Illinois Journal of Mathematics
Volume 51, Number 4, Winter 2007, Pages 1123–1158
S 0019-2082

POINTWISE CONVERGENCE OF THE ERGODIC BILINEAR HILBERT TRANSFORM

CIPRIAN DEMETER

ABSTRACT. We prove that the ergodic bilinear Hilbert transform converges almost everywhere for pairs of bounded functions. We also give a proof along the same lines of Bourgain's analog result for averages.

1. Introduction

Let $\mathbf{X} = (X, \Sigma, m, \tau)$ be a dynamical system, i.e., a complete probability space (X, Σ, m) equipped with an invertible bimeasurable transformation $\tau : X \to X$ such that $m\tau^{-1} = m$. The starting point in this discussion is a result proved by Bourgain for bilinear averages.

THEOREM 1.1 ([3]). *For each $f, g \in L^\infty(X)$ the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} f(\tau^n x) g(\tau^{-n} x) \tag{1.1}$$

converge for almost every x.

Bourgain's method consists of turning the issue of almost everywhere convergence into a quantitative problem regarding multipliers on the torus, which are investigated by using classical Fourier analysis. An important reduction of ergodic theoretic nature in his argument concerns the fact that g can be assumed to be orthogonal to the linear space $L^2(\mathbf{K})$, where $\mathbf{K} = (X, \mathcal{K}, m)$ and $\mathcal{K} \subset \Sigma$ is the σ- algebra generated by the eigenfunctions of τ. This is because the convergence is trivial in the case g is an eigenfunction, as it is easily seen from Birkhoff's ergodic theorem [2]. This reduction has the following consequence for the spectral behavior of any g orthogonal to $L^2(\mathcal{K})$:

$$\lim_{N\to\infty} \sup_{|z|=1} \left| \frac{1}{N}\sum_{n=0}^{N-1} g(\tau^{-n}x) z^n \right| = 0; \tag{1.2}$$

Received January 11, 2006; received in final form June 12, 2006.
2000 *Mathematics Subject Classification.* 37A50, 42B20.

see [1] for a proof of this and of some related results. Using this, Bourgain identifies the limit to be 0 for such a g. More generally,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}f(\tau^n x)g(\tau^{-n}x)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}f(\tau^n x)P_{\mathbf{K}}g(\tau^{-n}x),$$

for each $g\in L^\infty(X)$, where $P_{\mathbf{K}}g$ is the projection of g onto $L^2(\mathbf{K})$. A different way of putting this is to say that the Kronecker algebra $\mathcal{K}$ is a characteristic factor for the almost everywhere convergence of the averages (1.1).

In this paper we will prove the convergence of the ergodic bilinear Hilbert transform.

THEOREM 1.2. *For each $f,g\in L^\infty(X)$ the series*

$$\sum_{n=-N}^{N}{}' \frac{f(\tau^n x)g(\tau^{-n}x)}{n}$$

converges for almost every x.

REMARK 1.3. As a consequence of the above result and of the bilinear maximal inequality for very general kernels in [7], it follows that Theorem 1.2 holds for all $f\in L^p(X)$, $g\in L^q(X)$, whenever $1<p,q\le\infty$ and $\frac{1}{p}+\frac{1}{q}<\frac{3}{2}$.

Bourgain's approach does not seem to be applicable to the context of series, in part due to the fact that the characteristic factors for weighted operators other than the usual averages are much less understood, and probably of less relevance to the essence of the problem. In particular, (1.2) fails if the averages are replaced with the truncations of the ergodic Hilbert transform.

We prove Theorem 1.2 using time-frequency harmonic analysis, and by a similar argument we also give a new proof of Theorem 1.1. Our methods will not perceive the difference between the differentiation and the singular integral versions of the above result, due to a common decomposition of both operators into discrete model sums.

Interestingly, our argument does not appeal to characteristic factors or in general to any concrete spectral analysis. Moreover, only a little ergodic theory is needed in the whole argument, when integration along individual orbits allows us to transfer certain oscillation inequalities from harmonic analysis. However, the structure of the Kronecker factor is deeply rooted into our approach. Since the (linear) exponentials $e^{i\lambda x}$ are the eigenfunctions for rotations on the torus, it is probably the case that their presence in the wave packet decomposition of g is reminiscent of the expansion of $P_{\mathbf{K}}g$ into a basis consisting of eigenfunctions for τ. This also suggests that, perhaps, a time-frequency approach to the similar open questions concerning trilinear averages will involve quadratic exponentials like $e^{i\lambda x^2}$, which are second order eigenfunctions for the rotations on the torus.

Both theorems above will be consequences of the following very general harmonic analysis result, as explained in Section 3. We will use the notation

$$\begin{aligned}\operatorname{Dil}_s^p h(x) &= s^{-1/p} h\left(\frac{x}{s}\right),\\ \operatorname{Mod}_\theta g(x) &= e^{2\pi i\theta x} g(x).\end{aligned}$$

THEOREM 1.4. *Let $K : \mathbf{R} \to \mathbf{R}$ be an L^2 kernel satisfying the following requirements:*

$$\widehat{K} \in C^\infty(\mathbf{R} \setminus \{0\}), \tag{1.3}$$

$$|\widehat{K}(\xi)| \lesssim \min\left\{1, \frac{1}{|\xi|}\right\}, \quad \xi \neq 0, \tag{1.4}$$

$$\left|\frac{d^n}{d\xi^n}\widehat{K}(\xi)\right| \lesssim \frac{1}{|\xi|^n} \min\left\{|\xi|, \frac{1}{|\xi|}\right\}, \quad \xi \neq 0,\ n \geq 1. \tag{1.5}$$

Then for each $d = 2^{1/n}$, $n \in \mathbf{N}$, each $f, g \in L^\infty(\mathbf{R})$ with bounded support and each finite sequence of integers $u_1 < u_2 < \cdots < u_J$,

$$\begin{aligned}&\Bigg\| \Bigg(\sum_{j=1}^{J-1} \sup_{\substack{k\in\mathbf{Z}\\ u_j \leq k < u_{j+1}}} \Bigg| \int f(x+y) g(x-y) \times \\ &\qquad\qquad \times \left(\operatorname{Dil}^1_{d^k} K(y) - \operatorname{Dil}^1_{d^{u_{j+1}}} K(y)\right) dy \Bigg|^2 \Bigg)^{1/2} \Bigg\|_{L_x^{1,\infty}} \\ &\quad \lesssim J^{1/4} \|f\|_{L^2} \|g\|_{L^2},\end{aligned} \tag{1.6}$$

with the implicit constants depending only on n.

REMARK 1.5. Due to the assumptions that f and g are bounded and have bounded support it follows that for each $x \in \mathbf{R}$ the integral

$$\int f(x+y) g(x-y) (\operatorname{Dil}^1_{d^k} K(y) - \operatorname{Dil}^1_{d^{u_{j+1}}} K(y)) dy \tag{1.7}$$

converges absolutely. The same remark applies to the following two theorems.

REMARK 1.6. The proof of this theorem is inspired by ideas from [4], [5] and [6]. The same techniques can extend the theorem to a larger range for p and q and eliminate the dependence on J of the bound in the above inequality. However, its current form suffices for our purposes. Moreover, the argument for a more general result as mentioned would have to be more technical and would require us to revisit most of the main results in [4], making the whole presentation much longer. In particular, one of the advantages of the current approach is that it does not rely on any type of interpolation, being a purely L^2 argument.

Our hope is that by keeping technicalities to a minimum, the whole argument will become easier to follow. The interested reader is certainly referred to [4] for details on some of the results we are quoting here.

As an immediate corollary of Theorem 1.4 we get a particular case of Lacey's inequality for the bilinear maximal function:

COROLLARY 1.7 ([7]). *The following inequality holds for each* $f, g \in L^2(\mathbf{R})$*:*

$$\left\| \sup_{\epsilon>0} \frac{1}{\epsilon} \int_{|y|>\epsilon} |f(x+y)g(x-y)|dy \right\|_{L_x^{1,\infty}} \lesssim \|f\|_2 \|g\|_2.$$

Theorem 1.4 will follow from two distinct results of dyadic analysis. The first one, Theorem 1.8, is the particular case $d = 2$ of the above result and captures the main difficulty of the problem. The second one, Theorem 1.9, is a square function estimate and will be used to control error terms.

To understand better the connection between these three theorems we introduce some notation. Let $x : (0, \infty) \to \mathbf{C}$. Let also $u_1 < \cdots < u_J$ be as in Theorem 1.4 and define $a_1 \leq \cdots \leq a_J$ such that $a_j n \leq u_j < (a_j + 1)n$. Then observe that

$$\left(\sum_{j=1}^{J-1} \sup_{\substack{k \in \mathbf{Z} \\ u_j \leq k < u_{j+1}}} \left| x\left(\frac{k}{n}\right) - x\left(\frac{u_{j+1}}{n}\right) \right|^2 \right)^{1/2}$$
$$\lesssim \sum_{i=0}^{n-1} \left(\sum_{j=1}^{J-1} \sup_{\substack{k \in \mathbf{Z} \\ a_j \leq k < a_{j+1}}} \left| x\left(k + \frac{i}{n}\right) - x\left(a_{j+1} + \frac{i}{n}\right) \right|^2 \right)^{1/2}$$
$$+ \sum_{\substack{i,j=0 \\ i \neq j}}^{n-1} \left(\sum_{k \in \mathbf{Z}} \left| x\left(k + \frac{i}{n}\right) - x\left(k + \frac{j}{n}\right) \right|^2 \right)^{1/2}.$$

Using this inequality and a dilation argument, Theorem 1.4 will follow immediately from the two results below.

THEOREM 1.8. *Let* $K : \mathbf{R} \to \mathbf{R}$ *be an* L^2 *kernel satisfying* (1.3), (1.4) *and* (1.5). *Then for each finite sequence of integers* $u_1 < u_2 < \cdots < u_J$ *and each* $f, g \in L^\infty(\mathbf{R})$ *with bounded support,*

$$\Bigg\| \Bigg(\sum_{j=1}^{J-1} \sup_{\substack{k\in\mathbf{Z}\\ u_j\le k<u_{j+1}}} \Bigg| \int f(x+y)g(x-y)\times \times \big(\mathrm{Dil}^1_{2^k} K(y) - \mathrm{Dil}^1_{2^{u_{j+1}}} K(y)\big)dy \Bigg|^2 \Bigg)^{1/2} \Bigg\|_{L^{1,\infty}_x} \lesssim J^{1/4}\|f\|_{L^2}\|g\|_{L^2},$$

with some universal implicit constant.

THEOREM 1.9. *Let $K:\mathbf{R}\to\mathbf{R}$ be an L^2 kernel satisfying* (1.3), (1.4) *and* (1.5) *and the extra requirement*

$$|\widehat{K}(\xi)| \lesssim |\xi|, \quad \xi \neq 0. \tag{1.8}$$

Then for each $f,g\in L^\infty(\mathbf{R})$ with bounded support,

$$\Bigg\| \Bigg(\sum_{k\in\mathbf{Z}} \Bigg| \int f(x+y)g(x-y)\,\mathrm{Dil}^1_{2^k} K(y)dy \Bigg|^2 \Bigg)^{1/2} \Bigg\|_{L^{1,\infty}_x} \lesssim \|f\|_{L^2}\|g\|_{L^2}, \tag{1.9}$$

with some universal implicit constant.

In Section 3 we indicate how the result of Theorem 1.4 can be transferred to a similar inequality in a dynamical system, and how this implies the convergence in Theorems 1.1 and 1.2. In Section 4 we discretize the operator in Theorem 1.8, while the remaining sections are concerned with proving its boundedness. In the last section we briefly sketch how the same procedure can be applied to prove Theorem 1.9.

We thank Camil Muscalu for pointing out an error in the original argument for Theorem 6.1.

2. Notation

In this section we set up some notation and terminology for the rest of the paper.

If I is an interval, then $c(I)$ denotes the center of I, while cI is the interval with the same center and length c times the length of I. By 1_A we denote the characteristic function of the set $A\subset\mathbf{R}$, while for any dyadic interval I,

$$\chi_I(x) = \left(1+\frac{(x-c(I))^2}{|I|^2}\right)^{-1/2}.$$

A tile P is a rectangle $P=I_P\times\omega_P$ such that I_P is a dyadic interval and $|I_P|\cdot|\omega_P|=1$. A multitile s is a box $s=I_s\times\omega_{s,1}\times\cdots\times\omega_{s,n}$ such that I_s is a dyadic interval, $|\omega_{s,1}|=\cdots=|\omega_{s,n}|$ and $|I_s|\cdot|\omega_{s,1}|=1$.

For each E of finite measure, $X(E)$ will denote the set of all functions supported in E with $\|f\|_\infty \leq 1$, while $X_2(E)$ will denote the L^2 normalized set of all functions supported in E with $\|f\|_\infty \leq |E|^{-1/2}$. Also $Mf(x) = \sup_{r>0} \frac{1}{2r}\int_{x-r}^{x+r} |f|(y)dy$ denotes the classical Hardy-Littlewood maximal function.

The notation $a \lesssim b$ means that $a \leq cb$ for some universal constant c, while $a \sim b$ means that $a \lesssim b$ and $b \lesssim a$. Sometimes we will write $a \lesssim_{\text{parameters}} b$ to indicate that $a \leq cb$ with c depending only on the specified parameters.

3. Pointwise convergence for averages and series

3.1. Bounded oscillation implies convergence. Assume we have a sequence W_k of weighted operators defined on a dynamical system $\mathbf{X} = (X, \Sigma, m, \tau)$ by the formula

$$W_k(f,g)(x) = \sum_{n\in\mathbf{Z}} w_{k,n} f(\tau^n x) g(\tau^{-n} x), \quad k \in \mathbf{N}.$$

LEMMA 3.1. *If for some $f, g \in L^2(X)$*

$$\left\| \left(\sum_{j=1}^{J-1} \sup_{\substack{k\in\mathbf{Z} \\ u_j \leq k < u_{j+1}}} \left| W_k(f,g)(x) - W_{u_{j+1}}(f,g)(x) \right|^2 \right)^{1/2} \right\|_{L^{1,\infty}} \lesssim J^{1/4} \|f\|_{L^2} \|g\|_{L^2} \tag{3.1}$$

uniformly in J and all finite sequences of positive integers $u_1 < u_2 < \cdots < u_J$, then

$$\lim_{k\to\infty} W_k(f,g)(x)$$

exists for almost every $x \in X$.

Proof. To see this, assume for contradiction that the convergence does not hold. It follows that there is a measurable set $X' \subset X$ with $m(X') > 0$ and some $\alpha > 0$, such that for each $x \in X'$

$$\limsup_{k\to\infty} W_k(f,g)(x) - \liminf_{k\to\infty} W_k(f,g)(x) > \alpha.$$

An elementary measure theoretic argument shows that one can then choose a subset $X'' \subseteq X'$ of positive measure and a sequence of positive integers $(u_j)_{j\in\mathbf{N}}$ such that

$$\sup_{u_j \leq k < u_{j+1}} |W_k(f,g)(x) - W_{u_{j+1}}(f,g)(x)| > \alpha,$$

for each $j \in \mathbf{N}$ and for each $x \in X''$. We immediately get that for each J

$$\left\| \left(\sum_{j=1}^{J} \sup_{u_j \le k < u_{j+1}} \left| W_k(f,g)(x) - W_{u_{j+1}}(f,g)(x) \right|^2 \right)^{1/2} \right\|_{1,\infty} \ge J^{1/2} \alpha m(X''),$$

which contradicts inequality (3.1). □

3.2. Proof of Theorem 1.1. Let $M \in \mathbf{N}$ be arbitrary. We apply Theorem 1.4 to a C^∞ kernel K_M satisfying

$$1_{[0,1]} \le K_M \le 1_{[-\frac{1}{M}, 1+\frac{1}{M}]}$$

Then we invoke standard transfer methods like in [4] to get the following corollary.

Corollary 3.2. *For each $d = 2^{1/n}$, $n \in \mathbf{N}$, each finite sequence of positive integers $u_1 < u_2 < \cdots < u_J$ and each $f, g \in L^2(X)$*

$$\left\| \left(\sum_{j=1}^{J-1} \sup_{\substack{k \in \mathbf{Z} \\ u_j \le k < u_{j+1}}} \left| (A_{d^k} - A_{d^{u_{j+1}}})(f,g) + (E_{d^k,M} - E_{d^{u_{j+1}},M})(f,g) \right|^2 \right)^{1/2} \right\|_{1,\infty} \lesssim_{M,n} J^{1/4} \|f\|_{L^2} \|g\|_{L^2},$$

where for each real $r > 1$ we denote

$$A_r(f,g)(x) = \frac{1}{[r]} \sum_{0 \le n \le r} f(\tau^n x) g(\tau^{-n} x),$$

while

$$E_{r,M}(f,g)(x) = \sum_{n \in \mathbf{Z}} w_{r,n,M} f(\tau^n x) g(\tau^{-n} x)$$

is some error term with

$$\sup_{r>1} \sum_{n \in \mathbf{Z}} |w_{r,n,M}| \lesssim M^{-1}. \tag{3.2}$$

By using this and Lemma 3.1 we find that there is a set X_0 of full measure such that for each $M \in \mathbf{N}$, each $d = 2^{1/n}$ and each $x \in X_0$ the following limit exists:

$$\lim_{k \to \infty} (A_{d^k}(f,g)(x) + E_{d^k,M}(f,g)(x)).$$

Fix now some $f, g \in L^\infty(X)$ with $\|f\|_\infty, \|g\|_\infty \le 1$. Based on (3.2) we now get that for almost every $x \in X_0$

$$\begin{aligned}\limsup_{k\to\infty} & A_k(f,g)(x) - \liminf_{k\to\infty} A_k(f,g)(x) \\ &\le \limsup_{k\to\infty}(A_{d^k}(f,g)(x) + E_{d^k,M}(f,g)(x)) \\ &\quad - \liminf_{k\to\infty}(A_{d^k}(f,g)(x) + E_{d^k,M}(f,g)(x)) \\ &\quad + \limsup_{k\to\infty} E_k(f,g)(x) - \liminf_{k\to\infty} E_k(f,g)(x) + 10n^{-1} \\ &\lesssim n^{-1} + M^{-1}.\end{aligned}$$

Since we can take M and n to be as large as we want, it follows that

$$\lim_{k\to\infty} A_k(f,g)(x)$$

exists almost everywhere.

3.3. Proof of Theorem 1.2. Since the Hilbert kernel involved in Theorem 1.2 is not integrable, the route towards proving convergence in this case poses some extra difficulties. As a consequence, we will present a more detailed argument in this case.

The point is again to prove the almost everywhere convergence of the series

$$\sum_{n=-N}^{N}{}' \frac{f(\tau^n x)g(\tau^{-n}x)}{n}$$

along lacunary sequences and then invoke the boundedness of both f and g to get the convergence along the full sequence of positive integers. For simplicity we choose to present the argument in the particular case $N = 2^k$.

Let $M \in \mathbf{N}$ be arbitrary. We apply Theorem 1.4 to a C^∞ kernel K_M satisfying

$$K_M(x) = \frac{1}{x} \ \text{ for } |x| \ge 1,$$
$$|K_M 1_{[-1,1]}| \le 2 \times 1_{1-\frac{1}{M} \le |x| \le 1}.$$

Introduce the discrete kernels $H_{k,M} : \mathbf{R} \to \mathbf{R}$, $k \ge 1$, defined by

$$H_{k,M}(x) = \sum_{-2^k \le i \le 2^k - 1} 1_{[i,i+1)}(x) \frac{1}{2^k} K_M\left(\frac{i}{2^k}\right) + \sum_{i \in \mathbf{Z}\setminus[-2^k, 2^k-1]} 1_{[i,i+1)}(x)\frac{1}{i}.$$

Let $k_1 < k_2 < \cdots < k_J$ be an arbitrary sequence of positive integers. The fact that the terms of the sequence are positive is a crucial fact, exploited in the following. Indeed, we note that

$$|H_{k,M}(y) - \mathrm{Dil}^1_{2^k} K_M(y)| \lesssim_M \begin{cases} \frac{1}{2^{2k}}, & |y| \le 2^k, \\ \frac{1}{y^2}, & |y| \ge 2^k. \end{cases}$$

From the boundedness of the maximal averages in Corollary 1.7 we deduce that

$$\begin{aligned} &\left\| \left(\sum_{k\ge 1} \left| \int f(x+y)g(x-y)(H_{k,M}(y) - \mathrm{Dil}^1_{2^k} K_M(y))dy \right|^2 \right)^{1/2} \right\|_{1,\infty} \\ &\quad \lesssim_M \| \sup_{\epsilon>0} \frac{1}{\epsilon} \int_{|y|>\epsilon} |f(x+y)g(x-y)|dy\|_{1,\infty} \\ &\quad \lesssim \|f\|_2 \|g\|_2. \end{aligned}$$

As a consequence of this and Theorem 1.4 applied to K_M, we get that

$$\begin{aligned} (3.3)\quad &\left\| \left(\sum_{j=1}^{J-1} \sup_{k_j \le k < k_{j+1}} \left| \int f(x+y)g(x-y)\big(H_{k,M}(y) \right.\right.\right. \\ &\qquad\qquad \left.\left.\left. - H_{k_{j+1},M}(y)\big)dy \right|^2 \right)^{1/2} \right\|_{1,\infty} \\ &\quad \lesssim_M J^{1/4} \|f\|_2 \|g\|_2. \end{aligned}$$

The next step consists of transferring (3.3) to the integers. By considering functions like $f : \mathbf{R} \to \mathbf{R}$ with

$$f(x) = \begin{cases} \phi([x]), & [x] + \frac{1}{4} \le x \le [x] + \frac{1}{2}, \\ 0, & \text{otherwise,} \end{cases}$$

and $g : \mathbf{R} \to \mathbf{R}$ with

$$g(x) = \begin{cases} \psi([x]), & [x] + \frac{1}{4} \le x \le [x] + \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

we get that for each $\phi, \psi : \mathbf{Z} \to \mathbf{Z}$ with finite support

$$\begin{aligned} (3.4)\quad &\left\| \left(\sum_{j=1}^{J-1} \sup_{k_j \le k < k_{j+1}} \left| \sum_{b\in\mathbf{Z}} \phi(a+b)\psi(a-b) \times \right.\right.\right. \\ &\qquad\qquad \left.\left.\left. \times \big(H_{k,M}(b) - H_{k_{j+1},M}(b)\big) \right|^2 \right)^{1/2} \right\|_{l^{1,\infty}(\mathbf{Z})} \\ &\quad \lesssim_M J^{1/4} \|\phi\|_{l^2(\mathbf{Z})} \|\psi\|_{l^2(\mathbf{Z})}. \end{aligned}$$

For each $k \geq 1$ introduce the kernels $A_{k,M} : \mathbf{Z} \to \mathbf{Z}$ and $S_{k,M} : \mathbf{Z} \to \mathbf{Z}$ defined by

$$A_{k,M}(i) = \begin{cases} H_{k,M}(i), & -2^k \leq i \leq 2^k, \\ 0, & \text{otherwise}, \end{cases}$$

$$S_{k,M}(i) = \begin{cases} \frac{1}{i}, & -2^k \leq i \leq 2^k,\ i \neq 0, \\ 0, & \text{otherwise}, \end{cases}$$

and note that for each $k < k'$

$$H_{k,M} - H_{k',M} = O_{k,M} - O_{k',M} := (A_{k,M} - S_{k,M}) - (A_{k',M} - S_{k',M}).$$

Thus (3.4) gives

$$\Bigg\| \Bigg(\sum_{j=1}^{J-1} \sup_{k_j \leq k < k_{j+1}} \Bigg| \sum_{b \in \mathbf{Z}} \phi(a+b)\psi(a-b) \times \times \big(O_{k,M}(b) - O_{k_{j+1},M}(b) \big) \Bigg|^2 \Bigg)^{1/2} \Bigg\|_{l^{1,\infty}(\mathbf{Z})} \lesssim_M J^{1/4} \|\phi\|_{l^2(\mathbf{Z})} \|\psi\|_{l^2(\mathbf{Z})}.$$

Standard transfer to a dynamical system $\mathbf{X} = (X, \Sigma, \mu, \tau)$, as in [4], leads to

$$\Bigg\| \Bigg(\sum_{j=1}^{J-1} \sup_{k_j \leq k < k_{j+1}} \Bigg| \sum_{n \in \mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) \times \times \big(O_{k,M}(n) - O_{k_{j+1},M}(n) \big) \Bigg|^2 \Bigg)^{1/2} \Bigg\|_{1,\infty} \lesssim_M J^{1/4} \|f\|_2 \|g\|_2. \tag{3.5}$$

By invoking Lemma 3.1 it follows that if $\|f\|_{L^\infty}, \|g\|_{L^\infty} \leq 1$, then

$$\lim_{k \to \infty} \sum_{n \in \mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) O_{k,M}(n)$$

exists for almost every $x \in X$. Finally,

$$\begin{aligned}
&\limsup_{k\to\infty} \sum_{n\in\mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) S_{k,M}(n) \\
&\qquad - \liminf_{k\to\infty} \sum_{n\in\mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) S_{k,M}(n) \\
&\quad \le \limsup_{k\to\infty} \sum_{n\in\mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) O_{k,M}(n) \\
&\qquad - \liminf_{k\to\infty} \sum_{n\in\mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) O_{k,M}(n) \\
&\qquad + \limsup_{k\to\infty} \sum_{n\in\mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) A_{k,M}(n) \\
&\qquad - \liminf_{k\to\infty} \sum_{n\in\mathbf{Z}} f(\tau^n x) g(\tau^{-n} x) A_{k,M}(n) \\
&\quad \lesssim CM^{-1}.
\end{aligned}$$

Since this holds for arbitrary M, we get that

$$\lim_{k\to\infty} {\sum_{n=-2^k}^{2^k}}' \frac{f(\tau^n x) g(\tau^{-n} x)}{n}$$

exists almost everywhere.

4. Discretization

DEFINITION 4.1. A set $\mathcal{G}'$ of (not necessarily dyadic) intervals is called a grid if

- $I, I' \in \mathcal{G}'$ and $|I'| \le |I|$ imply that either $I' \subset I$ or $I \cap I' = \emptyset$;
- $I \in \mathcal{G}'$ implies that $|I| = 2^k$, for some $k \in \mathbf{Z}$.

The standard dyadic grid is

$$\mathcal{S} = \{\, [2^i l, 2^i (l+1)] : \; i, l \in \mathbf{Z}\}. \tag{4.1}$$

We will also be interested in a more general type of grid. For each odd integer $N \ge 3$, each $0 \le t \le N-2$ and $0 \le L \le N-1$, the collection

$$\mathcal{G}_{N,t,L} := \left\{ \left[2^i \left(l + \frac{L}{N} \right), 2^i \left(l + \frac{L}{N} + 1 \right) \right] \; : \; i \equiv t \pmod{N-1}, \; l \in \mathbf{Z} \right\}$$

is a grid, as easily follows from the fact that $2^{N-1} \equiv 1 \pmod{N}$. We note that for each fixed N the grids $\mathcal{G}_{N,t,L}$ are pairwise disjoint, for $0 \le t \le N-2$ and $0 \le L \le N-1$.

For each L^2 kernel K, each $f, g \in L^\infty(\mathbf{R})$ with bounded support and each sequence $\mathbf{U} = (u_j)_{j=1}^J$ define the quantity

$$O_{K,\mathbf{U}}(f,g)(x) = \left(\sum_{j=1}^{J-1} \sup_{\substack{k\in\mathbf{Z}\\ u_j \le k < u_{j+1}}} \left| \int f(x+y)g(x-y) \big(\mathrm{Dil}^1_{2^k} K(y) - \mathrm{Dil}^1_{2^{u_{j+1}}} K(y)\big) dy \right|^2 \right)^{1/2}.$$

In the following we indicate how to discretize it. Choose $\eta : \mathbf{R} \to \mathbf{R}$ such that $\widehat{\eta}$ is a $C^\infty(\mathbf{R}\setminus\{0\})$ function which equals $\lim_{\xi\to 0^+} \widehat{K}(\xi)$ on $\big(0, \frac{1}{2}\big]$, $\lim_{\xi\to 0^-} \widehat{K}(\xi)$ on $\big[-\frac{1}{2}, 0\big)$, and 0 outside $[-1,1]$. The two limits exist due to the fact that $|\frac{d}{d\xi} K(\xi)| \lesssim 1$ for $\xi \neq 0$. It suffices to prove

$$\|O_{\eta,\mathbf{U}}(f,g)\|_{1,\infty} \lesssim J^{1/4} \|f\|_2 \|g\|_2, \tag{4.2}$$

$$\|O_{K-\eta,\mathbf{U}}(f,g)\|_{1,\infty} \lesssim J^{1/4} \|f\|_2 \|g\|_2. \tag{4.3}$$

The proofs for the above inequalities will follow from a more general principle, as explained below. The crucial property of the multiplier $\widehat{K-\eta}$ that will be used later is the following:

$$\left| \frac{d^n}{d\xi^n} \widehat{K-\eta}(\xi) \right| \lesssim \frac{1}{|\xi|^n} \min\left\{ |\xi|, \frac{1}{|\xi|} \right\}, \quad n \geq 0. \tag{4.4}$$

Note that the additional inequality $|\widehat{K-\eta}(\xi)| \lesssim |\xi|$ for $\xi \neq 0$ is a consequence of the fact that $|\frac{d}{d\xi} \widehat{K}(\xi)| \lesssim 1$ for $\xi \neq 0$. Write

$$\widehat{K-\eta}(\xi) = \sum_{j=-\infty}^{\infty} \widehat{K-\eta}(\xi) q\left(\frac{\xi}{2^j}\right), \tag{4.5}$$

where q is some Schwartz function supported in the annulus $\frac{1}{2} < |\xi| < 2$ such that

$$\sum_{j\in\mathbf{Z}} q\left(\frac{\xi}{2^j}\right) = 1, \quad \xi \neq 0.$$

Define $g_j = \widehat{K-\eta}(\xi) q(\xi/2^j)$. As a consequence of (4.4) we get that both the function η and each function $\mathrm{Dil}^1_{2^j} \check{g}_j$ satisfy

$$|\eta(x)| \lesssim_M \frac{1}{(1+|x|)^M},$$

$$\left| \frac{1}{2^j} \check{g}_j\left(\frac{x}{2^j}\right) \right| \lesssim_M \frac{2^{-|j|}}{(1+|x|)^M},$$

for all $M \geq 0$ and $x \in \mathbf{R}$, uniformly in $j \in \mathbf{Z}$. Moreover, each of the above functions has the Fourier transform constant on both $\{0 < \xi \leq \frac{1}{2}\}$ and $\{-\frac{1}{2} \leq \xi < 0\}$, as well as on $\{|\xi| \geq 2\}$.

For each $j \in \mathbf{Z}$ define the shifted sequence $\mathbf{U}'_j = \mathbf{U} - j$. Since the operators $O_{\check{g}_j, \mathbf{U}}$ and $O_{\mathrm{Dil}^1_{2^j} \check{g}_j, \mathbf{U}'_j}$ coincide, inequalities (4.2) and (4.3) will immediately follow if we prove that

$$\|O_{\check{\theta}, \mathbf{U}}(f,g)\|_{1,\infty} \lesssim J^{1/4} \|f\|_2 \|g\|_2, \tag{4.6}$$

uniformly in all $C^\infty(\mathbf{R} \setminus \{0\})$ functions θ which are constant on both $\{0 < \xi \leq \frac{1}{2}\}$ and $\{-\frac{1}{2} \leq \xi < 0\}$, as well as on $\{|\xi| \geq 2\}$, and which satisfy

$$|\check{\theta}(x)| \lesssim_M \frac{1}{(1+|x|)^M} \tag{4.7}$$

for all $M \geq 0$ and $x \in \mathbf{R}$.

By a similar reduction we can assume instead that θ is constant on $\{|x| \leq 1000\}$ and on $\{|x| \geq 4000\}$. This extra assumption will serve later for the purpose of creating disjointness of some sort between multitiles. Note that

$$\theta(2^k \xi) = \sum_{i \geq k} \theta_i(\xi),$$

where $\theta_i(\xi) = \theta(2^i \xi) - \theta(2^{i+1}\xi)$ is supported in the annulus $\{500 \times 2^{-i} \leq |\xi| \leq 4000 \times 2^{-i}\}$.

Pick a Schwartz function ψ such that $\widehat{\psi}$ is supported in $[0, \frac{2}{5}]$ and satisfies the following property for every $\xi \in \mathbf{R}$:

$$\sum_{l \in \mathbf{Z}} \left| \widehat{\psi}\left(\xi - \frac{l}{5}\right)\right|^2 = 1.$$

For each scale i use the following expansion for both f and g, valid in every L^p norm, $1 < p < \infty$:

$$f = \sum_{m,l \in \mathbf{Z}} \langle f, \psi_{i,m,\frac{l}{5}} \rangle \psi_{i,m,\frac{l}{5}},$$
$$g = \sum_{m,l \in \mathbf{Z}} \langle g, \psi_{i,m,\frac{l}{5}} \rangle \psi_{i,m,\frac{l}{5}},$$

where

$$\psi_{i,m,l}(x) = 2^{-i/2} \psi(2^{-i}x - m) e^{2\pi i 2^{-i} x l}.$$

The fundamental properties of $\psi_{i,m,l}$ that will be used in the following are

$$\left|\frac{d^n}{dx^n}\operatorname{Mod}_{-l2^{-i}}\psi_{i,m,l}(x)\right| \lesssim_M 2^{(-1/2-n)i}\frac{1}{(1+|2^{-i}x-m|)^M},\quad M\geq 0, \tag{4.8}$$

$$\operatorname{supp}\widehat{\psi}_{i,m,l}\subset\left[\frac{l}{5}2^{-i},\left(\frac{l}{5}+1\right)2^{-i}\right]. \tag{4.9}$$

The above properties can be summarized by saying that the tile $[m2^i,(m+1)2^i]\times[\frac{l}{5}2^{-i},(\frac{l}{5}+1)2^{-i}]$ is a Heisenberg box for $\psi_{i,m,l}$.

With the notation

$$\varphi_{i,\vec{m},\vec{l}}(x):=2^{i/2}\int\psi_{i,m_1,\frac{l_1}{5}}(x+y)\psi_{i,m_2,\frac{l_2}{5}}(x-y)\check{\theta}_i(y)dy$$

it follows that

$$O_{\check{\theta},\mathbf{U}}(f,g)(x)=\left(\sum_{j=1}^{J-1}\sup_{\substack{k\in\mathbf{Z}\\k_j\leq k<k_{j+1}}}\left|\sum_{i=k}^{k_{j+1}-1}\sum_{\vec{m},\vec{l}\in\mathbf{Z}^2}2^{-i/2}\langle f,\psi_{i,m_1,\frac{l_1}{5}}\rangle\langle g,\psi_{i,m_2,\frac{l_2}{5}}\rangle\varphi_{i,\vec{m},\vec{l}}(x)\right|^2\right)^{1/2}.$$

The triangle inequality then shows that Theorem 1.8 follows once we prove that

$$\left\|\sum_{m\geq 0}\left(\sum_{j=1}^{J-1}\sup_{\substack{k\in\mathbf{Z}\\u_j\leq k<u_{j+1}}}\left|\sum_{i=k}^{u_{j+1}-1}\sum_{\substack{\vec{m},\vec{l}\in\mathbf{Z}^2\\|m_1-m_2|=m}}2^{-i/2}\langle f,\psi_{i,m_1,\frac{l_1}{5}}\rangle\langle g,\psi_{i,m_2,\frac{l_2}{5}}\rangle\varphi_{i,\vec{m},\vec{l}}\right|^2\right)^{1/2}\right\|_{1,\infty}\lesssim J^{1/4}\|f\|_2\|g\|_2. \tag{4.10}$$

The computations that follow are meant to reveal the decay and localization of the functions $\varphi_{i,\vec{m},\vec{l}}$.

We first observe that since

$$\widehat{\varphi}_{i,\vec{m},\vec{l}}(\xi)=\int\widehat{\psi}_{i,m_1,\frac{l_1}{5}}(\eta)\widehat{\psi}_{i,m_2,\frac{l_2}{5}}(\xi-\eta)\theta(2\eta-\xi)d\eta,$$

it turns out that

$$\operatorname{supp}\widehat{\varphi}_{i,\vec{m},\vec{l}}\subseteq\operatorname{supp}\widehat{\psi}_{i,m_1,l_1}+\operatorname{supp}\widehat{\psi}_{i,m_2,l_2}\subset\left[\frac{l_3}{5}2^{-i},\left(\frac{l_3}{5}+1\right)2^{-i}\right], \tag{4.11}$$

where from now on we denote

$$l_3=l_1+l_2. \tag{4.12}$$

We next observe that for each $M \geq 0$

$$|2^{i/2}\varphi_{i,\vec{m},\vec{l}}(2^i x)| \lesssim \int |\psi(x+y-m_1)\psi(x-y-m_2)\check{\theta}(y)|dy$$
$$\lesssim_M \frac{1}{(1+|x-m_1|)^M(1+|x-m_2|)^M}$$
$$\lesssim_M \frac{1}{(1+|m_1-m_2|)^M(1+|x-\frac{m_1+m_2}{2}|)^M}.$$

A similar estimate holds for all derivatives, and we conclude that

$$\left|\frac{d^n}{dx^n}\operatorname{Mod}_{-\frac{l_3}{5}2^{-i}}\varphi_{i,\vec{m},\vec{l}}(x)\right| \lesssim_M 2^{(-1/2-n)i}\frac{1}{(1+|m_1-m_2|)^M(1+|2^{-i}x-\frac{m_1+m_2}{2}|)^M}, \tag{4.13}$$

for each $n, M \geq 0$.

We thus see that $\varphi_{i,\vec{m},\vec{l}}$ satisfies the same type of properties as $\psi_{i,m,l}$, with some extra uniform decay in $m = |m_1 - m_2|$. In particular, if $I_{i,\vec{m}}$ is one of the (at most two) dyadic intervals of length 2^i which contains $\frac{m_1+m_2}{2}$, then $I_{i,\vec{m}} \times [\frac{l_3}{5}2^{-i}, \left(\frac{l_3}{5}+1\right)2^{-i}]$ is certainly a Heisenberg box for $\varphi_{i,\vec{m},\vec{l}}$.

Finally, since

$$\varphi_{i,\vec{m},\vec{l}}(x) = \int \widehat{\psi}_{i,m_1,\frac{l_1}{5}}(\xi_1)\widehat{\psi}_{i,m_2,\frac{l_2}{5}}(\xi_2)\theta_i(\xi_2-\xi_1)e^{2\pi i x(\xi_1+\xi_2)}d\xi_1 d\xi_2,$$

it follows that in order for $\varphi_{i,\vec{m},\vec{l}}$ to not be the zero function we must have

$$10^2 < \left|\frac{l_1}{5} - \frac{l_2}{5}\right| < 10^4. \tag{4.14}$$

The next reduction concerns the fact that there is a finite universal set $E \in \mathbf{Z}$, such that every l_1, l_2, l_3 satisfying (4.12) and (4.14) will also satisfy the following:

$$l_2 = l_1 + e \tag{4.15}$$

for some $e \in E$. We will restrict the summation in (4.10) to those vectors $\vec{l}$ satisfying (4.15) for some fixed $e \in E$.

For each $\vec{l} \in \mathbf{Z}^3$ and each $i \in \mathbf{Z}$ define the cubes $Q_{\vec{l},i} = \prod_{j=1}^3 [\frac{l_j}{5}2^{-i}, (\frac{l_j}{5}+1)2^{-i}]$. Note that every frequency interval $[2^{-i}\frac{l_j}{5}, 2^{-i}(\frac{l_j}{5}+1)]$ of a cube $Q_{\vec{l},i}$ belongs to one of the grids $\mathcal{G}_{5,t,L}$, $0 \leq t \leq 3$, $0 \leq L \leq 4$. This allows us to further restrict the summation in (4.10) to those $\vec{l}$ and i for which each interval $[2^{-i}\frac{l_j}{5}, 2^{-i}(\frac{l_j}{5}+1)]$ is in a fixed grid depending on j.

Denote by $\mathbf{D}'$ the union over all $i \in \mathbf{Z}$ and all $l_1, l_2, l_3 \in \mathbf{Z}$ which are subject to all the indicated restrictions, of the set of all the cubes $Q_{\vec{l},i}$. For each $m \geq 0$

introduce the set of generalized multitiles $\mathbf{P}_m$ to be

$$\Big\{P=\prod_{j=1}^{3} I_{P,j}\times Q_{\vec{l},i}:\ I_{P,j}=[m_j2^i,(m_j+1)2^i],\ j\le 2,\\ I_{P,3}=I_{i,\vec{m}},\ Q_{\vec{l},i}\in\mathbf{D'},\ |m_1-m_2|=m\Big\}.$$

Each such multitile can be thought of as the product of 3 tiles $P_j:=I_{P,j}\times [l_j2^{-i},(l_j+1)2^{-i}]$. For each multitile P as above define $\psi_{P,j}=\psi_{i,m_j,l_j}$ for $j=1,2$ and $\psi_{P,3}=\varphi_{i,\vec{m},\vec{l}}$. Thus (4.10) will follow if we prove that

$$\Bigg\|\Bigg(\sum_{j=1}^{J-1}\sup_{\substack{k\in\mathbf{Z}\\ u_j\le k<u_{j+1}}}\Bigg|\sum_{\substack{P\in\mathbf{P}_m\\ 2^k\le|I_P|<2^{u_{j+1}-1}}}|I_P|^{-1/2}\langle f,\psi_{P,1}\rangle\langle g,\psi_{P,2}\rangle\psi_{P,3}\Bigg|^2\Bigg)^{1/2}\Bigg\|_{1,\infty}\\ \lesssim(1+m)^{-2}J^{1/4}\|f\|_2\|g\|_2,$$

where $|I_P|$ is defined as the common value of all $|I_{P,j}|$.

As will easily follow from our later analysis, it is enough to prove the above for $m=0$. The extra decay in m for the other terms will be a consequence of the extra decay in (4.13). The choice of working with $m=0$ will simplify the notation, not the argument. Indeed, this implies that for each P, $I_{P,1}=I_{P,2}=I_{P,3}$.

A last harmless reduction consists of sparsifying the set of the dyadic intervals in the grids. More precisely, we will assume that if for two time intervals I,I' we have $\frac{|I'|}{|I|}>1$, then $\frac{|I'|}{|I|}\ge 2^{\Delta}$, where Δ is a sufficiently large constant to be chosen later ($\Delta=1000$ will certainly suffice). Moreover, we can also assume that two frequency intervals of equal length are separated by at least 10 times their length.

We now summarize all the various reductions made so far in the following theorem, which implies Theorem 1.8.

THEOREM 4.2. *Let $\mathcal{G}$, $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ be four grids with $\mathcal{G}$ satisfying*

$$\mathcal{G}\subset\mathcal{S}, \tag{4.16}$$

$$I,I'\in\mathcal{G}\Rightarrow\max\{|I||I'|^{-1},|I'||I|^{-1}\}\ge 2^{\Delta}, \tag{4.17}$$

$$\omega,\omega'\in\mathcal{G}_j,\ |\omega|=|\omega'|\Rightarrow\operatorname{dist}(|\omega|,|\omega'|)\ge 10|\omega|. \tag{4.18}$$

Let e be a number with $10^2\le|e|\le 10^5$ and define

$$\mathbf{D}=\Bigg\{\prod_{j=1}^{3}\Big[\frac{l_j}{5}2^i,\Big(\frac{l_j}{5}+1\Big)2^i\Big]\in\prod_{j=1}^{3}\mathcal{G}_j:\ l_2=l_1+e,\ l_3=l_1+l_2\Bigg\}.$$

Define also the set of multitiles

$$\mathbf{S} = \left\{ s = I_s \times Q_s : I_s \in \mathcal{G},\ \ Q_s \in \mathbf{D} \textit{ with sidelength } \frac{1}{|I_s|} \right\}.$$

Assume that each multitile $s = I_s \times \prod_{j=1}^{3} \omega_{s,j} \in \mathbf{S}$ *is associated with three functions* $(\psi_{s,j})_{j=1}^{3}$ *satisfying*

$$\left| \frac{d^n}{dx^n} \operatorname{Mod}_{-c(\omega_{s,j})} \psi_{s,j}(x) \right| \lesssim_{n,M} |I_s|^{-1/2-n} \chi_{I_s}^{M}(x), \quad n, M \geq 0, \tag{4.19}$$

$$\operatorname{supp} \widehat{\psi}_{s,j} \subset \omega_{s,j}, \tag{4.20}$$

for each $j = 1, 2, 3$.

Then for each $f, g \in L^2(\mathbf{R})$ *and each finite sequence of integers* $\mathbf{U} := u_1 < u_2 < \cdots < u_J$ *we have the estimate*

(4.21)
$$\left\| \left(\sum_{j=1}^{J-1} \sup_{\substack{k \in \mathbf{Z} \\ u_j \leq k < u_{j+1}}} \left| \sum_{\substack{s \in \mathbf{S} \\ 2^k \leq |I_s| < 2^{u_{j+1}}}} |I_s|^{-1/2} \langle f, \psi_{s,1} \rangle \langle g, \psi_{s,2} \rangle \psi_{s,3} \right|^2 \right)^{1/2} \right\|_{1,\infty} \lesssim J^{1/4} \|f\|_2 \|g\|_2,$$

with the implicit constant depending only on the implicit constants in (4.19).

For each $1 \leq j \leq J-1$ let $\kappa_j : \mathbf{R} \to \{u_j, u_j + 1, \ldots, u_{j+1} - 1\}$ be some arbitrary stopping time. For each $s \in \mathbf{S}$ with $2^{u_1} \leq |I_s| < 2^{u_J}$ define

$$\phi_{s,j}(x) = \psi_{s,j}(x)$$

for $j \in \{1, 2\}$ and

$$\phi_{s,3}(x) = \psi_{s,3} 1_{2^{\kappa_j(x)} \leq |I_s| < 2^{u_{j+1}}}(x),$$

where j is the unique integer such that $2^{u_j} \leq |I_s| < 2^{u_{j+1}}$. An equivalent formulation for (4.21) that we will sometimes find easier to handle is

$$\left\| \left(\sum_{j=1}^{J-1} \left| \sum_{\substack{s \in \mathbf{S} \\ 2^{u_j} \leq |I_s| < 2^{u_{j+1}}}} |I_s|^{-1/2} \langle f, \phi_{s,1} \rangle \langle g, \phi_{s,2} \rangle \phi_{s,3} \right|^2 \right)^{1/2} \right\|_{1,\infty} \lesssim J^{1/4} \|f\|_2 \|g\|_2. \tag{4.22}$$

From now on we will fix e, Δ, $\mathbf{S}$, the wave packets $\psi_{s,j}$, the sequence $\mathbf{U}$ and the stopping times κ_j, and we will implicitly assume that for each $s \in \mathbf{S}$ we have $2^{u_1} \leq |I_s| < 2^{u_J}$.

5. The combinatorics of the multitiles

In this section we start by defining a relation of order between multitiles. This will alow us to split $\mathbf{S}$ into structured collections like trees and forests. The model sum restricted to each tree is essentially a Littlewood-Paley dyadic decomposition modulated with a frequency from the frequency interval of the top of the tree. The estimation of the model sum restricted to each such tree involves classical Calderon-Zygmund theory. This will be seen in Section 6. The modern time-frequency side of the whole approach manifests itself in the fact that $\mathbf{S}$ consists of many trees, modulated with possibly different frequencies. The goal of this section is to prove that the model sums corresponding to distinct trees are almost orthogonal, a principle quantified in various Bessel type inequalities. The almost orthogonality will follow if the trees are selected to be strongly disjoint, a combinatorial property which, as the name suggests, is stronger than mere disjointness.

Definition 5.1. For two multitiles $s, s' \in \mathbf{S}$ we write

- $s'_j < s_j$ if $I_{s'} \subsetneq I_s$ and $3\omega_{s,j} \subsetneq 3\omega_{s',j}$;
- $s'_j \leq s_j$ if $s'_j < s_j$ or $s_j = s'_j$;
- $s'_j \lesssim s_j$ if $I_{s'} \subseteq I_s$ and $\omega_{s,j} \subset 10e\omega_{s',j}$;
- $s'_j \lesssim' s_j$ if $s'_j \lesssim s_j$ and $10\omega_{s,j} \cap 10\omega_{s',j} = \emptyset$.

Lemma 5.2. *Given any two multitiles $s, s' \in \mathbf{S}$ such that $s'_i < s_i$ for some $i \in \{1, 2, 3\}$, it follows that $s'_j \lesssim' s_j$ for each $j \in \{1, 2, 3\} \setminus \{i\}$.*

Proof. We illustrate the proof on a particular case which is otherwise representative for the general argument. Consider some $Q_s = \prod_{j=1}^{3}[\frac{l_j}{5}2^i, (\frac{l_j}{5}+1)2^i]$, $Q_{s'} = \prod_{j=1}^{3}[\frac{l'_j}{5}2^{i'}, (\frac{l'_j}{5}+1)2^{i'}]$ and assume that $s'_1 < s_1$. This implies that

$$|l'_1 - 2^{i-i'}l_1| \leq 15. \tag{5.1}$$

Assume now for contradiction that $\omega_{s,2} \nsubseteq 10e\omega_{s',2}$. This in turn implies that

$$|l'_2 - 2^{i-i'}l_2| > 10e. \tag{5.2}$$

Inequalities (5.1) and (5.2) together with the fact that $l_2 - l_1 = l'_2 - l'_1 = e$ will immediately lead to the contradiction.

Assume next for contradiction that $10\omega_{s,2} \cap 10\omega_{s',2} \neq \emptyset$. This in turn implies that

$$|l'_2 - 2^{i-i'}l_2| \leq 50. \tag{5.3}$$

Inequalities (5.1) and (5.3) together with the fact that $l_2 - l_1 = l'_2 - l'_1 = e$ will immediately lead to the contradiction. □

Let $i, j \in \{1, 2, 3\}$.

Definition 5.3. A collection of multitiles $\mathbf{T} \subset \mathbf{S}$ is called an i-tree with top $T \in \mathbf{S}$ if

$$s_i < T_i \text{ for each } s \in \mathbf{T} \setminus \{T\}.$$

We say that $\mathbf{T}$ is a tree if it is an i-tree for some i. We call the tree $\mathbf{T}$ j-lacunary if

$$s_j \lesssim' T_j \text{ for each } s \in \mathbf{T} \setminus \{T\}.$$

Remark 5.4. A tree does not necessarily contain its top. By Lemma 5.2 we know that a tree is j-lacunary if and only if it is an i-tree for some $i \neq j$. Note also that by (4.18) each tree can contain at most one multitile with a given time interval.

A very useful tool for proving estimates for a single tree $\mathbf{T}$ is its size, a quantity which encodes the BMO properties of the model sum associated with $\mathbf{T}$.

Definition 5.5. Consider some $j \in \{1, 2, 3\}$ and a finite subset of multitiles $\mathbf{S}' \subset \mathbf{S}$. Assume that each $s \in \mathbf{S}'$ is associated with a function $F_s : \mathbf{R} \to \mathbf{C}$. Define the j-size of $\mathbf{S}'$ relative to the collection (F_s) by the formula

$$\mathrm{size}_j(\mathbf{S}') = \sup_{\mathbf{T}} \left(\frac{1}{|I_T|} \sum_{s \in \mathbf{T}} |\langle F_s, \phi_{s,j} \rangle|^2 \right)^{1/2},$$

where the supremum is taken over all the trees $\mathbf{T} \in \mathbf{S}'$ of lacunary type j.

To simplify notation, we will not add (F_s) as a superscript of size, since (F_s) will always be clear from the context. Actually the 1-size and the 2-size will always be understood with respect to the functions f and g, respectively.

Note that size is a monotone function with respect to $\mathbf{S}'$. The following lemma shows how to estimate tree paraproducts by using the size.

Lemma 5.6. *If* $\mathbf{T}$ *is a tree, then*

$$\sum_{s \in \mathbf{T}} |I_s|^{-1/2} \prod_{i=1}^{3} |\langle F_s, \phi_{s,i} \rangle| \lesssim |I_T| \prod_{i=1}^{3} \mathrm{size}_i(\mathbf{T}).$$

Proof. Assume $\mathbf{T}$ is an i-tree. Apply l_2 estimates for the terms corresponding to $j, k \in \{1, 2, 3\} \setminus \{i\}$ and l_∞ for the terms corresponding to i. □

Definition 5.7. Let $j \in \{1, 2, 3\}$. Two trees $\mathbf{T}$ and $\mathbf{T}'$ with tops T and T' are said to be strongly j-disjoint if

- $s_j \cap s'_j = \emptyset$ for each $s \in \mathbf{T} \cup \{T\}, s' \in \mathbf{T}' \cup \{T'\}$;

- whenever $s \in \mathbf{T} \cup \{T\}, s' \in \mathbf{T}' \cup \{T'\}$ are such that $\omega_{s,j} \subsetneq \omega_{s',j}$, then one has $I_T \cap I_{s'} = \emptyset$, and similarly with $\mathbf{T}$ and $\mathbf{T}'$ reversed.

A collection of trees is called mutually strongly j-disjoint if any two trees in the collection are strongly j-disjoint and each tree is j-lacunary.

REMARK 5.8. Each subset $\mathbf{S_T} \subset \mathbf{T}$ of an i-tree $\mathbf{T}$ can be decomposed in a unique way as the disjoint union of i-trees $\mathbf{T}'$ containing their tops

$$\mathbf{S_T} = \bigcup_{\mathbf{T}' \in D(\mathbf{S_T})} \mathbf{T}',$$

such that these tops have pairwise disjoint time intervals. Precisely, the tops of the trees in the collection $D(\mathbf{S_T})$ will be the maximal multitiles in $\mathbf{S_T}$ with respect to the order $<$. If one has a collection $\mathcal{F}$ of mutually strongly j-disjoint trees $\mathbf{T}$ and each $\mathbf{T}$ is associated with a subset $\mathbf{S_T} \subset \mathbf{T}$, then the decomposition

$$\mathcal{F} = \bigcup_{\mathbf{T} \in \mathcal{F}} \bigcup_{\mathbf{T}' \in D(\mathbf{S_T})}$$

gives rise to a family of mutually strongly j-disjoint trees.

The concept of strong disjointness is the key ingredient behind the phenomenon of almost orthogonality responsible for the following Bessel type inequalities.

PROPOSITION 5.9 (Nonmaximal Bessel's inequality; see [6]). *Let $j \in \{1, 2\}$. Consider a collection $\mathbf{S}' \subseteq \mathbf{S}$ of multitiles and let $\text{size}_j(\cdot)$ denote the j-size with respect to some function $F \in L^2(\mathbf{R})$. Then $\mathbf{S}'$ can be written as a disjoint union*

$$\mathbf{S}' = \mathbf{S}'_1 \cup \mathbf{S}'_2,$$

where

$$\text{size}_j(\mathbf{S}'_1) \le \frac{\text{size}_j(\mathbf{S}')}{2},$$

while $\mathbf{S}'_2$ consists of a family $\mathcal{F}_{\mathbf{S}'_2}$ of pairwise disjoint trees satisfying

$$\sum_{\mathbf{T} \in \mathcal{F}_{\mathbf{S}'_2}} |I_T| \lesssim \text{size}_j(\mathbf{S}')^{-2} \|F\|_2^2, \tag{5.4}$$

with the implicit constant independent of $\mathbf{S}'$ and F.

For $j = 3$ we will need the following version of the above result. We call forest any collection $\mathcal{F}$ of strongly 3-disjoint trees and we denote by

$$N_{\mathcal{F}}(x) = \sum_{\mathbf{T} \in \mathcal{F}} 1_{I_T}$$

the counting function of $\mathcal{F}$.

PROPOSITION 5.10 (Maximal Bessel's inequality). *Let $\mathbf{S}' \subseteq \mathbf{S}$ be a collection of multitiles and let $\mathbf{U} := u_1 < u_2 < \cdots < u_J$ be an arbitrary sequence of integers. For each $s \in \mathbf{S}'$ let $j(s)$ be the unique number in $\{1, 2, \ldots, J-1\}$ such that $2^{u_{j(s)}} \leq |I_s| < 2^{u_{j(s)+1}}$. Consider also an arbitrary sequence of functions $h_1, h_2, \ldots, h_{J-1} : \mathbf{R} \to \mathbf{C}$ satisfying*

$$\sum_{j=1}^{J-1} |h_j|^2 \equiv 1,$$

and a function $h \in X_2(E)$, where E is an arbitrary set of finite measure. Let $\text{size}_3(\cdot)$ *denote the 3-size with respect to the functions $H_s : \mathbf{R} \to \mathbf{C}$ defined by*

$$H_s = h h_{j(s)}.$$

Then $\mathbf{S}'$ can be written as a disjoint union

$$\mathbf{S}' = \mathbf{S}'_1 \cup \mathbf{S}'_2,$$

where

$$\text{size}_3(\mathbf{S}'_1) \leq \frac{\text{size}_3(\mathbf{S}')}{2},$$

while $\mathbf{S}'_2$ consists of a family $\mathcal{F}_{\mathbf{S}'_2}$ of pairwise disjoint trees satisfying

$$\sum_{\mathbf{T} \in \mathcal{F}_{\mathbf{S}'_2}} |I_T| \lesssim J^{1/8}\, \text{size}_3(\mathbf{S}')^{-2} \left(\frac{1}{\text{size}_3(\mathbf{S}')|E|^{1/2}} \right)^{\epsilon} \tag{5.5}$$

for each $\epsilon > 0$, with the implicit constant depending only on ϵ.

This proposition will follow from a chain of successive reductions, as in [4].

PROPOSITION 5.11 (Maximal Bessel's inequality, first reduction). *Assume $\mathbf{S}' \subseteq \mathbf{S}$ can be organized as a forest $\mathcal{F}'$ of trees $\mathbf{T}$ with tops T. Let $\mathbf{U} := u_1 < u_2 < \cdots < u_J$ be an arbitrary sequence of integers. Consider also an arbitrary sequence of functions $h_1, h_2, \ldots, h_{J-1} : \mathbf{R} \to \mathbf{C}$ satisfying*

$$\sum_{j=1}^{J-1} |h_j|^2 \equiv 1,$$

and a function $h \in X_2(E)$, where E is an arbitrary set of finite measure. Define the functions H_s as before. Assume also that

$$2^m \leq \left(\frac{1}{|I_T|} \sum_{s \in \mathbf{T}} |\langle H_s, \phi_{s,3} \rangle|^2 \right)^{1/2} \leq 2^{m+1}$$

for each $\mathbf{T} \in \mathcal{F}'$*, and*

$$\left(\frac{1}{|I_{T'}|}\sum_{\substack{s\in\mathbf{T}\\ I_s\subseteq I_{T'}}}|\langle H_s,\phi_{s,3}\rangle|^2\right)^{1/2}\le 2^{m+1}$$

for each $T' \in \mathbf{T} \in \mathcal{F}'$.

Then

$$\sum_{\mathbf{T}\in\mathcal{F}'}|I_T|\lesssim J^{1/8}2^{-2m}\left(\frac{1}{2^m|E|^{1/2}}\right)^{\epsilon}$$

for each $\epsilon > 0$*, with the implicit constant depending only on* ϵ.

Proposition 5.11 implies Proposition 5.10 by a standard tree selection algorithm; see [6].

The next reduction allows to replace the dependency on $|E|^{1/2}$ with a dependency on the counting function multiplicity. By linearity we can also eliminate the size parameter 2^m.

Proposition 5.12 (Maximal Bessel's inequality, second reduction). *Assume* $\mathbf{S}' \subseteq \mathbf{S}$ *can be organized as a forest* $\mathcal{F}'$ *of trees* $\mathbf{T}$ *with tops* T*. Let* $\mathbf{U} := u_1 < u_2 < \cdots < u_J$ *be an arbitrary sequence of integers. Consider also an arbitrary sequence of functions* $h_1, h_2, \ldots, h_{J-1} : \mathbf{R} \to \mathbf{C}$ *satisfying*

$$\sum_{j=1}^{J-1}|h_j|^2\equiv 1, \tag{5.6}$$

and a function $h \in L^2(\mathbf{R})$*. Define the functions* H_s *as before. Assume also that*

$$1\le\left(\frac{1}{|I_T|}\sum_{s\in\mathbf{T}}|\langle H_s,\phi_{s,3}\rangle|^2\right)^{1/2}\le 2 \tag{5.7}$$

for each $\mathbf{T} \in \mathcal{F}'$*, and*

$$\left(\frac{1}{|I_{T'}|}\sum_{\substack{s\in\mathbf{T}\\ I_s\subseteq I_{T'}}}|\langle H_s,\phi_{s,3}\rangle|^2\right)^{1/2}\le 2 \tag{5.8}$$

for each $T' \in \mathbf{T} \in \mathcal{F}'$*. Let* I_0 *be an interval which contains the support of* $N_{\mathcal{F}'}$.

Then

$$\sum_{\mathbf{T}\in\mathcal{F}'}|I_T|\lesssim J^{1/8}\|N_{\mathcal{F}'}\|_\infty^{\epsilon}\int|h|^2\chi_{I_0}^{10} \tag{5.9}$$

for each $\epsilon > 0$*, with the implicit constant depending only on* ϵ.

To see how Proposition 5.12 implies Proposition 5.11 we first introduce the BMO norm of a forest $\mathcal{F}$ as

$$\|\mathcal{F}\|_{\mathrm{BMO}} := \sup_I \frac{1}{|I|} \sum_{\mathbf{T}\in\mathcal{F}: I_T \subseteq I} |I_T|,$$

where the supremum is taken over all the dyadic intervals I. We then recall the following result from [4].

LEMMA 5.13. *Let $\mathcal{F}$ be a forest such that for some $\epsilon < 1$ and for some $A, B > 0$,*

$$\|N_{\mathcal{F}'}\|_1 \leq A\|N_{\mathcal{F}'}\|_\infty^\epsilon \text{ and } \|\mathcal{F}'\|_{\mathrm{BMO}} \leq B\|N_{\mathcal{F}'}\|_\infty^\epsilon$$

for all forests $\mathcal{F}' \subseteq \mathcal{F}$. Then we have

$$\|N_{\mathcal{F}}\|_1 \lesssim_\epsilon AB^{\frac{\epsilon}{1-\epsilon}}.$$

Proof of Proposition 5.11 assuming Proposition 5.12. Let $\mathcal{F}' \subset \mathcal{F}$ be arbitrary. From Proposition 5.12 with h replaced by $h/2^m$, and I_0 chosen to be so large as to contain all the time intervals arising from $\mathcal{F}'$, we have

$$\begin{aligned}
\|N_{\mathcal{F}'}\|_1 &= \sum_{\mathbf{T}\in\mathcal{F}'} |I_T| \\
&\lesssim \sum_{s\in\bigcup_{\mathbf{T}\in\mathcal{F}'}\mathbf{T}} |\langle H_s/2^m, \phi_{s,3}\rangle|^2 \\
&\lesssim_\epsilon J^{1/8}\|N_{\mathcal{F}'}\|_\infty^\epsilon \int |h/2^m|^2 \\
&\lesssim J^{1/8}2^{-2m}\|N_{\mathcal{F}'}\|_\infty^\epsilon
\end{aligned}$$

thanks to the L^2 normalization of $h \in X_2(E)$.

On the other hand, if I_0 is an arbitrary dyadic interval, then by replacing $\mathcal{F}'$ with $\{\mathbf{T} \in \mathcal{F}' : I_T \subseteq I_0\}$ in the above argument we see that

$$\begin{aligned}
\frac{1}{|I_0|} \sum_{\mathbf{T}\in\mathcal{F}': I_T\subseteq I_0} |I_T| &\lesssim_\epsilon \frac{J^{1/8}}{|I_0|}\|N_{\mathcal{F}'}\|_\infty^\epsilon \int |f/2^m|^2 \chi_{I_0}^{10} \\
&\lesssim J^{1/8}\|N_{\mathcal{F}'}\|_\infty^\epsilon 2^{-2m}|E|^{-1}
\end{aligned}$$

thanks to the uniform bound of $|E|^{-1/2}$ on $f \in X_2(E)$. Taking the suprema over I_0 we conclude that $\|\mathcal{F}'\|_{\mathrm{BMO}} \lesssim_\epsilon J^{1/8}\|N_{\mathcal{F}'}\|_\infty^\epsilon 2^{-2m}|E|^{-1}$. Applying Lemma 5.13 we conclude that

$$\sum_{\mathbf{T}\in\mathcal{F}} |I_T| = \|N_{\mathcal{F}}\|_1 \lesssim_\epsilon J^{1/8}2^{-2m}(2^{-2m}|E|^{-1})^{\frac{\epsilon}{1-\epsilon}}.$$

Since $\epsilon > 0$ is arbitrary, the proof of Proposition 5.11 follows. □

We next focus on proving Proposition 5.12. We will borrow some of the terminology from [4] in order to quote some results from there. Let $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$ be the dyadic grids

$$\begin{aligned} &\mathcal{D}_0 := \{[2^{-i}l, 2^{-i}(l+1)] : i, l \in \mathbf{Z}\} \text{ (i.e., the standard dyadic grid)},\\ &\mathcal{D}_1 := \{[2^{-i}(l+(-1)^i/3), 2^{-i}(l+1+(-1)^i/3)] : i, l \in \mathbf{Z}\},\\ &\mathcal{D}_2 := \{[2^{-i}(l-(-1)^i/3), 2^{-i}(l+1-(-1)^i/3)] : i, l \in \mathbf{Z}\}. \end{aligned} \tag{5.10}$$

One can easily verify that for every interval J (not necessarily dyadic) there exists a $d \in \{0,1,2\}$ and a shifted dyadic interval $J' \in \mathcal{D}_d$ such that $J \subseteq J' \subseteq 3J$; we will say that J is *d-regular*.

Let $A \geq 1$, and let $d \in \{0,1,2\}$. We shall say that a collection $\mathcal{I} \subset \mathcal{D}_0$ of time intervals is (A,d)-*sparse* if we have the following properties:

(i) If $I, I' \in \mathcal{I}$ are such that $|I| > |I'|$, then $|I| \geq 2^{100A}|I'|$.

(ii) If $I, I' \in \mathcal{I}$ are such that $|I| = |I'|$ and $I \neq I'$, then $\operatorname{dist}(I, I') \geq 100A|I'|$.

(iii) If $I \in \mathcal{I}$, then AI is d-regular, thus there exists an interval $I_A \in \mathcal{D}_d$ such that $AI \subseteq I_A \subseteq 3AI$. We refer to I_A as the A-*enlargement* of I.

If $\mathcal{I}$ is an (A,d)-sparse set of time intervals and P is a tile whose time interval I_P lies in $\mathcal{I}$, we write $I_{P,A}$ for the A-enlargement of I_P. We recall the following two results from [4].

THEOREM 5.14. *Let $A, D > 1$ and let $\mathcal{F}'$ be a forest with $\|N'_{\mathcal{F}}\|_\infty \leq D$. Let also $\mathbf{S}' := \bigcup_{\mathbf{T}\in\mathcal{F}'} \mathbf{T}$, and suppose that the time intervals*

$$\{I_s : s \in \mathbf{S}'\} \cup \{I_T : \mathbf{T} \in \mathcal{F}'\}$$

are (A,d)-sparse.

Then there exists an exceptional set $\mathbf{S}_* \subset \mathbf{S}'$ of multitiles with

$$\left|\bigcup_{s\in\mathbf{S}_*} I_s\right| \lesssim_\nu (A^{-\nu} + D^{-\nu}) \sum_{\mathbf{T}\in\mathcal{F}'} |I_T| \tag{5.11}$$

such that we have the Bessel-type inequality

$$\sum_{s\in\mathbf{S}'\setminus\mathbf{S}_*} |\langle f, \phi_{s,3}\rangle|^2 \lesssim_\nu ((\log(2+AD))^{10} + A^{10-\nu}D^{10})\|f\|_2^2$$

for each $\nu > 1$ and each $f \in L^2(\mathbf{R})$.

LEMMA 5.15 (Sparsification). *Let $\mathcal{I}$ be a collection of time intervals. Then we can split $\mathcal{I} = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_L$ with $L = O(A^2)$ such that each $\mathcal{I}_l$ for $1 \leq l \leq L$ is (A,d)-sparse for some $d = 0,1,2$.*

Proof of Proposition 5.12. We start by noting that it suffices to prove Proposition 5.12 without the localizing weight $\chi_{I_0}^{10}$. This is because $\chi_{I_0}^{-10}$ is a polynomial and hence $\chi_{I_0}^{-10}\psi_{s,3}$ satisfies the same properties (4.19) and

(4.20) as $\psi_{s,3}$. Let $\mu \geq 20$ be arbitrary. We first apply the above lemma with $A = C_\mu (J\|N_{\mathcal{F}'}\|_\infty)^{1/\mu}$ to the collection $\mathcal{I} := \{I_s : s \in \mathbf{S}'\}$ to split $\mathcal{I}$ into $\mathcal{I}_1, \ldots, \mathcal{I}_L$, for some $L = O(A^2)$. Each tree $\mathbf{T} \in \mathcal{F}'$ will be disintegrated over the collections $\mathcal{I}_l$. A further disintegration occurs by differentiating multitiles according to their time length, so in the end

$$\mathbf{T} = \bigcup_{1 \leq l \leq L} \bigcup_{1 \leq j \leq J-1} \mathbf{S}_{\mathbf{T},l,j}.$$

Here

$$\mathbf{S}_{\mathbf{T},l,j} = \{s \in \mathbf{T} : I_s \in \mathcal{I}_l,\ 2^{u_j} \leq |I_s| < 2^{u_{j+1}}\}.$$

According to Remark 5.8, each collection

$$\mathcal{F}_{l,j} = \bigcup_{\mathbf{T} \in \mathcal{F}'} \bigcup_{\mathbf{T}' \in D(\mathbf{S}_{\mathbf{T},l,j})} \mathbf{T}'$$

is a forest. It is easy to see that each $\mathcal{F}_{l,j}$ satisfies the requirements of Theorem 5.14 with A as above, and moreover

$$\|N_{\mathcal{F}_{l,j}}\|_\infty \leq \|N_{\mathcal{F}'}\|_\infty,$$
$$\|N_{\mathcal{F}_{l,j}}\|_1 \leq \|N_{\mathcal{F}'}\|_1.$$

Define

$$\mathbf{S}_{l,j} = \bigcup_{\mathbf{T} \in \mathcal{F}'} \mathbf{S}_{\mathbf{T},l,j}.$$

By applying Theorem 5.14 to each forest $\mathcal{F}_{l,j}$ with $D = C_\mu J \|N_{\mathcal{F}'}\|_\infty$, $f = hh_j$ and $\nu = 200\mu$ we get an exceptional set $\mathbf{S}_{l,j,*}$ such that

$$\left| \bigcup_{s \in \mathbf{S}_{l,j,*}} I_s \right| \lesssim_\mu C_\mu^{-200\mu} J^{-3} \|N_{\mathcal{F}'}\|_\infty^{-3} \sum_{\mathbf{T} \in \mathcal{F}_{l,j}} |I_T| \tag{5.12}$$

and we have the Bessel-type inequality

$$\sum_{s \in \mathbf{S}_{l,j} \setminus \mathbf{S}_{l,j,*}} |\langle hh_j, \phi_{s,3}\rangle|^2 \lesssim_\mu \left((\log(2 + \|N_{\mathcal{F}'}\|_\infty))^{10} + (\log(2+J))^{10}\right) \|hh_j\|_2^2. \tag{5.13}$$

Define now

$$\mathbf{S}_* := \bigcup_{l,j} \mathbf{S}_{l,j,*}$$

and note that if C_μ is chosen sufficiently large, then

$$\left| \bigcup_{s \in \mathbf{S}_*} I_s \right| \lesssim \frac{1}{10} \frac{\|N_{\mathcal{F}'}\|_1}{\|N_{\mathcal{F}'}\|_\infty}. \tag{5.14}$$

By summing up in (5.13) and invoking (5.6) we get

$$\sum_{s\in\mathbf{S}'\setminus\mathbf{S}_*}|\langle H_s,\phi_{s,3}\rangle|^2$$
$$\lesssim_\mu \|N_{\mathcal{F}'}\|_\infty^{2/\mu}J^{2/\mu}((\log(2+\|N_{\mathcal{F}'}\|_\infty))^{10}+(\log(2+J))^{10})\|h\|_2^2$$
$$\lesssim_\mu \|N_{\mathcal{F}'}\|_\infty^{3/\mu}J^{1/8}\|h\|_2^2.$$

Write $\Omega := \bigcup_{s\in\mathbf{S}_*} I_s$. Then

$$\sum_{s\in\mathbf{S}':I_s\nsubseteq\Omega}|\langle H_s,\phi_{s,3}\rangle|^2 \le \sum_{s\in\mathbf{S}'\setminus\mathbf{S}_*}|\langle H_s,\phi_{s,3}\rangle|^2.$$

To prove (5.9), it thus suffices to show that

$$\sum_{s\in\mathbf{S}':I_s\subseteq\Omega}|\langle H_S,\phi_{s,3}\rangle|^2 \le \frac{1}{2}\sum_{s\in\mathbf{S}'}|\langle H_s,\phi_{s,3}\rangle|^2. \tag{5.15}$$

From (5.7), it thus suffices to show that

$$\sum_{s\in\mathbf{S}':I_s\subseteq\Omega}|\langle H_s,\phi_{s,3}\rangle|^2 \le \frac{1}{2}\|N_{\mathcal{F}'}\|_1.$$

For each tree $\mathbf{T}$ in $\mathcal{F}'$, consider the multitile set $\{s\in\mathbf{T}: I_s\subseteq\Omega\}$. If s is any tile in this set with I_s maximal with respect to set inclusion, then $I_s\subseteq\Omega$ and from (5.8) we have

$$\sum_{s'\in\mathbf{T}:I_{s'}\subseteq I_s}|\langle H_{s'},\phi_{s',3}\rangle|^2 \le 4|I_s|.$$

Summing this over all such s upon noting that the I_s are disjoint by dyadicity and maximality, we conclude that

$$\sum_{s\in\mathbf{T}:I_s\subseteq\Omega}|\langle H_s,\phi_{s,3}\rangle|^2 \le 4|I_T\cap\Omega| = 4\int_\Omega 1_{I_T}.$$

Summing this over all $\mathbf{T}\in\mathcal{F}'$ we obtain

$$\sum_{s\in\mathbf{S}':I_s\subseteq\Omega}|\langle H_s,\phi_{s,3}\rangle|^2 \le 4\int_\Omega N_{\mathcal{F}'} \le 4|\Omega|\|N_{\mathcal{F}'}\|_\infty$$

and the claim (5.15) follows now from (5.14). $\square$

6. Single tree estimate

Consider a 3-lacunary tree $\mathbf{T}$ and some coefficients $(c_s)_{s\in\mathbf{T}}$. The following representation holds for each $2^k \le |I_T|$, assuming Δ is chosen large enough:

$$(6.1)\qquad \sum_{s\in\mathbf{T}:|I_s|\ge 2^k} c_s\psi_{s,3}(x)$$
$$= \int \mathcal{F}\left(\sum_{s\in\mathbf{T}} c_s\psi_{s,3}\right)(\xi)\zeta(2^k(\xi - c(\omega_{T,3})))e^{2\pi i\xi x}d\xi.$$

Here ζ is some universal function equal to 1 on $[-100e, 100e]$ and equal to 0 outside $[-200e, 200e]$.

THEOREM 6.1 (Single tree estimate). *Let $\mathbf{T}$ be a 3-lacunary tree in $\mathbf{S}$ with top T, and let*

$$\mathcal{P}_{\mathbf{T}} = \{I \text{ dyadic} : I_s \subseteq I \subseteq I_{s'} \text{ for some } s, s' \in \mathbf{T}\}$$

be the time convexification of $\mathbf{T}$. Consider a finite sequence of integers $u_1 < u_2 < \cdots < u_L$. For each $s \in \mathbf{T}$ let $l(s)$ be the unique number in $\{1, 2, \ldots, L-1\}$ such that $2^{u_{l(s)}} \le |I_s| < 2^{u_{l(s)+1}}$. Consider also an arbitrary sequence of functions $h_1, h_2, \ldots, h_{L-1} : \mathbf{R} \to \mathbf{C}$ satisfying

$$\sum_{l=1}^{L-1} |h_l|^2 \equiv 1,$$

and a function $h \in X(E)$, for some $E \subset \mathbf{R}$ of finite measure. For each $s \in \mathbf{T}$ define $H_s = hh_{l(s)}$.

Then

$$\left(\frac{1}{|I_T|}\sum_{s\in\mathbf{T}} |\langle H_s, \phi_{s,3}\rangle|^2\right)^{1/2} \lesssim \sup_{I\in\mathcal{P}_{\mathbf{T}}} \frac{1}{|I|}\int_E \chi_I^2,$$

with some universal implicit constant.

Proof. By frequency translation invariance we may assume that $0 \in \omega_{T,3}$.

By using the monotonicity of $\mathcal{P}_{\mathbf{T}'}$ relative to sub-trees $\mathbf{T}' \subseteq \mathbf{T}$, it suffices to prove that for each 3-lacunary tree satisfying the extra assumptions

$$\sum_{s\in\mathbf{T}} |\langle H_s, \phi_{s,3}\rangle|^2 = \alpha^2|I_T|,$$

and

$$|\langle H_s, \phi_{s,3}\rangle| \le \alpha|I_s|^{1/2}$$

for each $s \in \mathbf{T}$, we have

$$\alpha \lesssim \beta := \sup_{I\in\mathcal{P}_{\mathbf{T}}} \frac{1}{|I|}\int_E \chi_I^2.$$

Assume for contradiction that the above inequality does not hold; more precisely, assume

$$\alpha \geq K\beta \tag{6.2}$$

for some sufficiently large K, whose (implicit) value will become clear after a few lines of argument.

We first note that if $H_s^{(1)}$ denotes the restriction of H_s to the complement of $2I_T$, then from the decay of $\phi_{s,3}$ we get

$$|\langle H_s^{(1)}, \phi_{s,3}\rangle| \lesssim \left(\frac{|I_s|}{|I_T|}\right)^{10} |I_T|^{-1/2} \int_E \chi_{I_s}^2$$

for all $s \in \mathbf{T}$. So, if $H_s^{(2)}$ denotes the restriction of H_s to $2I_T$, we conclude that if K is large enough, then

$$\frac{\alpha^2}{2}|I_T| \leq \sum_{s\in\mathbf{T}} |\langle H_s^{(2)}, \phi_{s,3}\rangle|^2 \leq 2\alpha^2|I_T|, \tag{6.3}$$

$$|\langle H_s^{(2)}, \phi_{s,3}\rangle| \leq 2\alpha|I_s|^{1/2} \tag{6.4}$$

for each $s \in \mathbf{T}$.

We will next prove that

$$\frac{1}{|I_T|}\int_{2I_T} h\left(\sum_{l=1}^{L-1}\left|\sum_{\substack{s\in\mathbf{T}\\ 2^{u_l}\leq|I_s|<2^{u_{l+1}}}} \langle H_s^{(2)}, \phi_{s,3}\rangle\phi_{s,3}\right|^2\right)^{1/2} \lesssim \alpha\beta,$$

with an implicit constant independent of K. This together with the following consequence of the lower bound in (6.3)

$$\frac{\alpha^2}{2} \leq \frac{1}{|I_T|}\int_{2I_T} h\left(\sum_{l=1}^{L-1}\left|\sum_{\substack{s\in\mathbf{T}\\ 2^{u_l}\leq|I_s|<2^{u_{l+1}}}} \langle H_s^{(2)}, \phi_{s,3}\rangle\phi_{s,3}\right|^2\right)^{1/2}$$

will contradict (6.2).

In view of (6.1) we can estimate

$$\left|\sum_{s\in\mathbf{T}\ 2^{u_l}\leq|I_s|<2^{u_{l+1}}} \langle H_s^{(2)}, \phi_{s,3}\rangle\phi_{s,3}(x)\right|$$
$$\leq \sup_{u_l\leq k<u_{l+1}}\left|\sum_{\substack{s\in\mathbf{T}\\ 2^{k}\leq|I_s|<2^{u_{l+1}}}} \langle H_s^{(2)}, \phi_{s,3}\rangle\psi_{s,3}(x)\right| \lesssim M(F_l),$$

where

$$F_l(x) = \sum_{\substack{s\in\mathbf{T}\\ 2^{u_l}\leq|I_s|<2^{u_{l+1}}}} \langle H_s^{(2)}, \phi_{s,3}\rangle\psi_{s,3}(x).$$

It thus suffices to show that

$$\frac{1}{|I_T|}\int_{2I_T} h\left(\sum_{l=1}^{L-1} M(F_l)^2\right)^{1/2} \lesssim \alpha\beta. \tag{6.5}$$

For a dyadic interval J denote by J_1, J_2, J_3 the three dyadic intervals of the same length as J, sitting at the left of J, with J_3 being adjacent to J. Similarly let J_5, J_6, J_7 be the three dyadic intervals of the same length as J, sitting at the right of J, with J_5 being adjacent to J. Also define $J_4 = J$. Let $\mathcal{J}$ be the set of all dyadic intervals J with the following properties:

(a) $J \cap 2I_T \neq \emptyset$;
(b) $\nexists\, I \in \mathcal{P}_{\mathbf{T}} : |I| < |J|$ and $I \subset 3J$;
(c) $J_i \in \mathcal{P}_{\mathbf{T}}$ for some $1 \leq i \leq 7$.

We claim that $2I_T \subset \bigcup_{J\in\mathcal{J}} J$. Indeed, assume by contradiction that there exists some $x \in 2I_T \setminus \bigcup_{J\in\mathcal{J}} J$. Let $J^{(0)} \subset J^{(1)} \subset J^{(2)} \subset \cdots$ be the sequence of dyadic intervals of consecutive lengths containing x, with $|J^{(0)}| = \min_{I\in\mathcal{P}_{\mathbf{T}}} |I|$. Since $J^{(0)} \notin \mathcal{J}$ and since (a) and (b) are certainly satisfied for $J^{(0)}$, it follows that $J_i^{(0)} \notin \mathcal{P}_{\mathbf{T}}$ for each $1 \leq i \leq 7$. Moreover, note that for each $1 \leq i \leq 7$ there is no $I \in \mathcal{P}_{\mathbf{T}}$ with $I \subset J_i^{(0)}$. We proceed now by induction. Assume that for some $j \geq 0$ we have proved that for each $1 \leq i \leq 7$ we have $J_i^{(j)} \notin \mathcal{P}_{\mathbf{T}}$ and also that there is no $I \in \mathcal{P}_{\mathbf{T}}$ with $I \subset J_i^{(j)}$. Note that this implies the same for $j+1$. Indeed, since $3J^{(j+1)} \subset 7J^{(j)}$ and by induction hypothesis, it follows that (b) is satisfied for $J^{(j+1)}$. Hence $J_i^{(j+1)} \notin \mathcal{P}_{\mathbf{T}}$ for each $1 \leq i \leq 7$. We verify now the second statement of the induction. Note that if there were an $I \in \mathcal{P}_{\mathbf{T}}$ with $I \subset J_i^{(j+1)}$, then the hypothesis of the induction and the fact that $3J^{(j+1)} \subset 7J^{(j)}$ would imply that $i \in \{1, 2, 6, 7\}$. Hence $I \subset J_i^{(j+1)} \subset I_T$, and by convexity of $\mathcal{P}_{\mathbf{T}}$ it would follow that $J_i^{(j+1)} \in \mathcal{P}_{\mathbf{T}}$, which is impossible. This closes the induction. To see how the claim follows from here, observe that $I_T = J_i^{(j)}$ for some i, j, which certainly contradicts the fact that $I_T \in \mathcal{P}_{\mathbf{T}}$.

The next thing we prove is that on each interval $2J$ with $J \in \mathcal{J}$, the function

$$F(x) := \left(\sum_{l=1}^{L-1} |F_j(x)|^2\right)^{1/2}$$

is "essentially constant". More exactly, we will show that if $x \in 2J$, then $|F(x) - F(c(J))| \lesssim \alpha$.

Indeed, for each $x \in 2J$ we use (4.19) and the fact that $0 \in \omega_{T,3}$ to get

$$\left(\sum_{l=1}^{L-1} |F_l(x) - F_l(c(J))|^2\right)^{1/2}$$
$$\leq |J| \sum_{\substack{s \in \mathbf{T} \\ |I_s| \geq |J|}} \sup_{x \in 2J} \left|\frac{d}{dx}\psi_{s,3}(x)\right| \left|\langle H_s^{(2)}, \phi_{s,3}\rangle\right|$$
$$+ 2 \sum_{\substack{s \in \mathbf{T} \\ |I_s| < |J|, I_s \cap 3J = \emptyset}} \sup_{x \in 2J} |\psi_{s,3}(x)| \left|\langle H_s^{(2)}, \phi_{s,3}\rangle\right|,$$

since, by construction, there exists no $s \in \mathbf{T}$ such that $|I_s| < |J|$ and $I_s \subset 3J$. By using (6.4), this is further bounded by

$$\alpha|J| \sum_{\substack{s \in \mathbf{T} \\ |I_s| \geq |J|}} \frac{1}{|I_s|}\chi_{I_s}^{10}(c(J)) + \alpha \sum_{\substack{s \in \mathbf{T} \\ |I_s| < |J|, I_s \cap 3J = \emptyset}} \chi_{I_s}^{10}(c(J)) \lesssim \alpha.$$

Define now the measure space $X = \bigcup_{J \in \mathcal{J}} J$ and its sigma-algebra Υ generated by the maximal intervals $J \in \mathcal{J}$. Recall that $2I_T \subset \bigcup_{J \in \mathcal{J}} J = X \subset 10I_T$. We will see that for each $x \in J$

$$\left(\sum_{l=1}^{L-1} M(F_l)^2(x)\right)^{1/2} \lesssim \frac{1}{|J|}\int_J \left(\sum_{l=1}^{L-1} M(F_l)^2(z)\right)^{1/2} dz + \alpha. \tag{6.6}$$

Indeed, if $r_l > \frac{1}{2}|J|$, then

$$\left(\sum_{l=l}^{L-1}\left(\frac{1}{2r_j}\int_{x-r_l}^{x+r_l} |F_l|(z)dz\right)^2\right)^{1/2} \lesssim \inf_{y \in J}\left(\sum_{l=1}^{L-1} M(F_l)^2(y)\right)^{1/2}$$
$$\lesssim \frac{1}{|J|}\int_J \left(\sum_{l=1}^{L-1} M(F_l)^2(z)\right)^{1/2} dz.$$

On the other hand, if $r_l \leq \frac{1}{2}|J|$, then

$$\left(\sum_{l=1}^{L-1}\left(\frac{1}{2r_l}\int_{x-r_l}^{x+r_l} |F_l|(z)dz\right)^2\right)^{1/2} \lesssim \sup_{y \in 2J}\left(\sum_{l=1}^{L-1} |F_l|^2(y)\right)^{1/2}$$
$$\lesssim \inf_{y \in J}\left(\sum_{l=1}^{L-1} |F_l|^2(y)\right)^{1/2} + \alpha$$
$$\lesssim \frac{1}{|J|}\int_J \left(\sum_{l=1}^{L-1} M(F_l)^2(z)\right)^{1/2} dz + \alpha.$$

From (6.6) we can write

$$\begin{aligned}
&\frac{1}{|I_T|}\int h\left(\sum_{l=1}^{L-1} M(F_l)^2\right)^{1/2} \\
&\quad\lesssim \frac{1}{|I_T|}\int_X h\mathbb{E}\left(\left(\sum_{l=1}^{L-1} M(F_l)^2\right)^{1/2}|\Upsilon\right) + \alpha \sup_{J\in\mathcal{J}} \frac{1}{|J|}\int_J |h| \\
&\quad= \frac{1}{|I_T|}\int_X \mathbb{E}(h|\Upsilon)\mathbb{E}\left(\left(\sum_{l=1}^{L-1} M(F_l)^2\right)^{1/2}|\Upsilon\right) + \alpha \sup_{J\in\mathcal{J}} \frac{1}{|J|}\int_J |h| \\
&\quad\le \frac{1}{|I_T|}\|\mathbb{E}(h|\Upsilon)\|_{L^\infty}\int_X \mathbb{E}\left(\left(\sum_{l=1}^{L-1} M(F_l)^2\right)^{1/2}|\Upsilon\right) + \alpha \sup_{J\in\mathcal{J}} \frac{1}{|J|}\int_E \chi_J^2 \\
&\quad\lesssim \left(\alpha + \left[\frac{1}{|I_T|}\int_X \mathbb{E}\left(\left(\sum_{l=1}^{L-1} M(F_l)^2\right)^{1/2}|\Upsilon\right)^2\right]^{1/2}\right)\sup_{J\in\mathcal{J}} \frac{1}{|J|}\int_E \chi_J^2 \\
&\quad\lesssim \alpha \sup_{J\in\mathcal{J}} \frac{1}{|J|}\int_E \chi_J^2,
\end{aligned}$$

where $\mathbb{E}(\cdot|\Upsilon)$ denotes the conditional expectation relative to Υ. To get the last inequality above we have used the upper bound in (6.3) and the almost orthogonality of $\{\psi_{s,3}, s\in T\}$. Finally, note that since for each $J\in\mathcal{J}$, $J_i\in\mathcal{P}_{\mathbf{T}}$ for some i, we have that

$$\sup_{J\in\mathcal{J}} \frac{1}{|J|}\int_E \chi_J^2 \lesssim \sup_{I\in\mathcal{P}_{\mathbf{T}}} \frac{1}{|I|}\int_E \chi_I^2,$$

which certainly ends the proof of our theorem. □

7. The proof of Theorem 4.2

For each $f,g\in L^2(\mathbf{R})$ and each subset of tiles $\mathbf{S}'\subset\mathbf{S}$ define

$$M_{\mathbf{S}'}(f,g)(x) = \left(\sum_{j=1}^{J-1}\left|\sum_{\substack{s\in\mathbf{S}' \\ 2^{u_j}\le|I_s|<2^{u_{j+1}}}} |I_s|^{-1/2}\langle f,\phi_{s,1}\rangle\langle g,\phi_{s,2}\rangle\phi_{s,3}\right|^2\right)^{1/2}.$$

In order to prove inequality (4.22) it suffices to show that

$$|\{x: M_{\mathbf{S}}(f,g)(x)>\lambda\}| \lesssim \frac{J^{1/4}}{\lambda},$$

uniformly for each f,g with $\|f\|_2=\|g\|_2=1$ and each $\lambda>0$. We claim that it suffices to prove the above for $\lambda\sim 1$. Indeed, for arbitrary λ choose $k\in\mathbf{Z}$

such that $2^k \le \lambda < 2^{k+1}$. Also, for each multitile

$$s = [m2^i, (m+1)2^i] \times \prod_{j=1}^{3} \left[\frac{l_j}{5}2^{-i}, \left(\frac{l_j}{5}+1\right)2^{-i}\right] \in \mathbf{S}$$

define

$$s(k) := [m2^{i+k}, (m+1)2^{i+k}] \times \prod_{j=1}^{3} \left[\frac{l_j}{5}2^{-i-k}, \left(\frac{l_j}{5}+1\right)2^{-i-k}\right] \in \mathbf{S}.$$

Define also the collection of multitiles

$$\mathbf{S}(k) = \{s(k) : s \in \mathbf{S}\},$$

and the functions

$$\psi_{s(k),j}(x) = \frac{1}{2^{k/2}}\psi_{s,j}\left(\frac{x}{2^k}\right).$$

Note that $\mathbf{S}(k)$ and $\psi_{s(k),j}$ satisfy all the requirements of Theorem 4.2 (with a different choice of grids), with the same implicit constants as in (4.19). The claim now follows by noting that

$$\begin{aligned}|I_s|^{-1/2}&\langle f, \phi_{s,1}\rangle\langle g, \phi_{s,2}\rangle\phi_{s,3}(x)\\ &= |I_{s(k)}|^{-1/2}\langle \mathrm{Dil}^2_{2^k} f, \phi_{s(k),1}\rangle\langle \mathrm{Dil}^2_{2^k} g, \phi_{s(k),2}\rangle 2^k\phi_{s(k),3}(2^k x).\end{aligned}$$

Define now $\Gamma = \max\{[-\log_2(\mathrm{size}_1(\mathbf{S}))], [-\log_2(\mathrm{size}_2(\mathbf{S}))]\}$, where the 1-size is understood here with respect to the function f, while the 2-size is understood with respect to g. By iterating Proposition 5.9 simultaneously for $F = f$ and $F = g$, it follows that $\mathbf{S}$ can be written as a disjoint union $\mathbf{S} = \bigcup_{n\ge\Gamma}\mathbf{S}_n$, with

$$\mathrm{size}_j\left(\bigcup_{m\ge n}\mathbf{S}_m\right) \le 2^{-n} \tag{7.1}$$

for $j = 1, 2$, and each $\mathbf{S}_n$ consists of a family $\mathcal{F}_{\mathbf{S}_n}$ of pairwise disjoint trees satisfying

$$\sum_{\mathbf{T}\in\mathcal{F}_{\mathbf{S}_n}} |I_T| \lesssim 2^{2n}. \tag{7.2}$$

The contributions coming from the collections $\mathbf{S}' = \bigcup_{\Gamma\le n\le 0}\mathbf{S}_n$ and $\mathbf{S}'' = \bigcup_{n>0}\mathbf{S}_n$ will be evaluated quite differently.

In the case of $\mathbf{S}'$, crude estimates will suffice. Let $\mathbf{T} \in \mathcal{F}_{\mathbf{S}_n}$. By using (7.1), the decay in (4.19) and the triangle inequality we immediately get the estimate

$$M_{\mathbf{T}}(f,g)(x) \lesssim 2^{-2n}\chi^4_{I_T}(x) \tag{7.3}$$

for each $x \notin 2I_T$. For each $\Gamma \le n \le 0$ and each $\mathbf{T} \in \mathcal{F}_{\mathbf{S}_n}$ set $E_n = 2^{-n} I_T$. From (7.2) we get that the exceptional set $E = \bigcup_{\Gamma \le n \le 0} E_n$ has measure $\lesssim 1$. Also, (7.3) implies that

$$\begin{aligned}\|M_{\mathbf{S}'}(f,g)\|_{L_1(E^c)} &\le \sum_{\Gamma \le n \le 0} \|M_{\mathbf{S}_n}(f,g)\|_{L_1(E_n^c)} \\ &\lesssim \sum_{\Gamma \le n \le 0} 2^{-2n} 2^{3n} \\ &\lesssim 1.\end{aligned}$$

We conclude that

$$|\{x : M_{\mathbf{S}'}(f,g)(x) \gtrsim 1\}| \lesssim 1. \tag{7.4}$$

We will focus next on the estimates for the collection $\mathbf{S}''$. This time we will rely on the fact that $\text{size}_1(\mathbf{S}'') \le 1$ and $\text{size}_2(\mathbf{S}'') \le 1$. As before, for each $s \in \mathbf{S}$ let $j(s)$ denote the unique number in $\{1, 2, \ldots, J-1\}$ such that $2^{u_{j(s)}} \le |I_s| < 2^{u_{j(s)+1}}$. The 3-size will now intervene in a crucial way. Define

$$V := \{x : M_{\mathbf{S}''}(f,g)(x) \gtrsim 1\}.$$

If $|V| \le 1$ there is nothing to prove, so we will assume $|V| > 1$. Let $h_1, h_2, \ldots, h_{J-1} : \mathbf{R} \to \mathbf{C}$ be functions satisfying

$$\sum_{j=1}^{J-1} |h_j|^2 \equiv 1$$

such that

$$M_{\mathbf{S}''}(f,g)(x) = \sum_{s \in \mathbf{S}''} |I_s|^{-1/2} \langle f, \phi_{s,1} \rangle \langle g, \phi_{s,2} \rangle h_{j(s)} \phi_{s,3}(x).$$

From Theorem 6.1 we know that the 3-size of the collection $\mathbf{S}''$ with respect to the functions $H_s := |V|^{-1/2} 1_V h_{j(s)}$ is $\lesssim 1$. There is actually no restriction in assuming that it is ≤ 1. By applying iteratively Propositions 5.9 and 5.10 to the collection $\mathbf{S}''$ it follows that $\mathbf{S}''$ can be written as a disjoint union $\mathbf{S}'' = \bigcup_{n \ge 0} \mathbf{S}''_n$, with

$$\text{size}_j(\mathbf{S}''_n) \le 2^{-n} \tag{7.5}$$

for $j \in \{1, 2, 3\}$, and each $\mathbf{S}''_n$ consists of a family $\mathcal{F}_{\mathbf{S}''_n}$ of pairwise disjoint trees satisfying

$$\sum_{\mathbf{T} \in \mathcal{F}_{\mathbf{S}''_n}} |I_T| \lesssim J^{1/8} 2^{\frac{5}{2} n}. \tag{7.6}$$

Finally, by Lemma 5.6

$$\begin{aligned}|V|^{1/2} &\lesssim \langle M_{\mathbf{S}''}(f,g), |V|^{-1/2}1_V\rangle \\ &= \sum_{s\in\mathbf{S}''} |I_s|^{-1/2}\langle f,\phi_{s,1}\rangle\langle g,\phi_{s,2}\rangle\langle H_s,\phi_{s,3}\rangle \\ &\le \sum_{n\ge 0}\sum_{s\in\mathbf{S}''_n} |I_s|^{-1/2}\langle f,\phi_{s,1}\rangle\langle g,\phi_{s,2}\rangle\langle H_s,\phi_{s,3}\rangle \\ &\lesssim \sum_{n\ge 0} 2^{-3n}\sum_{\mathbf{T}\in\mathcal{F}_{\mathbf{S}''_n}} |I_T| \\ &\lesssim \sum_{n\ge 0} 2^{-3n}2^{\frac{5}{2}n}J^{1/8} \\ &\lesssim J^{1/8}.\end{aligned}$$

We conclude that

$$|\{x : M_{\mathbf{S}''}(f,g)(x) \gtrsim 1\}| \lesssim J^{1/4}. \tag{7.7}$$

The estimates in (7.4) and (7.7) end the proof of Theorem 4.2.

8. The proof of Theorem 1.9

By applying discretization techniques like in Section 4, Theorem 1.9 will be a consequence of the following.

THEOREM 8.1. *Let $\mathcal{G}$, $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ be four grids with $\mathcal{G}$ satisfying*

$$\mathcal{G} \subset \mathcal{S},$$

$$I, I' \in \mathcal{G} \Rightarrow \max\{|I||I'|^{-1}, |I'||I|^{-1}\} \ge 2^{\Delta}.$$

Let e be a number with $10^2 < |e| \le 10^5$ and define

$$\mathbf{D} = \left\{\prod_{j=1}^{3}\left[\frac{l_j}{5}2^i, \left(\frac{l_j}{5}+1\right)2^i\right] \in \prod_{j=1}^{3}\mathcal{G}_j :\ l_2 = l_1 + e,\ l_3 = l_1 + l_2\right\}.$$

Define also the set of multitiles

$$\mathbf{S} = \left\{s = I_s \times Q_s : I_s \in \mathcal{G},\ Q_s \in \mathbf{D} \text{ with sidelength } \frac{1}{|I_s|}\right\}.$$

Assume that each multitile $s = I_s \times \prod_{j=1}^{3}\omega_{s,j} \in \mathbf{S}$ is associated with three functions $(\psi_{s,j})_{j=1}^{3}$ satisfying

$$\left|\frac{d^n}{dx^n}\operatorname{Mod}_{-c(\omega_{s,j})}\psi_{s,j}(x)\right| \lesssim_{n,M} |I_s|^{-1/2-n}\chi_{I_s}^M(x),\quad n, M \ge 0, \tag{8.1}$$

$$\operatorname{supp}\widehat{\psi}_{s,j} \subset \omega_{s,j}, \tag{8.2}$$

for each $j = 1, 2, 3$.

Then for each $f, g \in L^2(\mathbf{R})$ we have the estimate

$$\left\| \left(\sup_{k\in\mathbf{Z}} \left| \sum_{\substack{s\in\mathbf{S} \\ |I_s|=2^k}} |I_s|^{-1/2} \langle f, \psi_{s,1}\rangle \langle g, \psi_{s,2}\rangle \psi_{s,3} \right|^2 \right)^{1/2} \right\|_{1,\infty} \lesssim \|f\|_2 \|g\|_2, \tag{8.3}$$

with the implicit constant depending only on the implicit constants in (8.1).

To prove this theorem amounts to proving Theorem 4.2 in the case $u_1, \ldots, u_J$ are consecutive integers, with a bound independent of J. By analyzing the whole argument for Theorem 4.2, it follows that the dependency on J in there comes from a single source, namely, the Maximal Bessel inequality in Proposition 5.10. This dependency is eliminated by proving the following version of Proposition 5.11.

PROPOSITION 8.2. *Assume $\mathbf{S}' \subseteq \mathbf{S}$ can be organized as a forest $\mathcal{F}'$ of trees $\mathbf{T}$ with tops T. Consider an arbitrary sequence of functions $h_k : \mathbf{R} \to \mathbf{C}$, $k \in \mathbf{Z}$ satisfying*

$$\sum_{k\in\mathbf{Z}} |h_k|^2 \equiv 1,$$

and a function $h \in L^2(\mathbf{R})$. Define the functions H_s by $H_s = hh_{\log(|I_s|)}$. Assume also that

$$2^m \le \left(\frac{1}{|I_T|} \sum_{s\in\mathbf{T}} |\langle H_s, \psi_{s,3}\rangle|^2 \right)^{1/2}$$

for each $\mathbf{T} \in \mathcal{F}'$.

Then

$$\sum_{\mathbf{T}\in\mathcal{F}'} |I_T| \lesssim 2^{-2m} \|h\|_2^2.$$

Proof. It suffices to prove that

$$\sum_{\substack{s\in\mathbf{S}' \\ |I_s|=2^k}} |\langle f, \psi_{s,3}\rangle|^2 \lesssim \|f\|_2^2,$$

uniformly in all $k \in \mathbf{Z}$ and all $f \in L^2(\mathbf{R})$. This in turn will follow by duality from the following estimate

$$\left\| \sum_{\substack{s\in\mathbf{S}' \\ |I_s|=2^k}} a_s \psi_{s,3} \right\|_2^2 \lesssim \sum_{\substack{s\in\mathbf{S}' \\ |I_s|=2^k}} |a_s|^2,$$

which holds for all sequence (a_s).

Indeed,

$$\Big\|\sum_{\substack{s\in\mathbf{S}'\\|I_s|=2^k}} a_s\psi_{s,3}\Big\|_2^2 = \sum_{\omega\in\mathcal{G}_3:|\omega|=2^{-k}}\Big\|\sum_{\substack{s\in\mathbf{S}'\\ \omega_s=\omega}} a_s\psi_{s,3}\Big\|_2^2$$

$$= \sum_{\omega\in\mathcal{G}_3:|\omega|=2^{-k}}\sum_{\substack{s,s'\in\mathbf{S}'\\ \omega_s=\omega_{s'}=\omega}} a_s\bar{a}_{s'}\langle\psi_{s,3},\psi_{s',3}\rangle$$

$$\lesssim \sum_{\omega\in\mathcal{G}_3:|\omega|=2^{-k}}\sum_{\substack{s,s'\in\mathbf{S}'\\ \omega_s=\omega_{s'}=\omega}} |a_s a_{s'}|\left(1+\frac{\operatorname{dist}(I_s,I_{s'})}{2^k}\right)^{-10}$$

$$\lesssim \sum_{\omega\in\mathcal{G}_3:|\omega|=2^{-k}}\sum_{\substack{s\in\mathbf{S}'\\ \omega_s=\omega}} |a_s|^2$$

$$= \sum_{s\in\mathbf{S}'}|a_s|^2. \qquad \square$$

References

[1] I. Assani, *Wiener Wintner ergodic theorems*, World Scientific Publishing Co. Inc., River Edge, NJ, 2003. MR 1995517 (2004g:37007)

[2] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. USA **17** (1931), 656–660.

[3] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161. MR 1037434 (91d:28029)

[4] C. Demeter, T. Tao, and C. Thiele, *Maximal multilinear operators*, Trans. Amer. Math. Soc. **360** (2008), 4989–5042.

[5] C. Demeter, M. Lacey, T. Tao, and C. Thiele, *Breaking the duality in the return times theorem*, to appear in Duke Math. J.; available at `http://arxiv.org/abs/math.CA/0601455`.

[6] M. T. Lacey, *On the bilinear Hilbert transform*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), Doc. Math., Extra Vol. II (1998), 647–656 (electronic). MR 1648113 (99h:42015)

[7] ———, *The bilinear maximal functions map into L^p for $2/3 < p \le 1$*, Ann. of Math. (2) **151** (2000), 35–57. MR 1745019 (2001b:42015)

[8] C. Muscalu, T. Tao, and C. Thiele, *L^p estimates for the biest. II. The Fourier case*, Math. Ann. **329** (2004), 427–461. MR 2127985 (2005k:42054)

Ciprian Demeter, Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA

Current address: School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA

E-mail address: `demeter@math.ias.edu`