0}|\Delta_k f|\|_2^2
\end{aligned}$$

where the last equality is a consequence of (4.4). The corollary now follows from Proposition 4.2 once we verify that

$$\|\|f_k(x)\|_{V^r(k)}\|_{L^2_x} \lesssim_r \sigma\|f\|_2.$$

Note that for each x, $f_k(x) = a_{k,x} \star b_{k,x}$, where $(a_{k,x})_\xi = \epsilon_{\omega_{\xi,k}}$ and $(a_{k,x})_\xi = w_\xi(x)\mathbb{E}(fw_\xi|\mathcal{D}_k)(x)$. The above estimate is now a consequence of Lemma 2.2, Lemma 4.1 and inequality (4.3). ■

An argument very similar to the above also proves the following version of Corollary 4.3:

Corollary 4.4 *Consider a collection Ω of N disjoint dyadic intervals $\omega\in\mathbb{R}_+$. For each $\omega\in\Omega$ and each $k\in\mathbb{Z}$ let $\epsilon_{k,\omega}\in\mathbb{R}$. Define*

$$\Delta_k f(x) := \sum_{\omega\in\Omega} \epsilon_{k,\omega}(\widehat{f}1_\omega)\check{}(x).$$

Then for each $r > 2$

$$\big\|\sup_k |\Delta_k f|\big\|_{L^2} \lesssim_r N^{r/4-1/2} \sup_{\omega\in\Omega} \|\epsilon_{k,\omega}\|_{V^r(k)}\|f\|_{L^2}.$$

It turns out that the results of corollaries 4.3 and 4.4 are not general enough for our applications, and so we prove the following more general version. Consider now an arbitrary set $\Xi = \{\xi_1,\dots,\xi_N\}$ with no further restrictions on it, and for each $k\in\mathbb{Z}$ define Ω_k to be the set of all dyadic intervals of length 2^{-k} which contain some element of Ξ. We now associate to each $\omega\in\bigcup_k \Omega_k$ a number $\epsilon_\omega\in\mathbb{R}$ and define

$$\Delta_k f(x) := \sum_{\omega\in\Omega_k} \epsilon_\omega(\widehat{f}1_\omega)\check{}(x). \tag{4.5}$$

Proposition 4.5 *For each $r > 2$ we have the inequality*

$$\| \sup_k |\Delta_k f| \|_2 \lesssim_r N^{r/4-1/2} \sigma \|f\|_2,$$

where

$$\sigma = \sup_n \sup_{\xi_n \in \omega_k \in \Omega_k} \|\epsilon_{\omega_k}\|_{V^r(k)}.$$

Proof. It suffices as before to assume that the index k runs through a finite interval $\{a, a+1, \dots, b\}$ with $a, b \in \mathbb{Z}$. We can find a sequence $a = k_0 < k_1 < \dots < k_L = b$ with $L \leq N$, such that for each $0 \leq j \leq L-1$, Ω_k has the same cardinality when $k_j \leq k < k_{j+1}$. If $\widehat{f_j} := (\sum_{\omega \in \Omega_{k_j}} 1_\omega - \sum_{\omega \in \Omega_{k_{j+1}}} 1_\omega) \widehat{f}$, then the functions f_j are pairwise orthogonal. We can now bound $\| \sup_k |\Delta_k f| \|_2$ by

$$\Big\| \sup_j \sup_{k_j \leq k < k_{j+1}} \Big| \Big(\sum_{\omega \in \Omega_{k_{j+1}}} \epsilon_{\omega(k)} 1_\omega \sum_{j' > j} \widehat{f_{j'}} \Big)^\vee \Big| \Big\|_2 + \tag{4.6}$$

$$+ \Big\| \sup_j \sup_{k_j \leq k < k_{j+1}} \Big| \Big(\sum_{\omega \in \Omega_k} \epsilon_\omega 1_\omega \widehat{f_j} \Big)^\vee \Big| \Big\|_2 . \tag{4.7}$$

For each $\omega \in \Omega_{k_{j+1}}$ and each $k_j \leq k < k_{j+1}$, $\omega(k)$ is defined to be the interval in Ω_k containing ω. Corollary 4.3 and scaling invariance show that the term (4.7) can be bounded by

$$\begin{aligned} &\Big(\sum_j \Big\| \sup_{k_j \leq k < k_{j+1}} \Big| \Big(\sum_{\omega \in \Omega_k} \epsilon_\omega 1_\omega \widehat{f_j} \Big)^\vee \Big| \Big\|_2^2 \Big)^{1/2} \\ &\qquad \lesssim \Big(\sum_j N^{r/2-1} \sup_n \sup_{\substack{\xi_n \in \omega_k \in \Omega_k \\ k_j \leq k < k_{j+1}}} \|\epsilon_{\omega_k}\|_{V^r(k)}^2 \|f_j\|_2^2 \Big)^{1/2} \\ &\qquad \lesssim \sigma N^{r/4-1/2} \|f\|_2. \end{aligned}$$

To estimate the term in (4.6), define the maximal operators

$$O_j^*(h) := \sup_{k_j \leq k < k_{j+1}} \Big| \Big(\sum_{\omega \in \Omega_{k_{j+1}}} \epsilon_{\omega(k)} 1_\omega \widehat{h} \Big)^\vee \Big|.$$

We will argue that

$$\Big\| \sup_{1 \leq j \leq N} O_j^* \Big(\sum_{j \leq j' \leq N} f_{j'} \Big) \Big\|_2 \lesssim \sigma N^{r/4-1/2} \Big(\sum_{j=1}^L \|f_j\|_2^2 \Big)^{1/2}.$$

It suffices to consider only dyadic values of N so we will assume that $N = 2^M$, for some $M \geq 0$. For each $0 \leq m \leq M$, denote by A_m the best constant for which the following inequality holds for all discrete dyadic intervals $J = (j_1, j_2] := \{j_1+1, j_1+2, \dots, j_2\}$[1] $\subseteq \{1, 2 \dots, 2^M\}$ with 2^m elements

$$\Big\|\sup_{j\in J} O_j^*\Big(\sum_{j\leq j'\leq j_2} f_{j'}\Big)\Big\|_2 \lesssim A_m \Big(\sum_{j\in J}\|f_j\|_2^2\Big)^{1/2}.$$

We will use a reasoning similar to the one in the proof of the Rademacher-Menshov inequality, to argue that $A_M \lesssim B_M$, where

$$B_m := \sigma 2^{m(r/4-1/2)}.$$

We can write for each $0 \leq m \leq M-1$ and each discrete dyadic interval $J = (j_1, j_2] \subseteq \{1, 2, \dots, 2^M\}$ having 2^{m+1} elements and midpoint $j_3 := j_1 + 2^m$

$$\begin{aligned}
&\Big\|\sup_{j\in J} O_j^*\Big(\sum_{j\leq j'\leq j_2} f_{j'}\Big)\Big\|_2^2 \\
&\quad\leq \Big\|\sup_{j_3+1\leq j\leq j_2} O_j^*\Big(\sum_{j\leq j'\leq j_2} f_{j'}\Big)\Big\|_2^2 + \\
&\qquad + \Big(\Big\|\sup_{j_1+1\leq j\leq j_3} O_j^*\Big(\sum_{j\leq j'\leq j_3} f_{j'}\Big)\Big\|_2 + \Big\|\sup_{j_1+1\leq j\leq j_3} O_j^*\Big(\sum_{j_3+1\leq j'\leq j_2} f_{j'}\Big)\Big\|_2\Big)^2.
\end{aligned}$$

We then use the definition of A_m for the first two terms above and Corollary 4.4 for the third one, to bound the sum above by

$$\begin{aligned}
A_m^2 \sum_{j_3+1\leq j'\leq j_2}\|f_{j'}\|_2^2 + \Big(A_m\Big(\sum_{j_1\leq j'\leq j_3}\|f_{j'}\|_2^2\Big)^{1/2} + CB_m\Big(\sum_{j_3+1\leq j'\leq j_2}\|f_{j'}\|_2^2\Big)^{1/2}\Big)^2 \\
\leq (A_m + CB_m)^2 \sum_{j\in J}\|f_j\|_2^2.
\end{aligned}$$

We conclude that $A_{m+1} \leq A_m + CB_m$ for each $0 \leq m \leq M-1$, which together with the fact that $A_0 = 0$ proves that $A_M \lesssim B_M$. ■

Remark 4.6 *If in the above proposition we choose $\epsilon_\omega = 1$ for each ω, we recover the result of Bourgain from [1], with a slightly larger dependence on N of the bound. While Bourgain's bound is logarithmic in N, a bound of the form $N^{r/4-1/2}$ will suffice for our later applications, since we afford to take r as close to 2 as we want.*

[1] j_1 and j_2 are of the form $a2^b$ with $a, b \in \mathbb{Z}_+$

5. Pointwise estimates outside exceptional sets

5.1. An estimate for a collection of 2-trees

Assume we have a collection $\mathbf{S}' \subseteq \mathbf{S}$ of bitiles which can be written as a not necessarily disjoint union of 2-trees

$$\mathbf{S}' = \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}.$$

We shall assume that if $\mathbf{T} \in \mathcal{F}$, then $\mathbf{T}$ is indeed the maximal 2-tree in $\mathbf{S}'$ with the top $(I_{\mathbf{T}}, \xi_{\mathbf{T}})$, that is, all bitiles in $P \in \mathbf{S}'$ which satisfy $I_P \subseteq I_{\mathbf{T}}$ and $\xi_{\mathbf{T}} \in \omega_{P,2}$ are in $\mathbf{T}$.

Theorem 5.1 *For each $\beta \geq 1$, $\gamma > 0$ and each $(a_P)_{P \in \mathbf{S}'}$ define the exceptional sets*

$$E^{(1)} = \Big\{x : \sum_{\mathbf{T} \in \mathcal{F}} 1_{I_{\mathbf{T}}}(x) > \beta\Big\},$$
$$E^{(2)} = \bigcup_{\mathbf{T} \in \mathcal{F}} \Big\{x : \Big\| \sum_{\substack{P \in \mathbf{T} \\ |I_P| < 2^k}} a_P w_{P_1}(x) \Big\|_{V^r(k)} > \gamma\Big\}.$$

Then for each $x \notin E^{(1)} \cup E^{(2)}$ and each $r > 2$ we have

$$\Big\| \Big(\sum_{\substack{P \in \mathbf{S}' \\ |I_P| \leq 2^k}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \Big)_{k \in \mathbb{Z}} \Big\|_{M_2^*(\theta)} \lesssim_r \gamma \beta^{r/4 - 1/2}.$$

Proof. Fix x not in the union of the exceptional sets and let $\mathcal{F}_x$ be the family of all trees $\mathbf{T} \in \mathcal{F}$ with $x \in I_{\mathbf{T}}$. Define

$$\Xi_x = \{\xi_{\mathbf{T}}, \mathbf{T} \in \mathcal{F}_x\}.$$

For each $k \in \mathbb{Z}$ let Ω_k be the collection of dyadic frequency intervals of length 2^{-k} which contain an element of Ξ_x. Let $\tilde{\Omega}_k$ be the collection of all children of intervals in Ω_{k-1} that are not themselves in Ω_k. Observe that both $\bigcup_{k'} \tilde{\Omega}_{k'}$ and $\Omega_k \cup \bigcup_{k' \leq k} \tilde{\Omega}_{k'}$ are collections of pairwise disjoint intervals which cover $\mathbb{R}_+$ (with the possible exception of finitely many dyadic points). Moreover we can write

$$\sum_{P \in \mathbf{S}' : |I_P| < 2^k} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) = \sum_{\omega \in \Omega_k} 1_\omega(\theta) \sum_{\substack{P \in \mathbf{S}' \\ |I_P| < 2^k,\ \omega \cap \omega_{s,2} \neq \emptyset}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta)$$
$$+ \sum_{k' \leq k} \sum_{\omega \in \tilde{\Omega}_{k'}} 1_\omega(\theta) \sum_{\substack{P \in \mathbf{S}' \\ |I_P| < 2^k,\ \omega \cap \omega_{P,2} \neq \emptyset}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta).$$

Indeed, if $1_\omega(\theta) w_{P_1} 1_{\omega_{P,2}}(\theta) \not\equiv 0$ for some $\omega \in \Omega_k \cup \bigcup_{k' \leq k} \tilde{\Omega}_{k'}$ and $P \in \mathbf{S}'$, then this implies that $\omega \cap \omega_{P,2} \neq \emptyset$.

Moreover, when $\omega \in \Omega_k$, this latter restriction together with $|I_P| < 2^k$ is equivalent with just asking that $\omega \subseteq \omega_{P,2}$. Similarly, when $\omega \in \bigcup_{k' \le k} \tilde{\Omega}_{k'}$ then $\omega_{P,2} \subsetneq \omega$ is impossible, which in turn makes the requirement $|I_P| < 2^k$ superfluous. Indeed $\omega_{P,2} \subsetneq \omega$ would imply that $\omega_P \subseteq \omega$, contradicting the fact that ω_P contains an element from Ξ_x while ω does not. Hence we can rewrite

$$\sum_{P \in \mathbf{S}' : |I_P| < 2^k} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) =$$

$$= \sum_{\omega \in \Omega_k} 1_\omega(\theta) \sum_{\substack{P \in \mathbf{S}' \\ \omega \subseteq \omega_{P,2}}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \tag{5.1}$$

$$+ \sum_{k' \le k} \sum_{\omega \in \tilde{\Omega}_{k'}} 1_\omega(\theta) \sum_{\substack{P \in \mathbf{S}' \\ \omega \subseteq \omega_{P,2}}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta). \tag{5.2}$$

The multiplier in (5.2) can be written more conveniently as

$$\Big(1 - \sum_{\tilde{\omega} \in \Omega_k} 1_{\tilde{\omega}}\Big)\Big(\sum_{k'} \sum_{\omega \in \tilde{\Omega}_{k'}} 1_\omega(\theta) \sum_{P \in \mathbf{S}',\, \omega \subseteq \omega_{P,2}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta)\Big) =$$
$$= (1 - \sum_{\tilde{\omega} \in \Omega_k} 1_{\tilde{\omega}}) \sum_{P \in \mathbf{S}'} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta),$$

given the fact that

$$\Big(\bigcup_{I \in \Omega_k} I\Big)^c = \bigcup_{k' \le k} \bigcup_{I \in \tilde{\Omega}_{k'}} I \qquad \text{and} \qquad \Big(\bigcup_{k' \le k} \tilde{\Omega}_{k'}\Big) \bigcap \Big(\bigcup_{k' > k} \tilde{\Omega}_{k'}\Big) = \emptyset,$$

modulo some dyadic points. This maximal multiplier operator is now easily seen to be the composition of two operators. One is the identity minus Bourgain's maximal operator for which Proposition 4.5 provides good bounds. The second one is a linear operator associated with the multiplier $\sum_{P \in \mathbf{S}'} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta)$. To analyze the latter operator, we note that for each θ the contribution to the multiplier comes from a single tree. To see this note that the collection

$$\mathcal{A} := \big\{P \in \mathbf{S}' : x \in I_P,\ \theta \in \omega_{P,2}\big\}$$

is finite and totally ordered and hence it contains a maximum element P_θ. If $\mathbf{T}_\theta \in \mathcal{F}$ is one of the 2-trees to which P_θ belongs, then by the maximality condition in the hypothesis it follows that $P \in \mathbf{T}_\theta$ for each $P \in \mathcal{A}$. Moreover, there is some k such that

$$\sum_{P \in \mathbf{S}'} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) = \sum_{\substack{P \in \mathbf{T}_\theta \\ |I_P| \le 2^k}} a_P w_{P_1}(x).$$

By invoking Proposition 4.5 and the fact that $x \notin E^{(1)} \cup E^{(2)}$ we get that

$$\Big\| \Big(\Big(1 - \sum_{\tilde\omega \in \Omega_k} 1_{\tilde\omega}\Big) \sum_{P \in \mathbf{S}'} a_P w_{P_1}(x) 1_{\omega_{P,2}} \Big)_{k \in \mathbb{Z}} \Big\|_{M_2^*}$$
$$\leq \Big(1 + \Big\| \Big(\sum_{\tilde\omega \in \Omega_k} 1_{\tilde\omega}\Big)_{k\in\mathbb{Z}} \Big\|_{M_2^*}\Big) \Big\| \sum_{P \in \mathbf{S}'} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \Big\|_{L^\infty(\theta)}$$
$$\lesssim \gamma \beta^{r/4-1/2}.$$

The term (5.1) is clearly of the form

$$\sum_{\omega \in \Omega_k} 1_\omega(\theta) \epsilon_\omega$$

with

$$\epsilon_\omega = \sum_{\substack{P \in \mathbf{S}' \\ \omega \subseteq \omega_{P,2}}} a_P w_{P_1}(x).$$

We claim that for each nested sequence of intervals $\omega_k \in \Omega_k$ we have the variational norm estimate

$$\|\epsilon_{\omega_k}\|_{V^r(k)} \lesssim_r \gamma. \tag{5.3}$$

This follows immediately from the fact that all bitiles contributing to ϵ_{ω_k} belong to a single tree $\mathbf{T} \in \mathcal{F}$. Indeed, the collection

$$\mathcal{B} := \big\{P \in \mathbf{S}' : x \in I_P,\ \omega_k \subseteq \omega_{P,2} \text{ for some } \ k\big\}$$

is finite and totally ordered, so it has a maximum element P_x. If $\mathbf{T}_x \in \mathcal{F}_x$ is one of the 2-trees to which P_x belongs, then from the maximality condition in the hypothesis it follows that $P \in \mathbf{T}_x$ for each $P \in \mathcal{B}$. Moreover, for each k

$$\big\{P \in \mathbf{S}' : \omega_k \subseteq \omega_{P,2}\big\} = \big\{P \in \mathbf{T}_x : |I_P| \leq 2^k\big\}.$$

Thus (5.3) and Proposition 4.5 imply that

$$\Big\| \Big(\sum_{\omega \in \Omega_k} 1_\omega(\theta) \sum_{\substack{P \in \mathbf{S}' \\ \omega \subseteq \omega_{P,2}}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \Big)_{k \in \mathbb{Z}} \Big\|_{M_2^*(\theta)} \lesssim \gamma \beta^{r/4-1/2}.$$

■

5.2. An estimate for a collection of 1-trees

The discussion here is very similar to that for 2-trees. Assume we have a collection $\mathbf{S}'$ of bitiles which can be written as a not necessarily disjoint union of finitely many 1-trees

$$\mathbf{S}' = \bigcup_{\mathbf{T} \in \mathcal{F}} \mathbf{T}.$$

We shall assume that for every $P \in \mathbf{S}'$ there does not exist a tree $\mathbf{T} \in \mathcal{F}$ with $I_P \subset I_{\mathbf{T}}$ and $\xi_{\mathbf{T}} \in \omega_{P,2}$. This assumption does in particular imply that the upper tiles P_2 are pairwise disjoint. For assume not and $I_P \subsetneq I_{P'}$ and $\omega_{P',2} \subsetneq \omega_{P,2}$ for some P, P', then it is easy to see that the upper tile P_2 violates the above assumption with respect to any tree to which P' belongs.

Theorem 5.2 *Let $(a_P)_{P\in\mathbf{S}'}$ satisfy*

$$\sup_{P\in\mathbf{S}'} \frac{|a_P|}{|I_P|^{1/2}} \leq \sigma. \tag{5.4}$$

For each $\alpha \geq 1$ define the exceptional set

$$E = \Big\{x : \sum_{\mathbf{T}\in\mathcal{F}} 1_{I_{\mathbf{T}}}(x) > \beta\Big\}.$$

Then for each $x \notin E$ and each $r > 2$ we have

$$\Big\|\Big(\sum_{\substack{P\in\mathbf{S}' \\ |I_P|\leq 2^k}} a_P w_{P_1}(x) 1_{\omega_{P,2}}\Big)_{k\in\mathbb{Z}}\Big\|_{M_2^*} \lesssim_r \sigma\beta^{r/4-1/2}.$$

Proof. As before we write

$$\sum_{P\in\mathbf{S}':|I_P|<2^k} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta) = \sum_{\omega\in\Omega_k} 1_\omega(\theta) \sum_{\substack{P\in\mathbf{S}' \\ \omega\subseteq\omega_{P,2}}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta)$$
$$+ \sum_{k'\leq k*} \sum_{\omega\in\tilde{\Omega}_{k'}} 1_\omega(\theta) \sum_{\substack{P\in\mathbf{S}' \\ \omega\subseteq\omega_{P,2}}} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta).$$

The argument continues as in the previous section. Since the upper tiles P_2 are pairwise disjoint, the collections $\mathcal{A}$ and $\mathcal{B}$ contain at most one bitile. This observation together with (5.4) implies that

$$\Big\|\sum_{P\in\mathbf{S}'} a_P w_{P_1}(x) 1_{\omega_{P,2}}(\theta)\Big\|_{L^\infty(\theta)} \leq \sigma \qquad \text{and} \qquad \|\epsilon_{\omega_k}\|_{V^r(k)} \lesssim_r \sigma$$

An application of Proposition 4.5 ends the proof. ∎

5.3. Arbitrary collection of trees

Let $\mathbf{S}'$ be an arbitrary collection of bitiles which can be written as a not necessarily disjoint union of finitely many trees

$$\mathbf{S}' = \bigcup_{\mathbf{T}\in\mathcal{F}} \mathbf{T}.$$

We next show that $\mathbf{S}'$ can be split into a collection of 2-trees like in Section 5.1 and a collection of 1-trees like in Section 5.2.

For each $\mathbf{T} \in \mathcal{F}$ let $\mathbf{T}^{(2)}$ be the collection of all bitiles $P \in \mathbf{S}'$ such that $I_P \subseteq I_{\mathbf{T}}$ and $\xi_{\mathbf{T}} \in \omega_{P,2}$. If $\mathbf{S}^{(2)}$ denotes the union of all trees $\mathbf{T}^{(2)}$, then $\mathbf{S}^{(2)}$ qualifies as a collection of trees as in Section 5.1.

For each $\mathbf{T} \in \mathcal{F}$ let $\mathbf{T}^{(1)}$ be the collection of all bitiles $P \in \mathbf{S}' \setminus \mathbf{S}^{(2)}$ such that $I_P \subseteq I_{\mathbf{T}}$ and $\xi_{\mathbf{T}} \in \omega_{P,1}$. If $\mathbf{S}^{(1)}$ be the union of all trees $\mathbf{T}^{(1)}$, then $\mathbf{S}^{(1)}$ qualifies as a collection of trees as in Section 5.2. The additional geometric assumption is satisfied since we have exhausted all 2-trees first.

We will denote by $\mathcal{F}^{(2)}$ and $\mathcal{F}^{(1)}$ respectively the two families of trees that arise from the above procedure. An immediate consequence of the results in the previous two subsections is the following theorem.

Theorem 5.3 *Let* $(a_P)_{P\in\mathbf{S}'}$ *satisfy*

$$\sup_{P\in\mathbf{S}'} \frac{|a_P|}{|I_P|^{1/2}} \le \sigma.$$

For each $\beta \ge 1$ *and* $\gamma > 0$ *define the exceptional sets*

$$E^{(1)} = \Big\{x : \sum_{\mathbf{T}\in\mathcal{F}} 1_{I_{\mathbf{T}}}(x) > \beta\Big\},$$

$$E^{(2)} = \bigcup_{\mathbf{T}\in\mathcal{F}^{(2)}} \Big\{x : \Big\| \sum_{\substack{P\in\mathbf{T} \\ |I_P|<2^k}} a_P w_{P_1}(x) \Big\|_{V^r(k)} > \gamma\Big\}.$$

Then for each $x \notin E^{(1)} \cup E^{(2)}$ *and each* $r > 2$ *we have*

$$\Big\| \Big(\sum_{\substack{P\in\mathbf{S}' \\ |I_P|\le 2^k}} a_P w_{P_1}(x) 1_{\omega_{P,2}} \Big)_{k\in\mathbb{Z}} \Big\|_{M_2^*} \lesssim_r (\sigma+\gamma)\beta^{r/4-1/2}.$$

6. Main argument

In this section we present the proof of Theorem 1.3. For each collection of bitiles $\mathbf{S}' \subseteq \mathbf{S}$ define the following operator.

$$V_{\mathbf{S}'} f(x) = \Bigg\| \Bigg(\sum_{\substack{P\in\mathbf{S}' \\ |I_P|<2^k}} \langle f, w_{P_1}\rangle w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \Bigg)_{k\in\mathbb{Z}} \Bigg\|_{M_2^*(\theta)}.$$

Note that for each $\mathbf{S}'$ the operator $V_{\mathbf{S}'}$ is sublinear as a function of f. Also, for each f and x the mapping $\mathbf{S}' \to T_{\mathbf{S}'} f(x)$ is sublinear as a function of the bitile set $\mathbf{S}'$. We will prove in the following that

$$m\{x : V_{\mathbf{S}} 1_F(x) \gtrsim \lambda\} \lesssim_p \frac{|F|}{\lambda^p}, \tag{6.1}$$

for each $F \subseteq \mathbb{R}_+$ of finite measure, each $\lambda > 0$ and each $1 < p < \infty$.

Then, by invoking the Marcinkiewicz interpolation theorem and restricted weak type interpolation we get for each $1 < p < \infty$ that

$$\|V_{\mathbf{S}} f\|_p \lesssim_p \|f\|_p.$$

Fix F and λ. We first prove (6.1) in the case $\lambda \le 1$. Define the first exceptional set

$$E := \{x : M_p 1_F(x) \ge \lambda\}$$

and note that

$$|E| \lesssim \frac{|F|}{\lambda^p}.$$

Since the range of p is open, it thus suffices to prove that for each $\epsilon > 0$

$$m\{x \in \mathbb{R} : V_{\mathbf{S}_1} 1_F(x) \gtrsim \lambda^{1-\epsilon}\} \lesssim_{\epsilon,p} \frac{|F|}{\lambda^p}, \tag{6.2}$$

where

$$\mathbf{S}_1 = \{P \in \mathbf{S} : I_P \cap E^c \ne \emptyset\}.$$

Proposition 3.4 guarantees that $\text{size}(\mathbf{S}_1) \lesssim \lambda$, where the size is understood here with respect to the function 1_F. Define $\Delta := [-\log_2(\text{size}(\mathbf{S}_1))]$. Use the result of Proposition 3.5 to split $\mathbf{S}_1$ as a disjoint union $\mathbf{S}_1 = \bigcup_{n\ge\Delta} \mathbf{P}_n$, where $\text{size}(\mathbf{P}_n) \le 2^{-n}$ and each $\mathbf{P}_n$ consists of a family $\mathcal{F}_{\mathbf{P}_n}$ of trees satisfying

$$\sum_{\mathbf{T}\in\mathcal{F}_{\mathbf{P}_n}} |I_{\mathbf{T}}| \lesssim 2^{2n}|F|. \tag{6.3}$$

Let $\epsilon > 0$ be an arbitrary positive number. For each $n \ge \Delta$ define $\sigma := 2^{-n}$, $\beta := 2^{3n}\lambda^p$, $\gamma := 2^{-n/2}\lambda^{1/2-\epsilon}$. Define $a_P := \langle 1_F, w_{P_1}\rangle$ for each $P \in \mathbf{P}_n$ and note that the collection $\mathbf{P}_n$ together with the coefficients $(a_P)_{P\in\mathbf{P}_n}$ satisfy the requirements of Theorem 5.3. Let $\mathcal{F}^{(2)}_{\mathbf{P}_n}$ be the collection of all the 2-trees $\mathbf{T}^{(2)}$ obtained from the trees $\mathbf{T} \in \mathcal{F}_{\mathbf{P}_n}$ by the procedure described in the beginning of Section 5.3. Define the corresponding exceptional sets

$$E_n^{(1)} = \Big\{x : \sum_{\mathbf{T}\in\mathcal{F}} 1_{I_{\mathbf{T}}}(x) > \beta\Big\},$$

$$E_n^{(2)} = \bigcup_{\mathbf{T}\in\mathcal{F}^{(2)}} \Big\{x : \|\sum_{\substack{P\in\mathbf{T}\\ |I_P|<2^k}} a_P w_{P_1}(x)\|_{V^r(k)} > \gamma\Big\}.$$

By (6.3) and the fact that $\lambda \le 1$ we get

$$|E_n^{(1)}| \lesssim 2^{-n}\lambda^{-p}|F|.$$

By Theorem 3.7 and the fact that $\lambda \leq 1$, for each $1 < s < \infty$ we get
$$|E_n^{(2)}| \lesssim \gamma^{-s}\sigma^{s-2}|F| \lesssim 2^{-n(s/2-2)}\lambda^{-s(1/2-\epsilon)}|F|.$$
Define
$$E^* := \bigcup_{n\geq\Delta}(E_n^{(1)} \cup E_n^{(2)}).$$
Note that since $\Delta \gtrsim \log_2(\lambda^{-1})$, we have $|E^*| \lesssim \lambda^{-p}|F|$, an estimate which can be seen by using a sufficiently large s.

For each $x \notin E^*$, Theorem 5.3 guarantees that
$$\begin{aligned}
&\Big\| \sum_{\substack{P\in\mathbf{S}_1\\ |I_P|<2^k}} \langle 1_F, w_{P_1}\rangle w_{P_1}(x) 1_{\omega_{P,2}}(\theta)\Big\|_{M_2^*(\theta)}\\
&\qquad \leq \sum_{n\geq\Delta}\Big\| \sum_{\substack{P\in\mathbf{P}_n\\ |I_P|<2^k}} \langle 1_F, w_{P_1}\rangle w_{P_1}(x) 1_{\omega_{P,2}}(\theta)\Big\|_{M_2^*(\theta)}\\
&\qquad \lesssim \sum_{n\geq\Delta} n\Big[2^{(3(r/2-1)-1)n}\lambda^{p(r/2-1)} + 2^{(3(r/2-1)-1/2)n}\lambda^{p(r/2-1)+1/2-\epsilon}\Big]\\
&\qquad \lesssim \lambda^{1-2\epsilon},
\end{aligned}$$
if r is chosen sufficiently close to 2, depending on p and ϵ. This ends the proof of (6.2), and hence the proof of (6.1) in the case $\lambda \leq 1$.

We next focus on proving (6.1) in the case $\lambda > 1$. In this remaining part of the discussion the size will be understood with respect to the function $\lambda^{-1}1_F$. Proposition 3.4 implies that $\mathrm{size}(\mathbf{S}) \lesssim \lambda^{-1}$. Define $\Delta := [-\log_2(\mathrm{size}(\mathbf{S}))]$. Split $\mathbf{S}$ as before, as a disjoint union $\mathbf{S} = \bigcup_{n\geq\Delta}\mathbf{P}_n$, where $\mathrm{size}(\mathbf{P}_n) \leq 2^{-n}$ and each $\mathbf{P}_n$ consists of a family $\mathcal{F}_{\mathbf{P}_n}$ of trees satisfying
$$\sum_{\mathbf{T}\in\mathcal{F}_{\mathbf{P}_n}} |I_{\mathbf{T}}| \lesssim 2^{2n}\lambda^{-2}|F|. \tag{6.4}$$

For each $n \geq \Delta$ define $\sigma := 2^{-n}$, $\beta := 2^{(p+1)n}$ and $\gamma := 2^{-n/2}$. Define also $a_P := \langle \lambda^{-1}1_F, w_{P_1}\rangle$ for each $P \in \mathbf{P}_n$ and note that the collection $\mathbf{P}_n$ together with the coefficients $(a_P)_{P\in\mathbf{P}_n}$ satisfy the requirements of Theorem 5.3. Let $\mathcal{F}_{\mathbf{P}_n}^{(2)}$ the collection of all the 2-trees $\mathbf{T}^{(2)}$ obtained from the trees $\mathbf{T} \in \mathcal{F}_{\mathbf{P}_n}$ by the procedure described in the beginning of the Section 5.3. Define the corresponding exceptional sets
$$E_n^{(1)} = \Big\{x : \sum_{\mathbf{T}\in\mathcal{F}} 1_{I_{\mathbf{T}}}(x) > \beta\Big\},$$
$$E_n^{(2)} = \bigcup_{\mathbf{T}\in\mathcal{F}^{(2)}}\Big\{x : \|\sum_{\substack{P\in\mathbf{T}\\ |I_P|<2^k}} a_P w_{P_1}(x)\|_{V^r(k)} > \gamma\Big\}.$$

By (6.4) and the fact that $\lambda \geq 1$ we get

$$|E_n^{(1)}| \lesssim 2^{-(p-1)n}\lambda^{-2}|F|.$$

By Theorem 3.7 and the fact that $\lambda \geq 1$, for each $1 < s < \infty$ we get

$$|E_n^{(2)}| \lesssim \gamma^{-s}\sigma^{s-2}\lambda^{-2}|F| \lesssim 2^{-n(s/2-2)}\lambda^{-2}|F|.$$

Define

$$E^* := \bigcup_{n \geq \Delta} (E_n^{(1)} \cup E_n^{(2)}).$$

Note that since $\Delta \gtrsim \log_2(\lambda)$, we have

$$|E^*| \lesssim \lambda^{-p}|F|,$$

an estimate which can be seen by using a sufficiently large s.

For each $x \notin E^*$, Theorem 5.3 guarantees that

$$\begin{aligned}
&\Bigg\| \Bigg(\sum_{\substack{P \in \mathbf{S} \\ |I_P| < 2^k}} \langle \lambda^{-1} 1_F, w_{P_1} \rangle w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \Bigg)_{k \in \mathbb{Z}} \Bigg\|_{M_2^*(\theta)} \\
&\qquad \leq \sum_{n \geq \Delta} \Bigg\| \Bigg(\sum_{\substack{P \in \mathbf{P}_n \\ |I_P| < 2^k}} \langle \lambda^{-1} 1_F, w_{P_1} \rangle w_{P_1}(x) 1_{\omega_{P,2}}(\theta) \Bigg)_{k \in \mathbb{Z}} \Bigg\|_{M_2^*(\theta)} \\
&\qquad \lesssim \sum_{n \geq \Delta} n 2^{(p+1)(r/2-1)n} (2^{-n} + 2^{-n/2}) \lesssim 1,
\end{aligned}$$

if r is chosen sufficiently close to 2, depending only on p. This ends the proof of (6.1) in the case $\lambda > 1$.

References

[1] BOURGAIN, J.: Pointwise ergodic theorems for arithmetic sets. *Inst. Hautes Études Sci. Publ. Math.* **69** (1989), 5–45.

[2] BOURGAIN, J.: Return times of dynamical systems *C. R. Acad. Sci. Paris Sér. I Math.* **306** (1988), 483–485.

[3] DEMETER, C., LACEY, M., TAO, T. AND THIELE, C.: Breaking the duality in the return times theorem. *Duke Math. J.* **143** (2008), no. 2, 281–355.

[4] MUSCALU, C., TAO, T. AND THIELE, C.: L^p estimates for the biest. I. The Walsh case. *Math. Ann.* **329** (2004), no. 3, 401–426.

[5] THIELE, C.: *On the Bilinear Hilbert Transform.* Habilitationschrift, 1998.

[6] Thiele, C.: The quartile operator and pointwise convergence of Walsh series. *Trans. Amer. Math. Soc.* **352** (2000), no. 12, 5745–5766.

[7] Thiele, C.: The maximal quartile operator. *Rev. Mat. Iberoamericana* **17** (2001), no. 1, 107–135.

Recibido: 7 de noviembre de 2006
Revisado: 8 de diciembre de 2007

Ciprian Demeter
Indiana University
Department of Mathematics
831 East 3rd St.
Bloomington, IN 47405, USA
demeterc@indiana.edu

Michael Lacey
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332, USA
lacey@math.gatech.edu

Terence Tao
Department of Mathematics, UCLA
Los Angeles CA 90095-1555, USA
tao@math.ucla.edu

Christoph Thiele
Department of Mathematics, UCLA
Los Angeles CA 90095-1555, USA
thiele@math.ucla.edu

The first author was supported by NSF Grant DMS-0556389. The second author was supported by an NSF Grant. The third author was supported by a grant from the Macarthur Foundation. The fourth author was supported by NSF Grant DMS-0400879