Forum of Mathematics, Sigma (2015), Vol. 3, e25, 27 pages
doi:10.1017/fms.2015.25



# DYADIC TRIANGULAR HILBERT TRANSFORM OF TWO GENERAL FUNCTIONS AND ONE NOT TOO GENERAL FUNCTION

VJEKOSLAV KOVAČ[1], CHRISTOPH THIELE[2] and PAVEL ZORIN-KRANICH[2]

[1] Department of Mathematics, Faculty of Science, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia;
email: vjekovac@math.hr

[2] Department of Mathematics, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany;
email: thiele@math.uni-bonn.de, pzorin@math.uni-bonn.de

Received 7 July 2015; accepted 11 October 2015

**Abstract**

The so-called triangular Hilbert transform is an elegant trilinear singular integral form which specializes to many well-studied objects of harmonic analysis. We investigate $L^p$ bounds for a dyadic model of this form in the particular case when one of the functions on which it acts is essentially one dimensional. This special case still implies dyadic analogues of boundedness of the Carleson maximal operator and of the uniform estimates for the one-dimensional bilinear Hilbert transform.

2010 Mathematics Subject Classification: 42B20

## 1. Introduction

**1.1. Motivation.** In this article we begin the study of a dyadic model of the so-called triangular Hilbert transform. In order to motivate its definition, consider the family of trilinear forms (dual to two-dimensional bilinear Hilbert transforms)

$$\Lambda_{\vec{\beta}_0,\vec{\beta}_1,\vec{\beta}_2}(F_0, F_1, F_2) := \iint_{\mathbb{R}^2} \text{p.v.} \int_{\mathbb{R}} \prod_{i=0}^{2} F_i(\vec{x} - \vec{\beta}_i t) \frac{dt}{t} \, d\vec{x}, \tag{1.1}$$

where $\vec{\beta}_i \in \mathbb{R}^2$ are distinct points. When all three points $\vec{\beta}_i$ lie on the same line, these forms reduce to integrals of one-dimensional bilinear Hilbert transforms, and by the results of [**LT97**, **LT99**] we have the $L^p$ bounds

$$|\Lambda_{\vec{\beta}_0,\vec{\beta}_1,\vec{\beta}_2}(F_0, F_1, F_2)| \lesssim_{\vec{\beta}_0,\vec{\beta}_1,\vec{\beta}_2} \prod_{i=0}^{2} \|F_i\|_{p_i} \tag{1.2}$$

for all $1 < p_i < \infty$ with

$$\sum_{i=0}^{2} \alpha_i = 1, \quad \alpha_i = \frac{1}{p_i}. \tag{1.3}$$

By scaling, condition (1.3) is necessary.

Bounds of the form (1.2) in the more general situation when the $\vec{\beta}_i$ are in general position (that is, not on the same line) would unify some of the central results in time–frequency analysis.

(1) From the general case, one could recover $L^p$ bounds for the Carleson maximal operator by making an appropriate choice of the functions $F_i$; see Appendix B.

(2) From the general case, one could also recover the uniform bounds for the one-dimensional bilinear Hilbert transform in [**GL04**]. In fact, if the estimate (1.2) is true for any triple of $\vec{\beta}_i$ in general position, then it is true for every triple of $\vec{\beta}_i$, and the implied constant does not depend on the triple. This calculation is carried out in Appendix B, where it is verified that the constant will be the same as in the corresponding estimate for the trilinear form,

$$\Lambda_{\Delta}(F_0, F_1, F_2) := \text{p.v.} \iiint_{\mathbb{R}^3} F_0(x, y) F_1(y, z) F_2(z, x) \frac{d(x, y, z)}{x + y + z}, \tag{1.4}$$

which can be called the *triangular Hilbert transform*.

(3) By the method of rotations, a hypothetical estimate for (1.1) also implies $L^p$ estimates for the 'less singular' bilinear singular integrals from [**DT10**] uniformly over all choices of the 'direction matrices', at least for odd two-dimensional kernels; see Appendix B.

Unfortunately, the desired estimates for the triangular singular form (1.4) still seem to be out of reach of the current techniques and, from what we have said, they are expected to be highly nontrivial. In this paper we work in a dyadic model instead of the classical one, and we consider a particular case when one of the functions $F_0$, $F_1$, $F_2$ takes a special form. This case still turns out to be general enough to imply (dyadic versions of) both the $L^p$ bounds for the Carleson operator and uniform bounds for the bilinear Hilbert transform. In this sense our result has stronger one-dimensional consequences than the dyadic version of the argument from [DT10] which appears in [Dem15]: the latter does not contain the uniform bounds for the bilinear Hilbert transform.

**1.2. Notation.** Let us now introduce the dyadic model for the triangular Hilbert transform. In this model, the real line is replaced by the (Walsh) field $\mathbb{W} = \mathbb{F}_2((1/t))$ of one-sidedly infinite power series with coefficients in the two-element field $\mathbb{F}_2$. The field $\mathbb{W}$ is traditionally identified with $[0, \infty)$ via the map $\sum_k a_k t^k \mapsto \sum_k a_k 2^{-k}$, where $\mathbb{F}_2$ is identified with $\{0, 1\}$. This map is one-to-one on a conull set, and we normalize the Haar measure on $\mathbb{W}$ in such a way that this map becomes measure preserving. Under this identification the addition $\oplus$ and the multiplication $\circledast$ on $\mathbb{W}$ correspond to addition and multiplication of binary numbers without carrying over digits. We refer for instance to [Thi95, Section 1] for more details.

The sets

$$A_k := [0, 2^k), \quad k \in \mathbb{Z}$$

then become additive subgroups, and their cosets are simply *dyadic intervals* of length $2^k$, the collection of which will be denoted by $\mathbf{I}_k$. Some dyadic intervals (typically denoted by Latin letters, such as $I$) will be interpreted as time intervals, and they will always be subsets of the unit interval $[0, 1)$. Other dyadic intervals will be interpreted as frequency intervals (typically denoted by Greek letters, such as $\omega$), and they will have integer endpoints. For a dyadic interval $I$, we write $I^1$ for its left half and $I^{-1}$ for its right half. The unique dyadic parent of $I$ will be denoted $\operatorname{par} I$. When we mention a *dyadic square* we will always mean a dyadic square contained in $[0, 1)^2$.

We work with real-valued functions, which is no restriction, since all systemic functions under consideration, most notably the Haar functions, are real valued. Let us then reserve the letter $i$ to denote an index $i \in \{0, 1, 2\}$. It is convenient to regard $i$ as an element of $\mathbb{Z}/3\mathbb{Z}$, and interpret $i + 1$ and $i - 1$ correspondingly. We shall also consider the set $\mathcal{I}_k$ of all triples $\vec{I} = (I_0, I_1, I_2)$ of dyadic intervals contained in $[0, 1)$ such that

$$|I_0| = |I_1| = |I_2| = 2^k, \quad 0 \in I_0 \oplus I_1 \oplus I_2,$$

and the set $\mathcal{I} = \bigcup_{k \leqslant 0} \mathcal{I}_k$. We write

$$(I_0, I_1, I_2) \subset (J_0, J_1, J_2)$$

if $I_i \subset J_i$ for $i = 0, 1, 2$.

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

Any function $F$ on the unit square shall be interpreted as the integral operator

$$(F\varphi)(x) := \int_0^1 F(x, y)\varphi(y)\, dy$$

on $L^2([0, 1))$, denoted by the same letter. For any dyadic interval $I$, we normalize the *Haar function* $\mathrm{h}_I$ in $L^\infty$, so that $\mathrm{h}_I = \sum_{j\in\{\pm 1\}} j\mathbb{1}_{I^j}$. We shall also write $\mathrm{h}_I$ for the spatial multiplier operator acting on $L^2([0, 1))$ and defined by

$$(\mathrm{h}_I\varphi)(x) := \mathrm{h}_I(x)\varphi(x).$$

The *dyadic triangular Hilbert transform* can be written as

$$\Lambda^\epsilon(F_0, F_1, F_2) := \sum_{\vec{I}\in\mathcal{I}} \epsilon_{\vec{I}} |I_i|^{-1} \mathrm{tr}(\mathrm{h}_{I_i} F_{i-1} \mathrm{h}_{I_{i+1}} F_i \mathrm{h}_{I_{i-1}} F_{i+1}), \tag{1.5}$$

where $(\epsilon_{\vec{I}})_{\vec{I}\in\mathcal{I}}$ is an arbitrary sequence of scalars bounded in magnitude by 1, and $i \in \{0, 1, 2\}$ is a fixed index. The expression does not depend on the specific choice of $i$ by cyclicity of the trace. If the reader prefers an explicit integral representation, then (1.5) can be rewritten as

$$\begin{aligned}&\Lambda^\epsilon(F_0, F_1, F_2)\\ &\quad = \sum_{\vec{I}\in\mathcal{I}} \epsilon_{\vec{I}}\, |I_0|^{-1} \iiint_{\mathbb{W}^3} \mathrm{h}_{I_1}(x) F_0(x, y) \mathrm{h}_{I_2}(y) F_1(y, z) \mathrm{h}_{I_0}(z) F_2(z, x)\, dx\, dy\, dz,\end{aligned} \tag{1.6}$$

but we will continue to use the convenient 'trace-operator' notation. We note that (1.6) is a perfect Calderón–Zygmund kernel analogue of (1.4); that is,

$$\sum_{\vec{I}\in\mathcal{I}} \epsilon_{\vec{I}} |I_0|^{-1} \mathrm{h}_{I_1}(x)\mathrm{h}_{I_2}(y)\mathrm{h}_{I_0}(z)$$

replaces $1/(x + y + z)$. It is necessary to insert the coefficients $\epsilon_{\vec{I}}$, as otherwise the above kernel would telescope to the Dirac mass $\delta_0$ evaluated at $x + y + z$, and the form would become trivial. Informally speaking, the Walsh model cannot distinguish between p.v.$1/t$ and $\delta_0(t)$, so it becomes faithful only after breaking the form into scales. We obtain the following strong type estimates (see Section 2.1 for the definition of the character $e$).

THEOREM 1.7. *Let $F_0, F_1, F_2$ be functions supported on $A_0^2$. Suppose that either*

$$F_0(x_1, x_2) = f(x_2 \oplus (a \circledast x_1)) \quad \textit{for all } x_1, x_2 \in A_0 \tag{1.8}$$

*holds with some $a \in \mathbb{W} \setminus A_0$ and some measurable $f : \mathbb{W} \to \mathbb{R}$, or*

$$F_0(x_1, x_2) = f(x_2) e(N_{x_2} \circledast x_1) \quad \textit{for all } x_1, x_2 \in A_0 \tag{1.9}$$

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

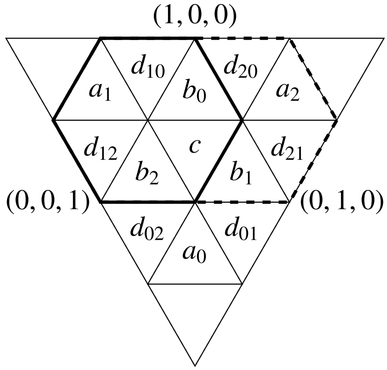


Figure 1. Ranges of exponents satisfying the scaling condition (1.3) in coordinates $(\alpha_0, \alpha_1, \alpha_2) = (1/p_0, 1/p_1, 1/p_2)$.

*holds with some measurable* $N : \mathbb{W} \to \mathbb{W}$ *and* $f : \mathbb{W} \to \mathbb{R}$. *Then*

$$|\Lambda^{\epsilon}(F_0, F_1, F_2)| \lesssim \|F_0\|_{p_0} \|F_1\|_{p_1} \|F_2\|_{p_2} \tag{1.10}$$

*for any* $1 < p_2 < \infty$ *and* $2 < p_0, p_1 < \infty$ *with* (1.3). *The implied constant does not depend on* $a$, $N$, *or the scalars* $|\epsilon_{\vec{I}}| \leqslant 1$ *with* $\epsilon_{\vec{I}} = 0$ *whenever some* $I_i \not\subseteq A_0$. *In case* (1.9), *we can relax the restriction on* $p_0$ *to* $1 < p_0 < \infty$. *In case* (1.8), $a \in A_1 \setminus A_0$, *we can relax the restrictions on both* $p_0$ *and* $p_1$ *to* $1 < p_0, p_1 < \infty$.

Since conditions (1.8) and (1.9) (with $a$ and $N$ fixed) describe subspaces of $L^{p_0}(\mathbb{W}^2)$ that are themselves $L^{p_0}$ spaces, Theorem 1.7 follows by real interpolation from (generalized) restricted weak type estimates. Such estimates also hold for certain negative values of $p_i$, the precise range of which is summarized below with the aid of Figure 1. Theorem 1.7 is the restriction of our results to the Banach triangle $c \cup b_0 \cup b_1 \cup b_2$ in Figure 1.

The local $L^2$ case (triangle $c$ in Figure 1) is covered by Proposition 2.4. In this case, the localization $\vec{I} \in \mathcal{I}_k$, $k \leqslant 0$, in definition (1.5) can be removed using the Loomis–Whitney inequality

$$\left| \iiint_{\mathbb{R}^3} F_0(x, y) F_1(y, z) F_2(z, x) \, d(x, y, z) \right| \leqslant \|F_0\|_2 \|F_1\|_2 \|F_2\|_2$$

to estimate contributions of scales $k > 0$.

Triangle $d_{12}$ is covered by Theorem 5.1; this gives the lower half of the solid hexagon in Figure 1. Triangle $d_{10}$ in cases (1.9) and (1.8), $a \in A_1 \setminus A_0$, is covered by Theorem 5.2; together with the previous result, this gives the full solid hexagon

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

in Figure 1. Finally, case (1.8), $a \in A_1 \setminus A_0$, is symmetric in indices 0, 2; in this case, we obtain estimates in the dashed extension of the solid hexagon in Figure 1.

Cases (1.8) and (1.9) are treated in a unified way, and they cover all types of function used to recover algebraically defined dyadic models for the Carleson operator and uniform estimates for the bilinear Hilbert transform; see Appendix A. However, note that already one type, namely case (1.8), $a \in A_1 \setminus A_0$, suffices to recover the bounds for both these operators. In particular, we recover the full range of exponents for which uniform estimates for the (dyadic) bilinear Hilbert transform are known (triangles $a_1, a_2$ have been treated in [**Li06**], and triangles $d_{12}, d_{21}$ in [**OT11**]). This range seems to be the best possible, because for $\alpha_{\pm 1} \leqslant -1/2$ there are indications that even nonuniform bounds for the bilinear Hilbert transform fail, whereas for $\alpha_0 \leqslant 0$ the bounds fail in the limiting case of the 1-linear Hilbert transform.

## 2. Tile decomposition

In this section, we describe a time–frequency decomposition for the form (1.5) that is well adapted both to diagonal functions (1.8) and to fiberwise characters (1.9). While the decomposition of the *form* is the same in both cases, the time–frequency projections of (one of) the *functions* differ. However, in both cases the time–frequency projections satisfy the same localization and scale compatibility properties, summarized in Definition 2.1. The proof of the local $L^2$ bounds uses only these properties and a single tree estimate. We will have to come back to the definition of time–frequency projections in the multifrequency Calderón–Zygmund decomposition in Section 5.

### 2.1. Wave packets.

The characters on the Walsh field $\mathbb{W}$ are the *Walsh functions*

$$w_N(x) := e(N \circledast x),$$

where $N \in \mathbb{W}$ and $e\colon \mathbb{W} \to \mathbb{R}$ is simply the periodization of $\mathrm{h}_{[0,1)}$. Their particular cases are the *Rademacher functions* $r_k := w_{2^{-k}}$, $k \in \mathbb{Z}$. The *Walsh wavepacket* associated with a dyadic rectangle $I \times \omega$ of area 1 is

$$w_{I\times\omega}(x) := |I|^{-1/2} 1_I(x) e(l(\omega) \circledast x),$$

where $l(\omega)$ is the left endpoint of $\omega$. This definition satisfies the usual recursive relations

$$w_{P_{\mathrm{up}}} = (w_{P_{\mathrm{left}}} - w_{P_{\mathrm{right}}})/\sqrt{2}, \quad w_{P_{\mathrm{down}}} = (w_{P_{\mathrm{left}}} + w_{P_{\mathrm{right}}})/\sqrt{2}$$

on every dyadic rectangle $P$ of area 2, and therefore coincides with the usual definition; see [**Thi95**, Section 1].

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

**2.2. Tile decomposition.** Our time–frequency analysis is $1\frac{1}{2}$-dimensional in the sense of [DT10]. We define *tiles* as dyadic boxes

$$\mathrm{p} = I_{\mathrm{p},0} \times I_{\mathrm{p},2} \times \omega_{\mathrm{p},1} \quad \text{where } |I_{\mathrm{p},0}| = |I_{\mathrm{p},2}| = |\omega_{\mathrm{p},1}|^{-1}.$$

A *bitile* is then any dyadic box of the form

$$P = I_{P,0} \times I_{P,2} \times \omega_{P,1} \quad \text{where } |I_{P,0}| = |I_{P,2}| = 2|\omega_{P,1}|^{-1}.$$

We will omit the subscripts p, $P$ if no confusion seems possible. For notational convenience we will throughout write $I_1 = I_0 \oplus I_2$.

Dyadic boxes are partially ordered by

$$P \leqslant P' :\iff I_i \subseteq I_i', \ \omega_i \supseteq \omega_i'.$$

Writing one of the Haar functions in (1.5) as a difference of two characteristic functions, we arrive at

$$\Lambda^{\epsilon}(F_0, F_1, F_2) = \sum_{\vec{I}} \epsilon_{\vec{I}} \sum_{j \in \{\pm 1\}} j|I_1|^{-1} \mathrm{tr}(1_{I_1^j} 1_{I_1^j} F_0 \mathrm{h}_{I_2} F_1 \mathrm{h}_{I_0} F_2),$$

where $1_I$ denotes, along with the characteristic function of the interval $I$, also the projection operator

$$(1_I \varphi)(x) = 1_I(x)\varphi(x).$$

Inserting identity operators (expanded in the Walsh basis) between characteristic functions, we obtain

$$\sum_{\vec{I}} \epsilon_{\vec{I}} \sum_{j \in \{\pm 1\}} j \sum_{\omega_1 : |\omega_1| = 2|I_1|^{-1}} 2|I_1|^{-2} \mathrm{tr}(1_{I_1^j}(w_{l_1} \otimes w_{l_1}) 1_{I_1^j} F_0 \mathrm{h}_{I_2} F_1 \mathrm{h}_{I_0} F_2).$$

Changing the order of summation, we obtain

$$\Lambda^{\epsilon}(F_0, F_1, F_2) = \sum_{P \text{ bitile}} \epsilon_{\vec{I}_P} \Lambda_P(F_0, F_1, F_2),$$

where

$$\Lambda_{\vec{I} \times \vec{\omega}}(F_0, F_1, F_2) := \sum_{j \in \{\pm 1\}} j 2|I_1|^{-2} \mathrm{tr}(1_{I_1^j}(w_l \otimes w_l) 1_{I_1^j} F_0 \mathrm{h}_{I_2} F_1 \mathrm{h}_{I_0} F_2).$$

Note that each $l$ can be replaced by any frequency from $\omega$, since this only multiplies the corresponding character by a constant on each of the intervals $I_i^j$.

**2.3. Time–frequency projections.** We begin by collecting desirable properties of time–frequency projections.

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

DEFINITION 2.1. We call orthogonal projections $\Pi_{\mathrm{p}}^{(i)}$, acting on $L^2(x_{i-1}, x_{i+1})$ and indexed by tiles p, *time–frequency projections* if they satisfy the following conditions.

(1) (Orthogonality) The projections $\Pi_{\mathrm{p}}^{(i)}$ corresponding to disjoint tiles are orthogonal.

(2) (Scale compatibility) Bitile projections $\Pi_P^{(i)}$ are well defined (there are two ways to write a bitile as a disjoint union of tiles, and the corresponding sums of tile projections are equal).

(3) (Support) $\operatorname{supp} \Pi_{\mathrm{p}}^{(i)} F_i \subset I_{i-1} \times I_{i+1}$.

A collection of bitiles $\mathbf{P}$ is called *convex* if $P, P'' \in \mathbf{P}$, $P \leqslant P' \leqslant P''$ implies that $P' \in \mathbf{P}$. The union of any finite convex collection of bitiles $\mathbf{P}$ can be written as the union of a collection of disjoint tiles $\mathbf{p}$ (this is proved by induction on the number of bitiles; see [Thi95, Lemma 1.7]). Given time–frequency projections, this allows us to consider the projections

$$\Pi_{\mathbf{P}}^{(i)} F_i := \sum_{\mathrm{p} \in \mathbf{p}} \Pi_{\mathrm{p}}^{(i)} F_i.$$

The scale compatibility property 2.1(2) implies that these projections do not depend on the choice of $\mathbf{p}$; see [Thi95, Corollary 1.9].

DEFINITION 2.2. We call time–frequency projections *adapted* to $F_0$ if, for every choice of $F_1, F_2$, every bitile $P$, and any convex collection of bitiles $\mathbf{P} \ni P$, we have

$$\Lambda_P(F_0, F_1, F_2) = \Lambda_P(\Pi_{\mathbf{P}}^{(0)} F_0, \Pi_{\mathbf{P}}^{(1)} F_1, \Pi_{\mathbf{P}}^{(2)} F_2). \tag{2.3}$$

The existence of adapted time–frequency projections suffices to establish restricted type bounds on the dyadic triangular Hilbert transform in the local $L^2$ range.

PROPOSITION 2.4. *Let $E_i \subset A_0^2$, $i \in \{0, 1, 2\}$, be measurable sets, and let $|F_i| \leqslant 1_{E_i}$ be functions for which there exist time–frequency projections adapted to $F_0$. Then*

$$|\Lambda^{\epsilon}(F_0, F_1, F_2)| \lesssim a_1^{1/2} a_2^{1/2} \left(1 + \log \frac{a_0}{a_1}\right),$$

*where $a_i = |E_{\sigma(i)}|$ is a decreasing rearrangement; that is, $\sigma$ is a permutation of $\{0, 1, 2\}$ and $a_0 \geqslant a_1 \geqslant a_2$. The implied constant is independent of the choices of the scalars $|\epsilon_{\vec{I}}| \leqslant 1$.*

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

We finish this section with the construction of time–frequency projections adapted to (1.8) and (1.9). For indices 0 and 2, we use the projections

$$\Pi_{\mathrm{p}}^{(2)} F_2(x_0, x_1) := 1_{I_0}(x_0)\langle F_2(x_0, \cdot), w_{I_1\times\omega_1}\rangle w_{I_1\times\omega_1}(x_1) \tag{2.5}$$

and

$$\Pi_{\mathrm{p}}^{(0)} F_0(x_2, x_1) := 1_{I_2}(x_2)\langle F_0(x_2, \cdot), w_{I_1\times\omega_1}\rangle w_{I_1\times\omega_1}(x_1). \tag{2.6}$$

The structural information given by (1.8) and (1.9) is encoded in the projections $\Pi^{(1)}$.

*2.3.1. One-dimensional functions* Suppose that (1.8) holds. Then we have

$$\Pi_{\mathrm{p}}^{(0)} F_0(x_1, x_2) = 1_{I_1}(x_1)(\Pi_{I_2\times a\circledast\omega_1} F_0(\cdot, x_1))(x_2),$$

where the projection on the right-hand side is a one-dimensional time–frequency projection (as defined for example in [**OT11**]) with a possibly multidimensional range. In this case, we define

$$\Pi_P^{(1)} F_1(x_2, x_0) := 1_{I_0}(x_0)(\Pi_{I_2\times a\circledast\omega_1} F_1(\cdot, x_0))(x_2).$$

*2.3.2. Fiberwise characters* Suppose that (1.9) holds. Then we have

$$\Pi_{\mathrm{p}}^{(0)} F_0(x_1, x_2) = 1_{I_1}(x_1)1_{I_2}(x_2)1_{\omega_1}(N_{x_2})F_0(x_1, x_2).$$

In this case, we define

$$\Pi_{\mathrm{p}}^{(1)} F_1(x_2, x_0) := 1_{I_0}(x_0)1_{I_2}(x_2)1_{\omega_1}(N_{x_2})F_1(x_2, x_0).$$

The projections $\Pi^{(1)}$ constructed above satisfy (2.3) only for bitiles with $I_i \subseteq A_0$, which explains the truncation in Theorem 1.7.

## 3. Single tree estimate

A *tree* $T$ is a convex set of bitiles that contains a maximal element

$$P_T = \vec{I}_T \times \vec{\omega}_T = I_{T,0} \times I_{T,2} \times \omega_{T,1}.$$

Equivalently, a tree can be described by a top frequency $\xi_{T,1}$ and a convex collection of space boxes $\mathcal{I}_T$. The corresponding tree $T$ then consists of all bitiles $P = \vec{I} \times \vec{\omega}$ with $\vec{I} \in \mathcal{I}_T$ and $\xi_{T,1} \in \omega_1$.

For a convex collection $\mathbf{P}$ of bitiles, define

$$\mathbf{size}^{(i)}(\mathbf{P}, F_i) := \sup_{T\subset\mathbf{P}\text{ tree}} |\vec{I}_T|^{-1/2} \|\Pi_T^{(i)} F_i\|_2. \tag{3.1}$$

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

For a collection $\mathbf{P}$ of bitiles, write

$$\Lambda^{\epsilon}_{\mathbf{P}}(F_0, F_1, F_2) := \sum_{P\in\mathbf{P}} \epsilon_{\vec{I}_P} \Lambda_P(F_0, F_1, F_2).$$

The objective of this section is to show that Definition 2.1 implies that

$$|\Lambda^{\epsilon}_T(F_0, F_1, F_2)| \lesssim |\vec{I}_T| \prod_{i=0}^{2} \mathbf{size}^{(i)}(T, F_i), \tag{3.2}$$

where $T$ is a tree and the implied constant is absolute. It follows from Definition 2.1 that

$$|\vec{I}_P|^{-1/2} \|\Pi^{(i)}_P F_i\|_{L^2(I_{i-1,P}\times I_{i+1,P})} \lesssim \mathbf{size}^{(i)}(T, F_i) \quad \text{for all } P \in T.$$

Thus in view of (2.3) it suffices to show that

$$|\Lambda^{\epsilon}_T(F_0, F_1, F_2)| \lesssim |\vec{I}_T| \prod_{i=0}^{2} \sup_{\vec{I}\in\mathcal{I}_T\cup\mathcal{L}_T} |\vec{I}|^{-1/2} \|F_i\|_{L^2(I_{i-1}\times I_{i+1})}, \tag{3.3}$$

where $\mathcal{L}_T$ denotes the collection of leaves of a tree, that is, maximal elements of $\mathcal{I}$ contained in a member of $T$ that are not themselves members of $\mathcal{I}_T$. By modulation, we may assume that $\xi_{T,1} = 0$. The tree operator can be written as

$$\sum_{\vec{I}\in\mathcal{I}_T} \epsilon_{\vec{I}} |I_1|^{-2} (\mathrm{tr}((1\otimes 1)1_{I_1} F_0 \mathrm{h}_{I_2} F_1 \mathrm{h}_{I_0} F_2 \mathrm{h}_{I_1}) + \mathrm{tr}((1\otimes 1)\mathrm{h}_{I_1} F_0 \mathrm{h}_{I_2} F_1 \mathrm{h}_{I_0} F_2 1_{I_1})).$$

The two summands are symmetric (under permuting the indices 0 and 2), and we consider only the first of them. With the convention that the domain of integration is $x_i, y_i \in I_i$ and the dyadic intervals have size $|I_i| = 2^k$, we have

$$\begin{aligned} &\mathrm{tr}((1\otimes 1)1_{I_1} F_0 \mathrm{h}_{I_2} F_1 \mathrm{h}_{I_0} F_2 \mathrm{h}_{I_1}) \\ &\quad = \int F_0(x_1, x_2) r_k(x_2) F_1(x_2, x_0) r_k(x_0) F_2(x_0, y_1) r_k(y_1)\, dx_1\, dx_2\, dx_0\, dy_1. \end{aligned}$$

The change of variables $x_1 = x_2 + y_0$, $y_1 = x_0 + y_2$ gives

$$\begin{aligned} &\int F_0(x_2 + y_0, x_2) r_k(x_2) F_1(x_2, x_0) r_k(x_0) \\ &\qquad \times F_2(x_0, x_0 + y_2) r_k(x_0 + y_2)\, dy_0\, dx_2\, dx_0\, dy_2 \\ &\quad = \int \tilde{F}_0(y_0, x_2) r_k(x_2) F_1(x_2, x_0) \tilde{F}_2(x_0, y_2) r_k(y_2)\, dy_0\, dx_2\, dx_0\, dy_2, \end{aligned}$$

where $\tilde{F}_0(y_0, x_2) := F_0(x_2 + y_0, x_2)$ and $\tilde{F}_2(x_0, y_2) := F_2(x_0, x_0 + y_2)$.

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

Thus the first half of the tree operator can be written as a single tree operator from [**Kov12**, Section 3] with square-dependent coefficients. The first step in the proof of [**Kov12**, Proposition 4] is an application of the Cauchy–Schwarz inequality in the sum over squares, so it still works in our situation. This, together with [**Kov12**, (2.2)], gives the required estimate.

## 4. Tree selection and local $L^2$ bounds

**4.1. The tree selection algorithm.** We organize bitiles into trees closely following the argument in [**OT11**, Lemma 2.2]. Here and later we use coordinate projections $\pi_{(i)} : \mathbb{W}^3 \to \mathbb{W}^2$, $(x_{i-1}, x_i, x_{i+1}) \mapsto (x_{i-1}, x_{i+1})$.

PROPOSITION 4.1. *Let $n \in \mathbb{Z}$, $i \in \{0, 1, 2\}$, a function $F_i$, and a system of (not necessarily adapted) time–frequency projections $\Pi^{(i)}$ be given. Then every finite convex collection of bitiles* $\mathbf{P}$ *can be partitioned into a convex collection of bitiles* $\mathbf{P}'$ *with*

$$\mathbf{size}_i(\mathbf{P}', F_i) \leqslant 2^{-n}$$

*and a further convex collection of bitiles that is the disjoint union of a collection of convex trees* $\mathbf{T}$ *with*

$$\sum_{T\in\mathbf{T}, \vec{I}_T \subset \vec{J}} |\vec{I}_T| \leqslant 9 \cdot 2^{2n} \|1_{\pi_{(i)}\vec{J}} F_i\|_2^2, \quad \vec{J} \in \mathcal{I}. \tag{4.2}$$

The latter bound includes both an $L^1$ estimate (taking $\vec{J}$ large enough to contain all time intervals in $\mathbf{P}$) and a BMO estimate (noting $\|1_{\pi_{(i)}\vec{J}} F_i\|_2^2 \leqslant |\vec{J}| \|F_i\|_\infty^2$) for the counting function $\sum_{T\in\mathbf{T}} 1_{\vec{I}_T}$.

*Proof.* We will remove three collections of trees, each of which satisfies (4.2) with a smaller constant. At each step we remove a tree that is also a down-set, thus ensuring that both the remaining collection $\mathbf{P}'$ and the collection of all removed tiles are convex.

Replacing $F_i$ by $2^n F_i$, we may assume that $n = 0$. We write every bitile $P$ as $P^{+1} \cup P^{-1}$, where the tiles $P^j$, $j = \pm 1$, are given by $\vec{I}_P \times \omega_{P,1}^j$.

For a tree $T$, write

$$T_j := \{P \in T : P^j \leqslant P_T\}, \quad j = \pm 1.$$

Then

$$\Pi_T^{(i)} F_i = \Pi_{P_T}^{(i)} F_i + \sum_{j=\pm 1} \sum_{P\in T_j} \Pi_{P^{-j}}^{(i)} F_i,$$

and this sum is orthogonal by Definition 2.1(1).

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

Let $\{P_1, \dots, P_n\}$ be the collection of maximal bitiles in $\mathbf{P}$ that satisfy

$$\|\Pi^{(i)}_{P_k} F_i\|_2^2 > 3^{-1}|\vec{I}_{P_k}|.$$

These bitiles are necessarily pairwise disjoint, so we have

$$\sum_{k:\vec{I}_{P_k}\subset\vec{J}} |\vec{I}_{P_k}| < 3 \sum_{k:\vec{I}_{P_k}\subset\vec{J}} \|\Pi^{(i)}_{P_k} F_i\|_2^2 \leqslant 3\|1_{\pi_{(i)}\vec{J}} F_i\|_2^2$$

for every $\vec{J} \in \mathcal{I}$, where the last inequality follows from parts (1) and (3) of Definition 2.1. Thus, removing the bitiles $P \leqslant P_k$ from $\mathbf{P}$, we may assume that

$$\|\Pi^{(i)}_{P} F_i\|_2^2 \leqslant 3^{-1}|\vec{I}_{P}|, \qquad P \in \mathbf{P}.$$

The next step will be done twice, for $j = \pm 1$. In each case, we remove a collection of trees $\mathbf{T}_j$ such that for every remaining tree $T$ we have

$$\sum_{P\in T_j} \|\Pi^{(i)}_{P^{-j}} F_i\|^2 \leqslant 3^{-1}|\vec{I}_T|^2. \tag{4.3}$$

The collection $\mathbf{T}_j = \{T_1, T_2, \dots\}$ is selected iteratively. Suppose that $T_1, \dots, T_k$ have been selected, and suppose that (4.3) is violated for some remaining tree $T \subset \mathbf{P} \setminus T_1 \cup \cdots \cup T_k$. Choose one such tree for which either the left endpoint of $\omega_{T,1}$ is minimal (if $j = -1$) or the right endpoint is maximal (for $j = +1$), and let $T_{k+1} \subset \mathbf{P}$ be the down-set spanned by the chosen tree.

We claim that the tiles of the form $P_m^{-j}$, $P_m \in (T_m)_j$, are pairwise disjoint. This is clear within each tree, so assume for contradiction that $P_k^{-j} < P_l^{-j}$, $k \neq l$. In particular, we have $P_k < P_l$, and this implies that $k < l$, since otherwise $P_k$ should have been included in $T_l$. On the other hand, $\omega^{-j}_{P_k,1} \supsetneq \omega^{-j}_{P_l,1}$ implies that $\omega^{-j}_{P_k,1} \supseteq \omega_{P_l,1} \supsetneq \omega^{j}_{P_l,1} \supseteq \omega_{T_l,1}$, whereas $\omega_{T_k,1} \subseteq \omega^{j}_{P_k,1}$. Thus $\omega_{T_k,1}$ is either to the right (if $j = -1$) or to the left (if $j = +1$) of $\omega_{T_l,1}$, in both cases contradicting the choice of $T_k$.

Violation of (4.3) for $T_k \in \mathbf{T}_j$ and parts (1) and (3) of Definition 2.1 give

$$\sum_{k:\vec{I}_{T_k}\subset\vec{J}} |\vec{I}_{T_k}| < \sum_{k:\vec{I}_{T_k}\subset\vec{J}} 3 \sum_{P\in(T_k)_j} \|\Pi^{(i)}_{P^{-j}} F_i\|_2^2 \leqslant 3\|1_{\pi_{(i)}\vec{J}} F_i\|_2^2,$$

as required. For each remaining tree, we will have

$$\|\Pi^{(i)}_{P_T} F_i\|^2 + \sum_{j=\pm1} \sum_{P\in T_j} \|\Pi^{(i)}_{P^{-j}} F_i\|^2 \leqslant (3^{-1} + 3^{-1} + 3^{-1})|\vec{I}_T|,$$

and this gives the required estimate for $\mathbf{size}_i(T, F_i)$. □

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

### 4.2. Local $L^2$ bounds (triangle $c$).

*Proof of Proposition 2.4.* Normalizing $\tilde{F}_i = F_i/|E_i|^{1/2}$, we have to show that

$$|\Lambda^{\epsilon}_{\mathbf{P}}(\tilde{F}_0, \tilde{F}_1, \tilde{F}_2)| \lesssim a_0^{-1/2}\left(1 + \log \frac{a_0}{a_1}\right) \tag{4.4}$$

with a constant independent of the (finite) convex collection of bitiles $\mathbf{P}$. We have $\mathbf{size}^{(i)}(\tilde{F}_i) \leqslant \|\tilde{F}_i\|_\infty \leqslant |E_i|^{-1/2} = a_{\sigma^{-1}(i)}^{-1/2}$ and $\|\tilde{F}_i\|_2 \leqslant 1$. Fix integers $n_i$ such that $2^{n_i-1} < a_i^{-1/2} \leqslant 2^{n_i}$; note that in particular $n_0 \leqslant n_1 \leqslant n_2$. Running the tree selection algorithm (Proposition 4.1) iteratively at each scale $n \leqslant n_2$ for each $i \in \{0, 1, 2\}$, we obtain collections of trees $\mathbf{T}_n$ with

$$\sum_{T \in \mathbf{T}_n} |I_{T,i}|^2 \lesssim 2^{-2n}$$

and

$$\mathbf{size}^{(i)}(T, \tilde{F}_i) \leqslant \min(2^n, 2^{n_{\sigma^{-1}(i)}}), \quad T \in \mathbf{T}_n.$$

Summing the single tree estimate (3.3) over all trees, we obtain

$$|\Lambda^{\epsilon}(\tilde{F}_0, \tilde{F}_1, \tilde{F}_2)| \lesssim \sum_{n \leqslant n_2} 2^{-2n} \prod_{i=0}^{2} \min(2^n, 2^{n_i}).$$

The sum over $n$ is an increasing geometric series for $n < n_0$ and a decreasing geometric series for $n > n_1$. In particular, the sum is dominated by the terms $n_0 \leqslant n \leqslant n_1$; that is, we have the estimate

$$2^{n_0}(1 + n_1 - n_0) \lesssim a_0^{-1/2}\left(1 + \log \frac{a_0}{a_1}\right),$$

as required. □

## 5. Fiberwise multifrequency Calderón–Zygmund decomposition and an extended range of exponents

In order to extend the range of exponents in our main result, we perform a fiberwise multifrequency Calderón–Zygmund decomposition. Here, in contrast to the local $L^2$ range, we have to use the special form of the time–frequency projections $\Pi^{(0)}$ and $\Pi^{(2)}$.

Our decomposition unites the main features of the one-dimensional multifrequency Calderón–Zygmund decomposition in [OT11] and the fiberwise single-frequency Calderón–Zygmund decomposition in [Ber12, Kov12].

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

A useful simplification with respect to [OT11] is that we do not attempt to control the size of the good function; this corresponds to the observation that the argument on [OT11, page 1709] works directly for $a$ in place of $a_m$.

### 5.1. Triangles $b_2$ and $d_{12}$.

THEOREM 5.1. *Let* $0 < \alpha_0 \leqslant 1/2 \leqslant \alpha_2 < 1$ *and* $-1/2 < \alpha_1 < 1/2$ *satisfy* (1.3). *Then for any measurable sets* $E_i \subset A_0^2$, $i \in \{0, 1, 2\}$ *there exists a major subset* $E_1' \subset E_1$ *(which can be taken equal to* $E_1$ *if* $\alpha_1 > 0$*) such that for any dyadic test functions* $|F_i| \leqslant 1_{E_i}$, $|F_1| \leqslant 1_{E_1'}$ *with* (1.8) *or* (1.9) *we have*

$$|\Lambda^{\epsilon}(F_0, F_1, F_2)| \lesssim_{\alpha_0,\alpha_1,\alpha_2} \prod_{i=0}^{2} |E_i|^{\alpha_i},$$

*where the implied constant is independent of the choices of the scalars* $|\epsilon_{\vec{I}}| \leqslant 1$ *with* $\epsilon_{\vec{I}} = 0$ *whenever* $I_i \not\subset A_0$.

*Proof.* The required estimate is invariant under rescaling by powers of 2, so we may normalize $|E_1| \approx 1$. The localization changes to $E_i \subset A_k^2$ for some $k \in \mathbb{Z}$, but all previous results still apply by scale invariance. When $|E_2| \gtrsim |E_1|$, the estimate with $E_1' = E_1$ follows from the local $L^2$ case $0 < \alpha_0, \alpha_1, \alpha_2 \leqslant 1/2$, which is given by Proposition 2.4. Thus we may assume that $|E_2| < 2^{-20}$.

Define the exceptional sets

$$B_0 := \{M_{p_0}(|E_0|^{-1/p_0} 1_{E_0}) > 2^{10}\}$$

and

$$B_2 := \{\tilde{M}_{p_2}(|E_2|^{-1/p_2} 1_{E_2}) > 2^{10}\},$$

where $\tilde{M}_{p_2}$ is the directional maximal function (in direction $x_1$). The set

$$B_1 := \pi_{(1)}((\pi_{(0)}^{-1} B_0 \cup \pi_{(2)}^{-1} B_2) \cap \Delta), \qquad \Delta := \{x_0 \oplus x_1 \oplus x_2 = 0\} \subset \mathbb{W}^3,$$

has measure $< 1/2$ by the Hardy–Littlewood maximal inequality. Consider the major subset $E_1' := E_1 \setminus B_1$.

Define normalized functions

$$\tilde{F}_i := |E_i|^{-1/p_i} F_i.$$

By construction of the major subset, only the bitiles $P$ with

$$\pi_{(1)} \vec{I}_P \not\subset B_1$$

contribute to the trilinear form $\Lambda$, so consider a finite convex collection $\mathbf{P}$ of such bitiles. Since the $M_{p_0}$ maximal function dominates the $M_2$ maximal function

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

pointwise, and by Definition 2.1(3), we have

$$\mathbf{size}^{(0)}(\mathbf{P}, \tilde{F}_0) \lesssim 1, \quad \mathbf{size}^{(1)}(\mathbf{P}, \tilde{F}_1) \lesssim 1.$$

By the tree selection algorithm in Proposition 4.1, we partition $\mathbf{P}$ into a sequence of pairwise disjoint convex unions of pairwise disjoint trees $\mathbf{P}_k = \bigcup_{T\in\mathbf{T}_k} T$ and a remainder set with zero contribution to $\Lambda$ in such a way that

$$\mathbf{size}^{(0)}(\mathbf{P}_k, \tilde{F}_0) \lesssim 2^{-k}$$

and

$$\|N_k\|_p \lesssim_p 2^{2k}\|\tilde{F}_0\|_2^{2/p}\|\tilde{F}_0\|_\infty^{2-2/p}, \quad N_k := \sum_{T\in T_k} 1_{\vec{I}_T}, \quad 1 \leqslant p < \infty.$$

Choosing $p = p_0/2$, we obtain the bound

$$\|N_k\|_p \lesssim_p 2^{2k}.$$

For a fixed $k$ we will show that

$$|\Lambda^{\epsilon}_{\mathbf{P}_k}(\tilde{F}_0, \tilde{F}_1, \tilde{F}_2)| \lesssim 2^{-\delta k}$$

for some $\delta > 0$, depending only on the $p_i$, to be determined later.

Let $\mathcal{I}_B$ denote the collection of the maximal one-dimensional dyadic intervals of the form $\{x_0\} \times J_1 \subset B_2$. For each one-dimensional interval $J = \{x_0\} \times J_1 \in \mathcal{I}_B$ let

$$\Omega_J := \{\omega : |\omega||J| = 1, \exists T \in \mathbf{T}_k : \vec{I}_T \supseteq J, \omega \supseteq \omega_T\}.$$

Let

$$G := \sum_{J\in\mathcal{I}_B} G_J, \quad G_J(x_0, x_1) := 1_J(x_0, x_1) \sum_{\omega\in\Omega_J} (\Pi_{J_1\times\omega}\tilde{F}_2(x_0, \cdot))(x_1).$$

The sum defining the function $G$ is pointwise finite, and $G$ is measurable since $\tilde{F}_2$ is a dyadic test function.

We claim that for every $P = \vec{I} \times \omega_1 \in \mathbf{P}_k$ we have

$$\Lambda_P(\tilde{F}_0, \tilde{F}_1, \tilde{F}_2) = \Lambda_P(\tilde{F}_0, \tilde{F}_1, G).$$

Since $E_2 \subset B_2$ by construction, and the collection $\mathcal{I}_B$ covers $B_2$, it suffices to show that

$$\int_{J_1} \tilde{F}_1(x_0, x_2)\mathrm{h}_{I_0}(x_0)\tilde{F}_2(x_0, x_1) w_{I_1^j\times\omega_1}(x_1)\, dx_1 = \int_{J_1} \tilde{F}_1(x_0, x_2)\mathrm{h}_{I_0}(x_0) G_J(x_0, x_1) w_{I_1^j\times\omega_1}(x_1)\, dx_1$$

for every $J = \{x_0\} \times J_1 \in \mathcal{I}_B$, every $x_2 \in I_2$, and every $j \in \{\pm 1\}$. If $I_0 \times I_1 \cap J = \emptyset$, then both sides vanish identically. therwise we must have $x_0 \in I_0$.

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

If now $I_1 \subseteq J_1$, then, by construction, $\tilde{F}_1$ vanishes on $\{x_0\} \times I_2$, so both sides again vanish identically. On the other hand, if $J_1 \subsetneq I_1$, then, by construction, $\Omega_J$ contains an ancestor of $\omega_1$, so the integrals coincide again. This finishes the proof of the claim.

Now we estimate $\|G\|_2$. By Hölder and Hausdorff–Young inequalities, we get

$$\begin{aligned}\|G_J\|_{L^2(J)}^2 &= \sum_{\omega\in\Omega_J} |\langle \tilde{F}_2(x_0,\cdot), w_{J_1\times\omega}\rangle|^2 \\ &\leqslant |\Omega_J|^{1-2/p_1'} \Big( \sum_{\omega\in\Omega_J} |\langle \tilde{F}_2(x_0,\cdot), w_{J_1\times\omega}\rangle|^{p_1'} \Big)^{2/p_1'} \\ &\leqslant |\Omega_J|^{1-2/p_1'} \|\tilde{F}_2\|_{L^{p_2}(J)}^2 |J_1|^{1-2/p_1}.\end{aligned}$$

Maximality of $J \subset B_2$ gives an upper bound on the above $L^{p_2}(J)$ norm, and we obtain

$$\|G_J\|_{L^2(J)}^2 \lesssim |\Omega_J|^{1-2/p_2'} |J_1| \leqslant \int_J N_k^{1-2/p_2'}.$$

Integrating these bounds, and using monotonicity of $L^p$ norms (recall $|B_2| \lesssim 1$), we get

$$\begin{aligned}\|G\|_2^2 &\lesssim \int_{B_2} N_k^{1-2/p_2'} \lesssim \Big( \int_{B_2} N_k^p \Big)^{(1-2/p_2')/p} \\ &\leqslant \|N_k\|_p^{1-2/p_2'} \lesssim 2^{2k(1-2/p_2')}.\end{aligned}$$

Normalize

$$\tilde{G} := 2^{-k(1-2/p_2')} G,$$

so that $\|\tilde{G}\|_2 \lesssim 1$. We claim that

$$|\Lambda^{\epsilon}_{\mathbf{P}_k}(\tilde{F}_0, \tilde{F}_1, \tilde{G})| \lesssim 2^{-k}(1+pk),$$

which would finish the proof. By the tree selection algorithm in Proposition 4.1 (beginning at some scale $l_0 \leqslant 0$ with $\mathbf{size}^{(2)}(\mathbf{P}_k, \tilde{G}) \leqslant 2^{-l_0}$), we partition

$$\mathbf{P}_k = \bigcup_{l=l_0}^{\lceil pk \rceil} \bigcup_{T\in\mathbf{T}_{k,l}} T \cup \mathbf{P}_k',$$

where

$$\mathbf{size}^{(2)}(T, \tilde{G}) \lesssim 2^{-l}, \quad \mathbf{size}^{(1)}(T, \tilde{F}_1) \lesssim \min(1, 2^{-l})$$

for $T \in \mathbf{T}_{k,l}$,

$$\sum_{T\in\mathbf{T}_{k,l}} |\vec{I}_T| \lesssim 2^{2l},$$

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

and

$$\mathbf{size}^{(2)}(\mathbf{P}'_k, \tilde{G}), \mathbf{size}^{(1)}(\mathbf{P}'_k, \tilde{F}_1) \lesssim 2^{-pk}.$$

By the single tree estimate (3.3), we obtain

$$\left|\sum_{l} \sum_{T \in \mathbf{T}_{k,l}} \Lambda^{\epsilon}_{T}(\tilde{F}_0, \tilde{F}_1, \tilde{G})\right| \lesssim \sum_{l=-\infty}^{\lceil pk \rceil} 2^{2l} 2^{-k} \min(1, 2^{-l}) 2^{-l} \lesssim 2^{-k}(1 + pk).$$

The remaining term can be written as

$$|\Lambda^{\epsilon}_{\mathbf{P}'_k}(\tilde{F}_0, \tilde{F}_1, \tilde{G})| = \left|\sum_{T \in \mathbf{T}_k} \Lambda^{\epsilon}_{T \cap \mathbf{P}'_k}(\tilde{F}_0, \tilde{F}_1, \tilde{G})\right|.$$

Each $T \cap \mathbf{P}'_k$ is the disjoint union of a set of trees the union of whose top squares has measure bounded by $|\vec{I}_T|$. We have

$$\sum_{T \in \mathbf{T}_k} |\vec{I}_T| \leqslant \left\| \sum_{T \in \mathbf{T}_k} 1_{\vec{I}_T} \right\|_p^p \lesssim 2^{2pk},$$

so, again by the single tree estimate (3.3),

$$|\Lambda^{\epsilon}_{\mathbf{P}'_k}(\tilde{F}_0, \tilde{F}_1, \tilde{G})| \lesssim 2^{2pk} 2^{-k} 2^{-pk} 2^{-pk} = 2^{-k},$$

finishing the proof of the claim. □

### 5.2. Triangles $b_0$ and $d_{10}$.

THEOREM 5.2. *Let* $0 < \alpha_2 \leqslant 1/2 \leqslant \alpha_0 < 1$ *and* $-1/2 < \alpha_1 < 1/2$ *satisfy* (1.3). *Then for any measurable sets* $E_i \subset A_0^2$, $i \in \{0, 1, 2\}$ *there exists a major subset* $E'_1 \subset E_1$ *(which can be taken equal to* $E_1$ *if* $\alpha_1 > 0$*) such that for any dyadic test functions* $|F_i| \leqslant 1_{E_i}$, $|F_1| \leqslant 1_{E'_1}$ *satisfying either* (1.8) *with* $a \in A_1 \setminus A_0$ *or* (1.9), *we have*

$$|\Lambda^{\epsilon}(F_0, F_1, F_2)| \lesssim_{\alpha_0,\alpha_1,\alpha_2} \prod_{i=0}^{2} |E_i|^{\alpha_i},$$

*where the implied constant is independent of the choices of the scalars* $|\epsilon_{\vec{I}}| \leqslant 1$ *with* $\epsilon_{\vec{I}} = 0$ *whenever* $I_i \not\subset A_0$.

*Proof.* We can assume that $|E_0| \leqslant 2^{-20}|E_1|$, since otherwise the conclusion follows from the local $L^2$ case with $E'_1 = E_1$.

In case (1.9), we can also without loss of generality assume that $E_0 = A_0 \times \tilde{E}_0$. Setting $E'_1 = E_1 \setminus \tilde{E}_0 \times A_0$, we get that the left-hand side of the conclusion vanishes identically.

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

In case (1.8), we argue as in the proof of Theorem 5.1, with the roles of indices 0 and 2 interchanged. The main difference from the previous case is that the time–frequency projections in general need not be adapted to the good function $G$. However, under the additional condition $a \in A_1 \setminus A_0$, we may assume that

$$\mathbb{1}_{B_0}(x_1, x_2) = \mathbb{1}_{\tilde{B}_0}(x_2 \oplus (a \circledast x_1)),$$

and then the directional maximal function $\tilde{M}_{p_0} \mathbb{1}_{B_0}$ coincides with the two-dimensional maximal function $M_{p_0} \mathbb{1}_{B_0}$. It follows that for every $J \in \mathcal{I}_B$ and every bitile $P = \vec{I} \times \omega_1 \in \mathbf{P}$ we have either $J \cap I_1 \times I_2 = \emptyset$ or $J_1 \subsetneq I_1$, which in turn implies that

$$\Pi^{(0)}_{\mathbf{P}_k} \tilde{F}_0 = \Pi^{(0)}_{\mathbf{P}_k} G.$$

Thus we may replace $\tilde{F}_0$ by $G$ in the single tree estimates. □

## Acknowledgements

V. K. was partially supported by the Croatian Science Foundation under project 3526. C. T. was partially supported by NSF grant DMS 1001535.

## Appendix A. Known special cases

Let us discuss briefly how our main result specializes to some cases that have already appeared in the literature in a very similar form.

**A.1. Maximally modulated Haar multiplier.** Since the ordinary Haar multipliers

$$(H^\epsilon f)(x) := \sum_I \epsilon_I |I|^{-1} \langle f, \mathrm{h}_I \rangle \mathrm{h}_I(x),$$

where $|\epsilon_I| \leqslant 1$ for each dyadic interval $I$, constitute a good dyadic model for the Hilbert transform, the maximally modulated Haar multipliers

$$(H^\epsilon_\star f)(x) := \sup_N |(H^\epsilon M_N f)(x)| \tag{A.1}$$

provide a reasonable algebraic model for the Carleson operator, albeit different from the model of truncated Walsh–Fourier series considered for example in [Bil67]. Here $M_N$ simply represents the Walsh modulation operator,

$$(M_N f)(x) := w_N(x) f(x).$$

Let $\epsilon_{\vec{I}} = \epsilon_{(I_0, I_1, I_2)}$ depend only on the interval $I_0$, and take two functions $f$ and $g$ on $A_0$. Suppose that $N\colon A_0 \to \{0, 1, 2, \ldots\}$ is a choice function that linearizes the supremum in (A.1). If we substitute

$$F_0(x_1, x_2) := f(x_1 \oplus x_2),$$
$$F_1(x_2, x_0) := \operatorname{sgn} g(x_0)\sqrt{|g(x_0)|}\, w_{N(x_0)}(x_2),$$
$$F_2(x_0, x_1) := \sqrt{|g(x_0)|}\, w_{N(x_0)}(x_1 \oplus x_0)$$

into (1.6), we will obtain for $\Lambda^{\epsilon}(F_0, F_1, F_2)$ the equal expression

$$\sum_{\vec{I}\in\mathcal{I}} \frac{\epsilon_{I_0}}{|I_0|} \iiint f(x_1 \oplus x_2) g(x_0) w_{N(x_0)}(x_1 \oplus x_0) w_{N(x_0)}(x_2) \mathrm{h}_{I_1}(x_1) \mathrm{h}_{I_2}(x_2) \mathrm{h}_{I_0}(x_0)\, d\vec{x}.$$

Here and later in this appendix we use the convention that $x_i, y_i \in I_i$ for integration domains, unless specified otherwise. By the character property of the Walsh functions and the fact that the Haar functions are simply restrictions of the Rademacher functions to the corresponding intervals, this equals

$$\sum_{\vec{I}\in\mathcal{I}} \frac{\epsilon_{I_0}}{|I_0|} \iiint f(x_1 \oplus x_2) g(x_0) w_{N(x_0)}(x_1 \oplus x_2 \oplus x_0) r_k(x_1 \oplus x_2 \oplus x_0)\, d\vec{x}.$$

By changing the variables $y_0 = x_1 \oplus x_2$ (for fixed $x_1$) and observing $y_0 \in I_1 \oplus I_2 = I_0$, the above equals

$$\sum_{\vec{I}\in\mathcal{I}} \frac{\epsilon_{I_0}}{|I_0|} \iiint f(y_0) g(x_0) w_{N(x_0)}(y_0 \oplus x_0) r_k(y_0 \oplus x_0)\, dy_0\, dx_1\, dx_0.$$

Observe that at each scale $k$ the integral $\sum_{I\in\mathbf{I}_k} \int_{x_1\in I_1}$ can be disregarded, as it simply integrates over the union of intervals $I_1$, which is $A_0$. Using the character property once again, we obtain

$$\begin{aligned}
&\sum_{I_0} \frac{\epsilon_{I_0}}{|I_0|} \iint f(y_0) g(x_0) w_{N(x_0)}(y_0) w_{N(x_0)}(x_0) \mathrm{h}_{I_0}(y_0) \mathrm{h}_{I_0}(x_0)\, dy_0\, dx_0 \\
&\quad = \int_{\mathbb{W}} \sum_{I_0} \frac{\epsilon_{I_0}}{|I_0|} \langle w_{N(x_0)} f, \mathrm{h}_{I_0}\rangle \mathrm{h}_{I_0}(x_0) w_{N(x_0)}(x_0) g(x_0)\, dx_0 \\
&\quad = \int (M_{N(x_0)} H^{\epsilon} M_{N(x_0)} f)(x_0) g(x_0)\, dx_0.
\end{aligned}$$

From the established bound for $\Lambda^{\epsilon}$ in Theorem 1.7 using duality, we deduce that

$$\|H^{\epsilon}_{\star} f\|_p \lesssim \|f\|_p \quad \text{for any } 1 < p < \infty.$$

**A.2. Walsh model of uniform bilinear Hilbert transform.** Theorem 1.7 implies a bound for the trilinear form

$$\Lambda^{\epsilon,L}_{\mathrm{BHT}}(f, g, h) := \int \sum_{k} \sum_{\substack{I\in\mathbf{I}_k \\ \omega:|\omega|=2^{-k}}} \epsilon_I (\Pi_{I\times(\omega\oplus 2^{-k})} f)(\Pi_{I\times(2^L\omega)} g)(\Pi_{I\times(2^L\omega\oplus\omega\oplus 2^{-k})} h),$$

where $\epsilon = (\epsilon_I)_I$ is a sequence of coefficients indexed by dyadic intervals and satisfying $|\epsilon_I| \leqslant 1$, while $L$ is an arbitrary positive integer. This observation is interesting, because a single estimate for the triangular Hilbert transform implies bounds for a sequence of one-dimensional trilinear forms $\Lambda_{\mathrm{BHT}}^{\epsilon,L}$ with constants independent of $\epsilon$ and $L$.

This form is similar to, but different from, the trilinear form studied in [**OT11**]. As in Appendix A.1, the discrepancy is due to the fact that our model is based on the algebraic structure of the Walsh field rather than on the order structure.

In order to apply Theorem 1.7, substitute

$$F_0(x_1, x_2) := f(x_1 \oplus x_2 \oplus 2^{-L}x_2),$$
$$F_1(x_2, x_0) := h(2^{-L}x_2 \oplus x_0),$$
$$F_2(x_0, x_1) := g(x_0 \oplus 2^{-L}x_0 \oplus 2^{-L}x_1)$$

into (1.6) to obtain

$$\begin{aligned}\Lambda^{\epsilon}(F_0, F_1, F_2) = \sum_k 2^{-k} \sum_{\vec{I}\in\mathcal{I}_k} \epsilon_{\vec{I}} \iiint & f(x_1 \oplus x_2 \oplus 2^{-L}x_2) \\ & \times g(x_0 \oplus 2^{-L}x_0 \oplus 2^{-L}x_1) \\ & \times h(2^{-L}x_2 \oplus x_0) r_k(x_1 \oplus x_2 \oplus x_0)\, dx_1\, dx_2\, dx_0.\end{aligned}$$

Observe that $x_i \in I_i$, $i = 0, 1, 2$, implies that

$$x_1 \oplus x_2 \oplus 2^{-L}x_2 \in I_1 \oplus I_2 \oplus 2^{-L}I_2 = I_0 \oplus 2^{-L}I_2,$$
$$x_0 \oplus 2^{-L}x_0 \oplus 2^{-L}x_1 \in I_0 \oplus 2^{-L}(I_0 \oplus I_1) = I_0 \oplus 2^{-L}I_2,$$
$$2^{-L}x_2 \oplus x_0 \in I_0 \oplus 2^{-L}I_2,$$

so we should expand $f, g, h$ into the Walsh–Fourier series on the dyadic interval $I = I_0 \oplus 2^{-L}I_2$ of length $2^k$, that is, into the wave packets with fixed eccentricity:

$$f(x_1 \oplus x_2 \oplus 2^{-L}x_2) = 2^{-k} \sum_{m_0=0}^{\infty} \langle f, 1_I w_{m_0 2^{-k}} \rangle w_{m_0 2^{-k}}(x_1 \oplus x_2 \oplus 2^{-L}x_2),$$

$$g(x_0 \oplus 2^{-L}x_0 \oplus 2^{-L}x_1) = 2^{-k} \sum_{m_2=0}^{\infty} \langle g, 1_I w_{m_2 2^{-k}} \rangle w_{m_2 2^{-k}}(x_0 \oplus 2^{-L}x_0 \oplus 2^{-L}x_1),$$

$$h(2^{-L}x_2 \oplus x_0) = 2^{-k} \sum_{m_1=0}^{\infty} \langle h, 1_I w_{m_1 2^{-k}} \rangle w_{m_1 2^{-k}}(2^{-L}x_2 \oplus x_0).$$

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

Inserting these into the previous expression for $\Lambda^{\epsilon}(F_0, F_1, F_2)$, we obtain

$$\sum_k 2^{-4k} \sum_{\vec{I} \in \mathcal{I}_k} \epsilon_{\vec{I}} \sum_{m_0, m_2, m_1} \langle f, 1_I w_{m_0 2^{-k}} \rangle \langle g, 1_I w_{m_2 2^{-k}} \rangle \langle h, 1_I w_{m_1 2^{-k}} \rangle$$
$$\left( \int_{I_1} e((m_0 \oplus m_2 2^{-L} \oplus 1) 2^{-k} \circledast x_1) dx_1 \right)$$
$$\left( \int_{I_2} e((m_0 \oplus m_0 2^{-L} \oplus m_1 2^{-L} \oplus 1) 2^{-k} \circledast x_2) dx_2 \right)$$
$$\left( \int_{I_0} e((m_2 \oplus m_2 2^{-L} \oplus m_1 \oplus 1) 2^{-k} \circledast x_0) dx_0 \right).$$

Since we are integrating over intervals of length $2^k$, the above summands vanish unless

$$m_0 \oplus m_2 2^{-L} \oplus 1, \quad m_0 \oplus m_0 2^{-L} \oplus m_1 2^{-L} \oplus 1, \quad \text{and} \quad m_2 \oplus m_2 2^{-L} \oplus m_1 \oplus 1$$

all belong to $A_0$, which is easily seen to be equivalent to the conditions

$$m_0 \oplus m_2 \oplus m_1 = 0 \quad \text{and} \quad m_0 \oplus m_2 2^{-L} \oplus 1 \in A_0.$$

Moreover, in that case the three functions under the integrals over $I_1, I_2, I_0$ are precisely the constants

$$2^k e((m_0 \oplus m_2 2^{-L} \oplus 1) 2^{-k} \circledast l(I_i)), \quad i = 1, 2, 0,$$

where $l(I_i)$ is the left endpoint of $I_i$. Because $0 \in I_0 \oplus I_1 \oplus I_2$, they multiply to $2^{3k}$. Allow the coefficients $\epsilon_{\vec{I}}$ to depend on $I = I_0 \oplus 2^{-L} I_2$ only, and observe that each interval $I \in \mathbf{I}_k$ appears for exactly $2^{-k}$ choices of $\vec{I}$ as they range over $\mathcal{I}_k$. (Indeed, $I_2$ is arbitrary, and $I_0$, $I_1$ are then uniquely determined.) We end up with

$$\sum_k 2^{-2k} \sum_{I \in \mathbf{I}_k} \epsilon_I \sum_{\substack{m_0, m_2, m_1 \\ m_0 \oplus m_2 \oplus m_1 = 0 \\ m_0 \oplus m_2 2^{-L} \oplus 1 \in A_0}} \langle f, 1_I w_{m_0 2^{-k}} \rangle \langle g, 1_I w_{m_2 2^{-k}} \rangle \langle h, 1_I w_{m_1 2^{-k}} \rangle;$$

that is, by substituting $m = m_0 \oplus 1$ and $n = m_2 \oplus (m_0 \oplus 1) 2^L$,

$$\sum_k 2^{-2k} \sum_{I \in \mathbf{I}_k} \epsilon_I \sum_{\substack{m,n \\ 0 \leqslant n < 2^L}} \langle f, 1_I w_{(m \oplus 1) 2^{-k}} \rangle \langle g, 1_I w_{(m 2^L \oplus n) 2^{-k}} \rangle$$
$$\langle h, 1_I w_{(m 2^L \oplus n \oplus m \oplus 1) 2^{-k}} \rangle. \tag{A.2}$$

On the other hand, we can start from $\Lambda_{\mathrm{BHT}}^{\epsilon, L}$, and write the dyadic interval $\omega$ explicitly as $\omega = [m 2^{-k}, (m+1) 2^{-k})$. The three time–frequency projections appearing in the definition can be expanded using vertical decompositions into tiles as

$$\Pi_{I\times(\omega\oplus 2^{-k})} f = 2^{-k}\langle f, \mathbb{1}_I w_{(m\oplus 1)2^{-k}}\rangle \mathbb{1}_I w_{(m\oplus 1)2^{-k}},$$

$$\Pi_{I\times(2^L\omega)} g = 2^{-k}\sum_{n=0}^{2^L-1}\langle g, \mathbb{1}_I w_{(m2^L\oplus n)2^{-k}}\rangle \mathbb{1}_I w_{(m2^L\oplus n)2^{-k}},$$

$$\Pi_{I\times(2^L\omega\oplus\omega\oplus 2^{-k})} h = 2^{-k}\sum_{n'=0}^{2^L-1}\langle h, \mathbb{1}_I w_{(m2^L\oplus n'\oplus m\oplus 1)2^{-k}}\rangle \mathbb{1}_I w_{(m2^L\oplus n'\oplus m\oplus 1)2^{-k}}.$$

Observe that the integral

$$\int (\Pi_{I\times(\omega\oplus 2^{-k})} f)(\Pi_{I\times(2^L\omega)} g)(\Pi_{I\times(2^L\omega\oplus\omega\oplus 2^{-k})} h)$$

is equal to

$$2^{-2k}\sum_{n=0}^{2^L-1}\langle f, \mathbb{1}_I w_{(m\oplus 1)2^{-k}}\rangle\langle g, \mathbb{1}_I w_{(m2^L\oplus n)2^{-k}}\rangle\langle h, \mathbb{1}_I w_{(m2^L\oplus n\oplus m\oplus 1)2^{-k}}\rangle,$$

since the terms with $n \neq n'$ disappear. That way we arrive at (A.2) once again, completing the proof of $\Lambda^\epsilon(F_0, F_1, F_2) = \Lambda^{\epsilon,L}_{\mathrm{BHT}}(f, g, h)$.

**A.3. Endpoint counterexample.** The observation from the previous section is also useful to explain the failure of some estimates at the boundary of the Banach triangle. By formally taking $L \to \infty$, we are motivated to substitute

$$F_0(x_1, x_2) := f(x_1 \oplus x_2), \quad F_1(x_2, x_0) := h(x_0), \quad F_2(x_0, x_1) := g(x_0),$$

in which case (1.6) becomes

$$\sum_{\vec{I}\in\mathcal{I}} \epsilon_{I_0}|I_0|^{-1}\Big(\iint f(x_1\oplus x_2)\mathrm{h}_{I_1}(x_1)\mathrm{h}_{I_2}(x_2)\,dx_1\,dx_2\Big)\Big(\int g(x_0)h(x_0)\mathrm{h}_{I_0}(x_0)\,dx_0\Big)$$
$$= \sum_{I_0} \epsilon_{I_0}|I_0|^{-1}\langle f, \mathrm{h}_{I_0}\rangle\langle gh, \mathrm{h}_{I_0}\rangle = \int f(x)H^\epsilon(gh)(x)\,dx.$$

Since Haar multipliers are generally not bounded on $L^1$, we see that estimate (1.10) cannot hold when $p_0 = \infty$.

The positive results in this limiting case do not reveal the true structural complexity of $\Lambda^\epsilon$. Indeed, when one of the functions depends on a single

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

variable alone (such as $F_0(x, y) = x$), then the triangle 'breaks' immediately. No techniques from time–frequency analysis are required to bound such degenerate cases, even though they correspond both to the limiting case $a \to \infty$ and to the special case $N \equiv 0$ in Theorem 1.7.

## Appendix B. Real triangular Hilbert transform

In this appendix we show the equivalence of (1.1) and (1.4), and indicate how to obtain the Carleson operator from (1.4).

**B.1. Equivalence of the definitions.** For $\vec{\beta}_0, \vec{\beta}_1, \vec{\beta}_2 \in \mathbb{R}^2$ in general position, consider the change of variables

$$\begin{pmatrix} t \\ u \\ v \end{pmatrix} = B \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}, \quad B := \begin{pmatrix} 1 & 1 & 1 \\ \vec{\beta}_0 & \vec{\beta}_1 & \vec{\beta}_2 \end{pmatrix}.$$

If $\pi_i$ denotes the projection $\pi_i : \mathbb{R}^3 \to \mathbb{R}^2$, $\pi_i(x_0, x_1, x_2) = (x_{i+1}, x_{i-1})$, then, for arbitrary functions $F_0, F_1, F_2$, we have

$$\begin{aligned} \Lambda_\Delta(F_0, F_1, F_2) &= \iiint \prod_{i=0}^{2} F_i(\pi_i(x_0, x_1, x_2)) \frac{1}{x_0 + x_1 + x_2} \, dx_0 \, dx_1 \, dx_2 \\ &= |\det B|^{-1} \iiint \prod_{i=0}^{2} F_i(\pi_i B^{-1}(t, u, v)) \frac{dt}{t} \, du \, dv \\ &= |\det B|^{-1} \iiint \prod_{i=0}^{2} \tilde{F}_i((u, v) - \vec{\beta}_i t) \frac{dt}{t} \, du \, dv \\ &= |\det B|^{-1} \Lambda_{\vec{\beta}_0, \vec{\beta}_1, \vec{\beta}_2}(\tilde{F}_0, \tilde{F}_1, \tilde{F}_2), \end{aligned}$$

where

$$\tilde{F}_i(u, v) := F_i(\pi_i B^{-1}(0, u, v)). \tag{B.1}$$

Here we have used the fact that

$$\pi_i B^{-1} \begin{pmatrix} 1 \\ \vec{\beta}_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The surprising observation is now that

$$\|\tilde{F}_i\|_{p_i} = |\det B|^{1/p_i} \|F_i\|_{p_i}.$$

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

Indeed, the change of variables in the definition of $\tilde{F}_i$ is given by the $2 \times 2$ submatrix of $B^{-1}$ obtained by crossing out the first column and the $i$th row. By Cramer's rule, the determinant of that submatrix equals $\det B^{-1}$ times the $(1, i)$th entry of $B$, up to the sign. Since the latter entry of $B$ is 1, the determinant of the change of variables is $\pm(\det B)^{-1}$. The ratio of $L^{p_i}$ norms equals the absolute value of the determinant to the power $-1/p_i$, as required.

This shows that an $L^{p_0} \times L^{p_1} \times L^{p_2}$ estimate for $\Lambda_\Delta$ cannot be worse than the corresponding estimate for $\Lambda_{\vec{\beta}_0,\vec{\beta}_1,\vec{\beta}_2}$. Running the above argument backwards, we also obtain the converse. Finally, a uniform (in $\vec{\beta}_i$) estimate for $\Lambda_{\vec{\beta}_0,\vec{\beta}_1,\vec{\beta}_2}$ with $\vec{\beta}_i$ in general position implies the same estimate for the $\vec{\beta}_i$ lying on a line by a limiting argument.

**B.2. Less singular two-dimensional forms.** The trilinear forms introduced in [DT10] can be written as

$$\Lambda^K_{B_0,B_1,B_2}(F_0, F_1, F_2) := \iint_{\mathbb{R}^2} \text{p.v.} \int_{\mathbb{R}^2} \prod_{i=0}^{2} F_i(\vec{x} - B_i\vec{t})K(\vec{t})\, d\vec{t}\, d\vec{x},$$

where $B_1, B_2, B_3$ are now $2 \times 2$ real matrices (interpreted as linear operators on $\mathbb{R}^2$) and $K$ is a two-dimensional Calderón–Zygmund kernel. If $K$ is odd and homogeneous of degree $-2$, then it takes the form

$$K(r\cos\theta, r\sin\theta) = \frac{\Omega(\theta)}{r^2},$$
$$\Omega(\theta + \pi) = -\Omega(\theta) \quad \text{for } 0 < r < \infty, \theta \in \mathbb{R}/(2\pi\mathbb{Z}).$$

Observe that

$$\begin{aligned} &\text{p.v.} \int_{\mathbb{R}^2} \prod_{i=0}^{2} F_i\big(\vec{x} - B_i\vec{t}\big)K(\vec{t})\, d\vec{t} \\ &\quad = \text{p.v.} \int_0^{2\pi} \int_0^{\infty} \prod_{i=0}^{2} F_i\big(\vec{x} - B_i(r\cos\theta, r\sin\theta)\big)\frac{\Omega(\theta)}{r^2} r\, dr\, d\theta \\ &\quad = \int_0^{\pi} \Omega(\theta)\, \text{p.v.} \int_{\mathbb{R}} \prod_{i=0}^{2} F_i\big(\vec{x} - rB_i(\cos\theta, \sin\theta)\big)\frac{dr}{r}\, d\theta, \end{aligned}$$

so

$$\Lambda^K_{B_0,B_1,B_2} = \int_0^{\pi} \Omega(\theta)\, \Lambda_{B_0(\cos\theta,\sin\theta),B_1(\cos\theta,\sin\theta),B_2(\cos\theta,\sin\theta)}\, d\theta;$$

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

that is, $\Lambda^{K}_{B_0,B_1,B_2}$ is a 'superposition' of the forms (1.1). Consequently, $L^p$ estimates for all cases of the matrices studied in [**DT10**] and the remaining case from [**Kov12**] would follow from a single estimate for (1.4), even uniformly over all choices of $B_0$, $B_1$, $B_2$.

**B.3. Carleson maximal operator.** Analogously to the dyadic case, but with an additional smooth cutoff, we consider

$$\begin{aligned} F(x,y) &:= f(-x-y)\, D^p_L\phi(x),\\ G(y,z) &:= e_{N(z)}(y)\operatorname{sgn} g(z)\sqrt{|g(z)|}\, D^{2p'}_L\phi(y+z),\\ H(z,x) &:= e_{N(z)}(z+x)\sqrt{|g(z)|}\, D^{2p'}_L\phi(x) \end{aligned}$$

for $f \in L^p(\mathbb{R})$ and $g \in L^{p'}(\mathbb{R})$, $2 < p < \infty$, where $e_N(x) = e^{2\pi i N x}$, $N$ is a measurable linearizing function for the Carleson operator, $\phi$ is a smooth positive function with compact support, and $D^p_L\phi(x) = L^{-1/p}\phi(x/L)$. Then

$$\begin{aligned} \Lambda_\Delta(F,G,H) = \iiint & f(-x-y)e_{N(z)}(x+y+z)g(z)\\ & \times D^p_L\phi(x)D^{2p'}_L\phi(x)D^{2p'}_L\phi(y+z)\frac{d(x,y,z)}{x+y+z}. \end{aligned}$$

The change of variables $t = x+y+z$ gives

$$\iiint f(z-t)e_{N(z)}(t)g(z)D^p_L\phi(x)D^{2p'}_L\phi(x)D^{2p'}_L\phi(t-x)\frac{d(x,z,t)}{t},$$

which converges to a constant times

$$\iint \frac{1}{t} f(z-t)e_{N(z)}(t)g(z)\, d(z,t)$$

as $L \to \infty$, and yields an $L^p$ bound for

$$(C_N f)(z) = \text{p.v.} \int_{\mathbb{R}} f(z-t)e^{2\pi i N(z) t}\frac{dt}{t}.$$

**B.4. Multilinear generalization.** For any positive integer $n$, one can also consider the straightforward $(n+1)$-linear generalization of (1.1) given by

$$\Lambda_{\vec{\beta}_0,\vec{\beta}_1,\ldots,\vec{\beta}_n}(F_0,F_1,\ldots,F_n) := \int_{\mathbb{R}^n} \text{p.v.} \int_{\mathbb{R}} \prod_{j=0}^{n} F_j(\vec{x}-\vec{\beta}_j t)\frac{dt}{t}\, d\vec{x}, \qquad \text{(B.2)}$$

where this time $\vec{\beta}_j \in \mathbb{R}^n$ and the functions $F_j$ are $n$-dimensional. This object is expected to be even more difficult, as conjectured bounds for the one-dimensional trilinear Hilbert transform

$$\Lambda_{3\mathrm{HT}}(f_0, f_1, f_2, f_3) := \int_{\mathbb{R}} \mathrm{p.v.} \int_{\mathbb{R}} f_0(x) f_1(x-t) f_2(x-2t) f_3(x-3t) \frac{dt}{t}\, dx \tag{B.3}$$

and its variants would follow from bounds for (B.2) when $n = 3$. No positive results are known for this operator; see [Dem08] for some negative results. However, an interesting observation is that the linearized polynomial Carleson operator

$$(C_{N_1,\ldots,N_{n-1}} f)(x) = \mathrm{p.v.} \int_{\mathbb{R}} f(x-t) e^{i(N_1(x)t + N_2(x)t^2 + \cdots + N_{n-1}(x)t^{n-1})} \frac{dt}{t}$$

can be encoded into (B.2). Let $\vec{\beta}_0$ be the origin, and let $\vec{\beta}_1, \ldots, \vec{\beta}_n$ constitute the standard basis for $\mathbb{R}^n$. The identity

$$\sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^m = \begin{cases} 0 & \text{for } m = 0, 1, \ldots, k-1, \\ k! & \text{for } m = k \end{cases}$$

can be shown easily by induction on a positive integer $k$, and its immediate consequence is that

$$\sum_{j=0}^{n-1} \sum_{k=\max\{j,1\}}^{n-1} (-1)^j \frac{1}{k!} \binom{k}{j} N_k(x_n) \Big( \sum_{l=1}^{k} l x_l - jt \Big)^k = \sum_{k=1}^{n-1} N_k(x_n) t^k.$$

It follows that, with

$$F_j(x_1, \ldots, x_n) = g_j(x_n) \prod_{k=\max\{j,1\}}^{n-1} \exp\Big( i(-1)^j \frac{1}{k!} \binom{k}{j} N_k(x_n) \Big( \sum_{l=1}^{k} l x_l \Big)^k \Big)$$

for $j = 0, 1, \ldots, n-1$ and

$$F_n(x_1, \ldots, x_n) = f(x_n),$$

the form (B.2) formally becomes $\int (C_{N_1,\ldots,N_{n-1}} f) g_0 g_1 \cdots g_{n-1}$. To be precise we should also include appropriate cutoffs of the functions $F_j$, similarly as we did in the previous subsection. It would be interesting to investigate which particular cases of $\Lambda_{\vec{\beta}_0, \vec{\beta}_1, \ldots, \vec{\beta}_n}$ can be resolved using the techniques from [Lie09] and [Lie11].

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press

## References

[Ber12] F. Bernicot, 'Fiber-wise Calderón–Zygmund decomposition and application to a bi-dimensional paraproduct', *Illinois J. Math.* **56**(2) (2012), 415–422.

[Bil67] P. Billard, 'Sur la convergence presque partout des séries de Fourier–Walsh des fonctions de l'espace $L^2(0, 1)$', *Studia Math.* **28** (1966/1967), 363–388.

[Dem08] C. Demeter, 'Divergence of combinatorial averages and the unboundedness of the trilinear hilbert transform', *Ergod. Th. & Dynam. Sys.* **28**(5) (2008), 1453–1464.

[Dem15] C. Demeter, 'A guide to Carleson's Theorem', *Rocky Mountain J. Math.* **45**(1) (2015), 169–212.

[DT10] C. Demeter and C. Thiele, 'On the two-dimensional bilinear Hilbert transform', *Amer. J. Math.* **132**(1) (2010), 201–256.

[GL04] L. Grafakos and X. Li, 'Uniform bounds for the bilinear Hilbert transforms I', *Ann. of Math.* (2) **159**(3) (2004), 889–933.

[Kov12] V. Kovač, 'Boundedness of the twisted paraproduct', *Rev. Mat. Iberoam.* **28**(4) (2012), 1143–1164.

[LT97] M. Lacey and C. Thiele, '$L^p$ estimates on the bilinear Hilbert transform for $2 < p < \infty$', *Ann. of Math.* (2) **146**(3) (1997), 693–724.

[LT99] M. Lacey and C. Thiele, 'On Calderón's conjecture', *Ann. of Math.* (2) **149**(2) (1999), 475–496.

[Li06] X. Li, 'Uniform bounds for the bilinear Hilbert transforms II', *Rev. Mat. Iberoam.* **22**(3) (2006), 1069–1126.

[Lie09] V. Lie, 'The (weak-$l^2$) boundedness of the quadratic carleson operator', *Geom. Funct. Anal.* **19**(2) (2009), 457–497.

[Lie11] V. Lie, 'The polynomial carleson operator', Preprint, 2011, arXiv:1105.4504.

[OT11] R. Oberlin and C. Thiele, 'New uniform bounds for a Walsh model of the bilinear Hilbert transform', *Indiana Univ. Math. J.* **60**(5) (2011), 1693–1712.

[Thi95] C. M. Thiele, 'Time-frequency analysis in the discrete phase plane', PhD Thesis, Yale University. ProQuest LLC, Ann Arbor, MI, 1995.

https://doi.org/10.1017/fms.2015.25 Published online by Cambridge University Press