# $\mathrm{L}^p$ estimates for entangled multilinear forms

Polona Durcik

20th February 2014
Revised version 2nd December 2014

Master's Thesis Mathematics

Advisor: Prof. Dr. Christoph Thiele

Mathematical Institute

Mathematisch-Naturwissenschaftliche Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

**Abstract**

We prove $L^p$ estimates for a class of dyadic multilinear forms associated to the unit cube in $\mathbb{R}^n$. These forms act on $n$-dimensional functions and can be seen as a higher dimensional generalization of the twisted paraproduct. The structure of the latter differs from the one of the classical paraproduct in that the functions involved are entangled in such a way that they depend on the same one-dimensional variables. This thesis is based on the approach of Kovač [7], who recently obtained boundedness for the twisted paraproduct.

# Contents

**1 Introduction** **3**

**2 Preliminaries** **7**
2.1 Dyadic cubes . . . . . 7
2.2 Interpolation . . . . . 9
2.3 Dyadic maximal function . . . . . 11
2.4 Haar basis . . . . . 12

**3 Proof of Theorem 1.1** **15**
3.1 Determining $d_j$ . . . . . 17
3.2 A single tree estimate . . . . . 20
3.3 Completing the proof . . . . . 26

**4 Examples** **29**

**5 Final remarks** **38**
5.1 A simpler proof if $V = V_n,\ p_j = 2^n$ . . . . . 38
5.2 Discussion on the exponent range . . . . . 40

# 1 Introduction

The *continuous twisted paraproduct* is a two-dimensional bilinear operator proposed by Demeter and Thiele [3] while investigating a two-dimensional analogue of the bilinear Hilbert transform. It is given by

$$T_{\mathrm{c}}(F,G)(x,y)=\sum_{k\in\mathbb{Z}}2^{2k}\Big(\int_{\mathbb{R}}F(x-s,y)\varphi(2^ks)ds\Big)\Big(\int_{\mathbb{R}}G(x,y-t)\psi(2^kt)dt\Big)$$

where $\varphi$ is an arbitrary Schwartz function and $\psi$ is a Schwartz function satisfying $\mathrm{supp}(\hat{\psi})\subseteq\{\xi\in\mathbb{R}:1/2\leq|\xi|\leq 2\}$.

For $k\in\mathbb{Z}$ we denote $\varphi_k(t)=2^k\varphi(2^kt)$ and $\psi_k(t)=2^k\psi(2^kt)$. Let $P_\varphi$ be the one-dimensional convolution operator, $P_\varphi f:=f*\varphi$. When we apply the convolution operator in only one variable of a two-dimensional function we denote that variable with a superscript, for instance,

$$(P_\varphi^{(1)}F)(x,y):=(P_\varphi F_y)(x)$$

where $F_y(x):=F(x,y)$. Then, the twisted paraproduct can be alternatively written as

$$T_{\mathrm{c}}(F,G)=\sum_{k\in\mathbb{Z}}(P_{\varphi_k}^{(1)}F)(P_{\psi_k}^{(2)}G)$$

Its dyadic variant, the *dyadic twisted paraproduct* is defined as

$$T_{\mathrm{d}}(F,G):=\sum_{k\in\mathbb{Z}}(\mathbb{E}_k^{(1)}F)(\Delta_k^{(2)}G)\tag{1.1}$$

where the martingale averages and differences are defined as

$$\mathbb{E}_kf:=\sum_{|I|=2^k}|I|^{-1}\int_If(t)dt\,\mathbf{1}_I\quad\text{and}\quad\Delta_kf:=\mathbb{E}_{k-1}f-\mathbb{E}_kf\tag{1.2}$$

respectively. The sum in (1.2) runs over all dyadic intervals in $\mathbb{R}$ of size $2^k$. We write $\mathbf{1}_A$ for the characteristic function of a set $A\subseteq\mathbb{R}^n$. The superscript above the average and difference operators in (1.1) has the same meaning as in the continuous case.

In [7], Kovač proved the boundedness result[1]

$$\|T_{\mathrm{c}}(F,G)\|_{L^{pq/(p+q)}(\mathbb{R}^2)}\lesssim_{p,q}\|F\|_{L^p(\mathbb{R}^2)}\|G\|_{L^q(\mathbb{R}^2)}$$

whenever $1<p,q<\infty$, $\frac{1}{p}+\frac{1}{q}>\frac{1}{2}$. His approach was to bound the dyadic version $T_{\mathrm{d}}$ first and then relate the continuous version to the dyadic version. To do that it was more convenient to dualize and consider the corresponding trilinear form

$$\Lambda_{\mathrm{d}}(F,G,H):=\int_{\mathbb{R}^2}T_{\mathrm{d}}(F,G)(x,y)H(x,y)dxdy$$

[1]For two nonnegative quantities $A$ and $B$ we write $A\lesssim B$ if there exists an absolute constant $C>0$ such that $A\leq CB$. We write $A\lesssim_P B$ if the constant depends on a set of parameters $P$. We write $A\sim_P B$ if both $A\lesssim_P B$ and $B\lesssim_P A$.

As $\mathbb{E}_k$ and $\Delta_k$ can be represented in the Haar basis, $\Lambda_{\mathrm{d}}$ can be rewritten as

$$\Lambda_{\mathrm{d}}(F,G,H) = \sum_{I\times J}\int_{\mathbb{R}^4} F(u,y)G(x,v)H(x,y) \chi_I(u)\chi_I(x)h_J(v)h_J(y)dudxdvdy \tag{1.3}$$

with the sum over all dyadic squares $I \times J$ in $\mathbb{R}^2$. The scaling function and the Haar function adapted to a dyadic interval $I$ are defined as

$$\chi_I := |I|^{-1/2}\mathbf{1}_I \text{ and } h_I := |I|^{-1/2}\left(\mathbf{1}_{I_{\text{left half}}} - \mathbf{1}_{I_{\text{right half}}}\right)$$

These functions can be seen as discrete analogues of the Schwartz functions. The bounds that were established are

$$|\Lambda_{\mathrm{d}}(F,G,H)| \lesssim_{p,q,r} \|F\|_{L^p(\mathbb{R}^2)}\|G\|_{L^q(\mathbb{R}^2)}\|H\|_{L^r(\mathbb{R}^2)}$$

in the range $2 < p,q,r < \infty$, $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$. The fiber-wise Calderon-Zygmund decomposition by Bernicot [2] extended the range to $1 < p$, $q < \infty$, $2 < r \leq \infty$.

The operator $T_{\mathrm{c}}$ (or $T_{\mathrm{d}}$) is called *twisted*, since there is a twist in the variables in which the convolutions (or martingale projections) are taken. Another reason for its name is that the functions in $\Lambda_{\mathrm{d}}$ are entangled in the way that the form does not split into Haar coefficients of each of the functions separately, as it does for classical paraproducts. Following the definitions in [8] and [9], a two-dimensional multilinear form $\Lambda$ is called a *(dyadic) classical paraproduct* if it is given by[2]

$$\Lambda(F_1,\dots,F_n) = \sum_{I\times J}|I|^{2-n}\Big(\prod_{i\in S}\langle F_i, h_I\otimes\chi_J\rangle_{L^2(\mathbb{R}^2)}\Big)\Big(\prod_{i\in S^c}\langle F_i,\chi_I\otimes\chi_J\rangle_{L^2(\mathbb{R}^2)}\Big)$$

where $(f\otimes g)(x,y) := f(x)g(y)$, the sum runs over all dyadic squares $I\times J$ in $\mathbb{R}^2$, $S\subseteq\{1,\dots,n\}$, $|S|\geq 2$ and $S^c = \{1,\dots,n\}\setminus S$. The estimate

$$|\Lambda(F_1,\dots,F_n)| \lesssim_{n,(p_j)} \prod_{j=1}^n \|F_j\|_{L^{p_j}(\mathbb{R}^2)}$$

whenever $\sum_{j=1}^n \frac{1}{p_j} = 1$ and $1 < p_j < \infty$ can be established with the tools of classical Calderón-Zygmund theory. See [9] for more details. The splitting into Haar coefficients allows an application of the Cauchy-Schwarz inequality, which immediately yields operators such as maximal or square functions. In the twisted case this approach fails, as we have the dependence of several functions on the same one-dimensional variable.

At this point let us also remark that the form $\Lambda_{\mathrm{d}}$ has a certain general modulation symmetry. Namely, as the functions $G$ and $H$ in (1.3) both depend on the variable $x$, for any $g\in L^\infty(\mathbb{R})$ we have

$$\Lambda_{\mathrm{d}}(F,(g\otimes 1)G,H) = \Lambda_{\mathrm{d}}(F,G,(g\otimes 1)H)$$

[2]This is a two-dimensional dyadic formulation in order to reflect its similarity and difference to $\Lambda_{\mathrm{d}}$. Paraproducts in other dimensions can be defined analogously by pairing with one-dimensional or higher dimensional Haar functions. Continuous versions can be defined replacing each $\chi_I$ and $h_I$ by suitable bump functions. We refer to [9] for definitions on $\mathbb{R}$.

This invariance is closely related with failure of the classical theory when addressing $\Lambda_{\mathrm{d}}$.

The approach of Kovač is based on the Bellman function technique and preserves the symmetries stemming form the entangled structure. The approach consists of alternating applications of the Cauchy-Schwarz inequality, which reduces the complexity of the involved forms, and a telescoping argument, which relates several multi-scaled quantities to a simpler, single-scaled expression. Kovač further developed his method in [8], where he discusses the multilinear forms associated to finite bipartite undirected graphs, generalizing both the twisted and classical paraproduct. A brief note on the higher dimensional theory can be found at the end of [7].

In this thesis we discuss a higher dimensional generalization of the twisted paraproduct in greater detail by considering multilinear forms associated to unit (hyper)cubes. We work in the dyadic setting only.

For any $V \subseteq V_n := \{0,1\}^n$ we define a multilinear form $\Lambda = \Lambda_V$ acting on $|V|$ functions $F_j : \mathbb{R}^n \to \mathbb{C}$ by

$$\begin{aligned}\Lambda((F_j)_{j\in V}) \quad := \quad & \sum_{Q \text{ dyadic cube}} \int_{\mathbb{R}^{2n}} \Big( \prod_{j\in V} F_j(x_1^{j_1}, \dots, x_n^{j_n}) \Big) \\ & \Big( \prod_{i=1}^{n-1} \chi_{I_i}(x_i^0)\chi_{I_i}(x_i^1) \Big) h_{I_n}(x_n^0) h_{I_n}(x_n^1) dx_1^0 \dots dx_n^0 \, dx_1^1 \dots dx_n^1\end{aligned}$$

where $j = (j_1, \dots, j_n)$ and $Q = I_1 \times \cdots \times I_n$ is a dyadic cube in $\mathbb{R}^n$. The functions $F_j$ are assumed to belong to a certain class of test functions which we will specify later. We prove the following bounds.

**Theorem 1.1.** *Let $V \subseteq V_n$. The form $\Lambda$ satisfies the estimate*

$$|\Lambda((F_j)_{j\in V})| \lesssim_{n,(p_j)} \prod_{j\in V} \|F_j\|_{L^{p_j}(\mathbb{R}^n)} \tag{1.4}$$

*whenever the exponents $(p_j)$ are such that $\sum_{j\in V} \frac{1}{p_j} = 1$ and $d_j < p_j < \infty$ for all $j \in V$. The number $d_j$ is contained in $\{0, 2^0, 2^1, \dots, 2^{n-1}\}$ and is defined in* (3.2) *and* (3.3).

The lower bound $d_j$ can be explicitly calculated for every vertex set $V$ and depends on the structure of $V$. Roughly speaking, the more vertices lie on the same faces as $j$, the higher is $d_j$. Since $d_j$ is at most $2^{n-1}$, if $|V| > 2^{n-1}$ one certainly finds exponents $(p_j)$ with $\sum_{j\in V} \frac{1}{p_j} = 1$. There are also examples of sets with $|V| > 2^{n-1}$ for which the theorem yields an improved range with $d_j < 2^{n-1}$. An instance of such a set in $\mathbb{R}^3$ is

$$\{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$$

for which $d_{(0,0,0)} = d_{(1,0,0)} = d_{(0,1,0)} = 4$ and $d_{(0,0,1)} = d_{(1,1,1)} = 2$.

For vertex sets of smaller cardinalities we have to ensure that some of the exponents are strictly below $2^{n-1}$, in order to have a non-empty range. An examination of (3.2) and (3.3) shows that this is always possible if $n = 2$. If $n = 3$, we are able to find a

non-emtpy range for any $V \subseteq V_3$ as well. For all but one symmetry class of sets this follows from Theorem 1.1. The exception is

$$\{(0,0,0),(1,0,0),(0,0,1),(0,1,1)\} \tag{1.5}$$

for which $d_j = 4$ for all vertices $j$. However, this case can be reduced to two other cases, for which Theorem (1.1) yields a good bound. For more discussion and possible extensions of the range of exponents as well as for the list of $d_j$'s for any three-dimensional vertex set we refer to Section 5.

The numbers $d_j$ tend to yield an empty range more frequently when proceeding to higher dimensions. For instance, if $n = 4$ this happens for the form associated to

$$\{(0,0,0,0),(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\}$$

where we obtain $d_{(0,0,0,1)} = 2$, while $d_j = 8$ for all $j \neq (0,0,0,0)$. The reduction used in (1.5) do not directly apply as well.

To deal with a more symmetric situation, the approach of Kovač [7] is to consider forms related to the twisted paraproduct involving 4 functions. Then the author gradually separates the functions using the Cauchy-Schwarz inequality and a telescoping identity, which switches between the variables. This procedure yields several related multilinear forms at each step. A straightforward generalization of this method to $\mathbb{R}^n$ is to consider forms involving $2^n$ functions. Such an approach yields $d_j = 2^{n-1}$ for all $j \in V$, independently of the structure of $V$. As we will see in Section 3 where we discuss the structure of $\Lambda$ more thoroughly, symmetrizing the form by adding functions corresponds to symmetrizing the vertex set $V$ by adding certain vertices. For example, the twisted paraproduct is the form associated to

$$\{(0,0),(1,0),(0,1)\} \tag{1.6}$$

A more symmetric problem is to include the vertex $(1,1)$ in (1.6), which results in $V_2$.

However, to find a larger range of exponents we do not symmetrize our vertex set at once. We rather do that gradually and at each step we add only as many new vertices as needed. In the end we still end up with forms associated to symmetric sets, meaning that each vertex $j$ is paired with its negative $(1,\dots,1) - j$, but our effort is that these final sets are as small as possible. To find the smallest symmetrization we search for the minimal number of applications of the Cauchy-Schwarz inequality needed to separate the involved functions. This number is often lower than the one needed to separate the functions in the forms associated to $V_n$, since smaller vertex sets make the forms less entangled and more closely related to lower dimensional problems. In certain degenerate cases we can even derive the estimate directly from the estimate for lower dimensional forms. This is possible, for instance, if $V$ equals

$$\{(0,0,0),(0,0,1)\}$$

where none of the vertices has a "1" in the first two components and the problem is essentially a one-dimensional problem.

A natural question is whether it is possible to extend the range beyond $d_j$. In certain instances one can proceed via fiber-wise Calderón-Zygmund decomposition as for the twisted paraproduct. However, applying this argument is in general not possible, as our forms do not have the necessary structure. We elaborate on this in Section 5, where we also consider a three-dimensional example which allows such a decomposition.

In the Section 5 we also discuss a simpler approach to the problem, which yields an estimate in only one point. However, often this establishes exactly the endpoint case, when for at least some of the exponents $p_j = d_j$. For instance, for the twisted paraproduct, $d_j = 2$ for all $j$. But with the simpler procedure we are able to obtain the estimates for $p_{(0,0)} = p_{(1,0)} = 4$, $p_{(0,1)} = 2$ and $p_{(0,0)} = p_{(0,1)} = 4$, $p_{(1,0)} = 2$. The same method also establishes the endpoint case $p_j = 4$ for all $j \in V$ when $V$ equals (1.5).

This shows that the range of exponents in Theorem 1.1 is not maximal. It is an unresolved problem to determine the full range for which (1.4), as well as the estimates obtained in [8], hold.

This thesis is structured in the following way. Section 2 contains some background material. In Section 3 we prove Theorem 1.1. We perform the most importants steps of the proof on several concrete examples in Section 4. In Section 5 we sketch a simpler proof in the case of $\Lambda_{V_n}$ and $p_j = 2^n$ for all $j \in V_n$. Finally we give a couple of remarks on possible improvements of the exponent range in certain cases and state the numbers $d_j$ for all possible subsets of $V_3$ up to symmetries.

**Acknowledgements.** I would like to express my gratitude to my advisor Prof. Christoph Thiele for his guidance, motivating discussions and many helpful comments on several drafts of this thesis. I would like to thank Joris for reading and providing me with useful remarks. For reading this thesis I am also grateful to Prof. Herbert Koch.

# 2 Preliminaries

In this part we discuss dyadic cubes, interpolation theorems, dyadic maximal function and Haar basis. We refer to [4] and [5].

## 2.1 Dyadic cubes

A *dyadic interval* in $\mathbb{R}$ is an interval of the form

$$[2^k m, 2^k(m+1))$$

with $k, m \in \mathbb{Z}$. The *length* $\ell(I)$ of a dyadic interval $I$ is $2^k$ for $k \in \mathbb{Z}$. The number $k$ is called the *scale* of the interval. We consider half open intervals, so that different dyadic intervals of the same scale are always disjoint.

A *dyadic cube* in $\mathbb{R}^n$ is a product of dyadic intervals of the same scale, i.e. it is of the form

$$\prod_{i=1}^{n} [2^k m_i, 2^k(m_i+1))$$

for $m_1, \ldots, m_n, k \in \mathbb{Z}$. By a *dyadic cuboid* in we mean a product of dyadic intervals (not necessarily of the same scale). We define the *length* $\ell(Q)$ of a dyadic cube $Q = I_1 \times \cdots \times I_n$ to be its side length $\ell(I_j)$ and we write $|Q|$ for its Lebesgue measure.

For $k \in Z$ we denote by $\mathcal{D}_k$ the set of all dyadic cubes in $\mathbb{R}^n$ of length $2^k$, also called the *k-th generation*. By $\mathcal{D} = \mathcal{D}(\mathbb{R}^n)$ we denote the set of all dyadic cubes in $\mathbb{R}^n$. Thus, $\mathcal{D} = \cup_{k \in Z} \mathcal{D}_k$ and each dyadic cube is contained in a unique $\mathcal{D}_k$. If $n = 1$, we write $\mathcal{I} = \mathcal{D}(\mathbb{R}^1)$ for the collection of all dyadic intervals in $\mathbb{R}$ and $\mathcal{I}_k$ for their $k-$th generation.

The following proposition collects some properties of dyadic cubes.

**Proposition 2.1.** *(1) For every $k \in \mathbb{Z}$, $\mathcal{D}_k$ forms a partition of $\mathbb{R}^n$.*

*(2) Each dyadic cube is the union of $2^n$ pairwise disjoint dyadic cubes of half the side length, called the children of the cube.*

*(3) Given any two dyadic cubes, they are either disjoint or one is contained in the other.*

*(4) Each dyadic cube has a unique parent.*

*(5) Let $\mathcal{C}$ be any collection of dyadic cubes with $\sup_{Q \in \mathcal{C}} \ell(Q) < \infty$ and let $\mathcal{C}_{\max}$ be the the collection of maximal cubes in $\mathcal{C}$ with respect to the set inclusion. The cubes in $\mathcal{C}_{\max}$ are pairwise disjoint and $\bigcup_{Q' \in \mathcal{C}_{\max}} Q' = \bigcup_{Q \in \mathcal{C}} Q$.*

*Proof.* Note that the set of all dyadic cubes is invariant under dilation $x \mapsto 2x$ of $\mathbb{R}^n$ and that the set of all dyadic cubes of scale at most 0 is invariant under translation $x \mapsto x + y$ for $y \in \mathbb{Z}^n$.

(1) By dilation invariance it suffices to prove that each point $x \in \mathbb{R}^n$ is contained in exactly one cube of length 1. Let $m_i$ be the greatest integer which is smaller or equal to $x_i$. Then $\prod_{i=i}^{m} [m_i, m_i + 1)$ is the unique cube of length 1 containing $x$.

(2) By translation and dilation invariance it suffices to show this for $Q = [0, 1)^n$. For $n = 1$, the children of $I = [0, 1)$ are $I_l = [0, \frac{1}{2})$ and $I_r = [\frac{1}{2}, 1)$. For $n > 1$, the children of $Q$ are given by $I_{k_1} \times \cdots \times I_{k_n}$, $k_j \in \{l, r\}$.

(3) Let $Q_1, Q_2 \in \mathcal{D}$. Without loss of generality we may assume that $|Q_1| \geq |Q_2|$ and thus $|Q_1| = 2^{nj}|Q_2|$ for some $j \in \mathbb{N}_0$. We prove the claim by induction on $j$. If $j = 0$, $Q_1$ and $Q_2$ belong to the same generation and we are done by (1). Now assume that $j > 0$ and that we have proven the claim for $Q_1, Q_2$ with $|Q_1| = 2^{n(j-1)}|Q_2|$. Consider a pair of cubes with $|Q_1| = 2^{nj}|Q_2|$. If $Q_1 \cap Q_2 \neq \emptyset$, let $x \in Q_1 \cap Q_2$. Let $Q_1'$ be the child of $Q_1$ which contains $x$. By the induction hypothesis, $Q_2 \subseteq Q_1'$ and so $Q_2 \subseteq Q_1$, which finishes the proof.

(4) Let $Q$ be a dyadic cube and let $x \in Q$. By (1), $x$ is contained in a unique dyadic cube $Q'$ with $\ell(Q') = 2\ell(Q)$. Since $Q$ intersects at least one of the children of $Q'$, by (3) $Q$ has to equal that child.

(5) Any two different cubes in $\mathcal{C}_{\max}$ are disjoint by (3), since by maximality neither of them is contained in the other. We have $\bigcup_{Q'\in\mathcal{C}_{\max}} Q' \subseteq \bigcup_{Q\in\mathcal{C}} Q$. Now let $x \in \bigcup_{Q\in\mathcal{C}} Q$. Since we have an upper bound on the length of cubes, there exists a maximal dyadic cube containing $x$. Thus $x \in \bigcup_{Q'\in\mathcal{C}_{\max}} Q'$.

□

We will always denote the left and the right child of a dyadic interval $I$ by $I_l$ and $I_r$, respectively.

**Definition 2.1** (Convex tree)**.** A collection of dyadic cubes $\mathcal{T}$ in $\mathbb{R}^n$ is called a *tree* if there exists $Q_{\mathcal{T}} \in \mathcal{T}$ called the *root* of the $\mathcal{T}$ satisfying $Q \subseteq Q_{\mathcal{T}}$ for every $Q \in \mathcal{T}$. A tree is called *convex* if for all dyadic cubes $Q_1$, $Q_2$, $Q_3$, $Q_1 \subseteq Q_2 \subseteq Q_3$ and $Q_1$, $Q_3 \in \mathcal{T}$ imply $Q_2 \in \mathcal{T}$. A *leaf* of $\mathcal{T}$ is a dyadic cube which is not contained in $\mathcal{T}$ but its parent is. We denote the set of leaves of $\mathcal{T}$ by $\mathcal{L}(\mathcal{T})$.

Informally, convex trees "do not skip any scales". From Proposition 2.1 it follows that the leaves of any finite convex tree $\mathcal{T}$ partition the root $Q_{\mathcal{T}}$.

## 2.2 Interpolation

Let $(X, \mu)$ be a measure space and $f : X \to \mathbb{C}$ measurable. We call the function $d_f : (0, \infty) \to [0, \infty]$, given by

$$d_f(\lambda) := \mu(\{x \in X : |f(x)| > \lambda\})$$

the *distribution function of* $f$ (associated to $\mu$).

The following properties follow from the definition.

**Lemma 2.2.** *For a function $f$ in $L^p(X, \mu)$, $0 < p < \infty$, we have*

*(1)*

$$\|f\|_{L^p}^p = p \int_0^\infty \lambda^{p-1} d_f(\lambda) d\lambda$$

*(2)*

$$\|f\|_{L^p}^p \sim_p \sum_{k\in\mathbb{Z}} 2^{kp} d_f(2^k)$$

*Proof.* (1) By Fubini's theorem,

$$\begin{aligned} p \int_0^\infty \lambda^{p-1} d_f(\lambda) d\lambda &= p \int_0^\infty \lambda^{p-1} \int_X \mathbf{1}_{\{x:|f(x)|>\lambda\}} d\mu(x) d\lambda \\ &= \int_X \int_0^{|f(x)|} p\lambda^{p-1} d\lambda d\mu(x) \\ &= \int_X |f(x)|^p d\mu(x) = \|f\|_{L^p}^p \end{aligned}$$

(2) Since $d_f$ is non-increasing,

$$\begin{aligned} p\int_0^\infty \lambda^{p-1} d_f(\lambda)d\lambda &= p\sum_{k\in\mathbb{Z}} \int_{2^{k-1}}^{2^k} \lambda^{p-1} d_f(\lambda)d\lambda \\ &\geq p\sum_{k\in\mathbb{Z}} 2^{(k-1)(p-1)} d_f(2^k)(2^k - 2^{k-1}) \\ &= \frac{p}{2^p}\sum_{k\in\mathbb{Z}} 2^{kp} d_f(2^k) \end{aligned}$$

Thus $\sum_{k\in\mathbb{Z}} 2^{kp} d_f(2^k) \lesssim_p \|f\|_{L^p}^p$. For the other direction we proceed similarly.

□

Let us denote by $M(X,\mu)$ the set of all measurable function $f : X \to \mathbb{C}$. For $0 < p < \infty$ we define the *weak* $L^p(X,\mu)$ space, denoted $L^{p,\infty}(X,\mu)$, as the set of all $f \in M(X,\mu)$ for which

$$\|f\|_{L^{p,\infty}(X,\mu)} := \sup_{\lambda>0} \lambda\, d_f(\lambda)^{1/p}$$

is finite. It holds that $\|\cdot\|_{L^{p,\infty}}$ is a quasi norm. Two functions in $L^{p,\infty}(X,\mu)$ are considered equal if they agree $\mu-$almost everywhere. The *weak* $L^\infty(X,\mu)$ space is by definition $L^\infty(X,\mu)$.

Let $(Y,\nu)$ be another measure space and let $T$ be an operator from $L^p(X,\mu)$ to $M(Y,\nu)$. We say that $T$ is of *weak* $(p,q)$ *type*, $0 < p,q \leq \infty$, if there exists a constant $C > 0$ such that

$$\|Tf\|_{L^{q,\infty}} \leq C\,\|f\|_{L^p}$$

for all $f \in L^p(X,\mu)$.

We call an operator $T$ mapping from the vector space of all measurable functions on $(X,\mu)$ to $M(Y,\nu)$ *sublinear*, if for all $f,g$ and all $\lambda \in \mathbb{C}$

$$\begin{aligned} |T(f+g)(x)| &\leq |Tf(x)| + |Tg(x)| \\ |T(\lambda f)| &= |\lambda||Tf| \end{aligned}$$

**Theorem 2.3** (Marcinkiewicz interpolation)**.** *Let $(X,\mu)$ and $(Y,\nu)$ be measure spaces and $1 \leq p_0 < p_1 \leq \infty$. Let $T$ be a sublinear operator from $L^{p_0}(X,\mu) + L^{p_1}(X,\mu)$ to $M(Y,\nu)$ which is of weak $(p_0,p_0)$ and weak $(p_1,p_1)$ type. Then there exists a constant $C = C_p > 0$ such that for all $p_0 < p < p_1$ and all $f \in L^p(X,\mu)$*

$$\|Tf\|_{L^p} \leq C\,\|f\|_{L^p}$$

The proof can be found in [4] or [5]. Another interpolation theorem which is between strong type estimates is the following

**Theorem 2.4** (Riesz-Thorin)**.** *Let $(X,\mu)$ and $(Y,\nu)$ be measure spaces and let $1 \leq p_0, p_1, q_0, q_1 \leq \infty$. For $0 < \theta < 1$ define $p$ and $q$ by*

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}$$

*Let $T$ be a linear operator from $L^{p_0}(X,\mu)+L^{p_1}(X,\mu)$ to $M(Y,\nu)$. Assume that there exist positive constants $C_0$, $C_1$ such that*

$$\|Tf\|_{L^{q_0}} \leq C_0 \|f\|_{L^{p_0}} \text{ for all } f \in L^{p_0}(X,\mu)$$

*and*

$$\|Tf\|_{L^{q_1}} \leq C_1 \|f\|_{L^{p_1}} \text{ for all } f \in L^{p_1}(X,\mu)$$

*Then for all $f \in L^p(X,\mu)$ we have*

$$\|Tf\|_{L^q} \leq C_0^{1-\theta} C_1^{\theta} \|f\|_{L^p}$$

For the proof consult [5]. Both interpolation theorems generalize to the multilinear setting. A generalization of the Riesz-Thorin interpolation[3] can be found in [1]. For the multilinear interpolation between (restricted) weak type estimates, consult [9].

### 2.3 Dyadic maximal function

For $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ we define the *dyadic maximal function* of $f$ as

$$Mf(x) := \sup_{x \in Q} \frac{1}{|Q|} \int_Q |f|$$

where the supremum is taken over all dyadic cubes containing $x$.

**Proposition 2.5.** *For $1 < p \leq \infty$ we have*

$$\|Mf\|_{L^p(\mathbb{R}^n)} \lesssim_p \|f\|_{L^p(\mathbb{R}^n)}$$

*Proof.* Note that $\|Mf\|_{L^\infty(\mathbb{R}^n)} \leq \|f\|_{L^\infty(\mathbb{R}^n)}$. To prove the claim we show the weak $(1,1)$ estimate

$$|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \leq \lambda^{-1} \|f\|_{L^1(\mathbb{R}^n)}$$

Denote $E_\lambda = |\{x \in \mathbb{R}^n : Mf(x) > \lambda\}|$ and let $\mathcal{Q}$ be the union of maximal dyadic cubes $Q$ such that

$$\frac{1}{|Q|} \int_Q |f| > \lambda$$

These exist, since $|Q|^{-1} \int_Q |f| \leq |Q|^{-1} \|f\|_{L^1(\mathbb{R}^n)}$. The cubes in $\mathcal{Q}$ are by maximality disjoint and $E_\lambda = \cup_{Q \in \mathcal{Q}} Q$. Thus

$$|E_\lambda| = \sum_{Q \in \mathcal{Q}} |Q| \leq \lambda^{-1} \|f\|_{L^1(\mathbb{R}^n)}$$

which is the desired weak $(1,1)$ estimate. The statement of the proposition follows by the Marcinkiewicz interpolation theorem. □

[3] This method is also known as the *complex interpolation*.

**Corollary 2.6.** *The "d-th power" dyadic maximal function defined by*

$$M_d f(x) := \sup_{x \in Q \in \mathcal{D}} \left( \frac{1}{|Q|} \int_Q |f|^d \right)^{1/d}$$

*is bounded on $L^p(\mathbb{R}^n)$ for $d < p \leq \infty$.*

*Proof.* Write $M_d f = (M|f|^d)^{1/d}$ and apply the previous proposition. □

## 2.4 Haar basis

In this section we discuss an orthonormal basis of $L^2(\mathbb{R}^n)$, the Haar basis.

The *Haar function* associated to a dyadic interval $I$ is defined as

$$h_I := |I|^{-1/2} \left( \mathbf{1}_{I_l} - \mathbf{1}_{I_r} \right)$$

We also define the *normalized characteristic function* of $I$ as

$$\chi_I := |I|^{-1/2} \mathbf{1}_I$$

Note that both $h_I$ and $\chi_I$ are normalized in $L^2$ and that $h_I$ has mean zero. For the rest of this section we denote by $\langle \cdot, \cdot \rangle$ the standard inner product on $L^2$.

**Lemma 2.7.** *For any two dyadic intervals $I$ and $J$ we have*

$$\langle h_I, h_J \rangle = \begin{cases} 1 & \text{if } I = J \\ 0 & \text{if } I \neq J \end{cases}$$

*Proof.* By construction $\int_{\mathbb{R}} h_I h_I = 1$. If $I \neq J$ we have the following two options. If $I \cap J = \emptyset$ we immediately obtain $\int_{\mathbb{R}} h_I h_J = 0$. If not, assume that $I \subset J$ and $I \neq J$. By symmetry we may assume that $I \subseteq J_l$ and thus

$$\int_{\mathbb{R}} h_I h_J = |I|^{-1/2} |J|^{-1/2} \int_{J_l} \mathbf{1}_{I_l} - \mathbf{1}_{I_r} = 0$$

□

Denote $\widetilde{V}_n := V_n \setminus \{(0, \ldots, 0)\}$. To each dyadic cube $Q = I_1 \times \cdots \times I_n$ and $j = (j_1, \ldots, j_n) \in \widetilde{V}_n$ we associate an $n$-dimensional Haar function given by

$$h_Q^j := \bigotimes_{k=1}^{n} h_{I_k}^{j_k}$$

where

$$h_{I_k}^0 := \chi_{I_k}$$
$$h_{I_k}^1 := h_{I_k}$$

In view of Lemma 2.7 and the fact that $\langle h_I, \chi_I \rangle = 0$ we see that the family of these $n$-dimensional Haar functions forms an orthonormal system in $L^2(\mathbb{R}^n)$ called the *Haar system.*

Now we investigate its completeness. For $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ define the expectation $\mathbb{E}_k = \mathbb{E}_k^n$ of $f$ as

$$\mathbb{E}_k f := \sum_{Q \in \mathcal{D}_k} \frac{1}{|Q|} \int_Q f(t)dt \, \mathbf{1}_Q$$

for all $k \in \mathbb{Z}$. Note that it can also be written as

$$\mathbb{E}_k f = \sum_{Q \in \mathcal{D}_k} \langle f, \chi_Q \rangle \chi_Q$$

Define also the difference operator $\Delta_k = \Delta_k^n$ by

$$\Delta_k f := \mathbb{E}_{k-1} f - \mathbb{E}_k f$$

The following lemma gives us an alternative description of $\Delta_k$.

**Lemma 2.8.** *(1) For every $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and for all $k \in \mathbb{Z}$ we have the identity*

$$\Delta_k f = \sum_{Q \in \mathcal{D}_k} \sum_{j \in \widetilde{V}_n} \langle f, h_Q^j \rangle h_Q^j$$

*(2) Every $f$ in $L^2(\mathbb{R}^n)$ can be written as*

$$f = \sum_{k \in \mathbb{Z}} \Delta_k f$$

*where the series converges almost everywhere and in $L^2$.*

*Proof.* (1) We induct on $n$. If $n = 1$ and $Q = I$ a dyadic interval, we have

$$\begin{aligned}
\Delta_k f &= \sum_{I \in \mathcal{I}_{k-1}} \frac{1}{|I|} \Big( \int_I f \Big) \mathbf{1}_I - \sum_{I \in \mathcal{I}_k} \frac{1}{|I|} \Big( \int_I f \Big) \mathbf{1}_I \\
&= \sum_{I \in \mathcal{I}_k} \frac{2}{|I|} \Big( \int_{I_l} f \Big) \mathbf{1}_{I_l} + \frac{2}{|I|} \Big( \int_{I_r} f \Big) \mathbf{1}_{I_r} - \sum_{I \in \mathcal{I}_k} \frac{1}{|I|} \Big( \int_{I_l} f + \int_{I_r} f \Big) (\mathbf{1}_{I_l} + \mathbf{1}_{I_r}) \\
&= \sum_{I \in \mathcal{I}_k} \frac{1}{|I|} \Big( \int_{I_l} f \Big) (\mathbf{1}_{I_l} - \mathbf{1}_{I_r}) + \frac{1}{|I|} \Big( \int_{I_r} f \Big) (\mathbf{1}_{I_r} - \mathbf{1}_{I_l}) \\
&= \sum_{I \in \mathcal{I}_k} \langle f, h_I \rangle h_I
\end{aligned}$$

For the induction step we write $Q = Q_* \times I_n$ and $x = (x_*, x_n) \in \mathbb{R}^{n-1} \times \mathbb{R}$. When we apply the average or the difference operator in $n-1$ variables of a $n$ dimensional function $f$ we denote these variables with a superscript, for instance

$$(\mathbb{E}_k^{(*)} f)(x_*, x_n) := (\mathbb{E}_k^{n-1} f_{x_n})(x_*)$$

where $f_{x_n}(x_*) = f(x_*, x_n)$. Similarly we mark the variables in which we take the scalar product with a subscript, e.g. $\langle f, \cdot \rangle_{(*)} := \langle f_{x_n}, \cdot \rangle$. Here, the brackets denote the scalar product on $L^2(\mathbb{R}^{n-1})$.

By Fubini and since $|Q| = |Q_*||I_n|$ and $\mathbf{1}_Q = \mathbf{1}_{Q_*} \otimes \mathbf{1}_{I_n}$, we can write

$$(\Delta_k f)(x) = \sum_{I_n \in \mathcal{I}_{k-1}} \frac{1}{|I_n|} \left( \int_{I_n} (\mathbb{E}_{k-1}^{(*)} f)(x_*, \cdot) \right) \mathbf{1}_{I_n}(x_n) - \sum_{I_n \in \mathcal{I}_k} \frac{1}{|I_n|} \left( \int_{I_n} (\mathbb{E}_k^{(*)} f)(x_*, \cdot) \right) \mathbf{1}_{I_n}(x_n)$$

A similar calculation as for $n = 1$ shows that this is equal to

$$\sum_{I_n \in \mathcal{I}_k} \langle (\mathbb{E}_{k-1}^{(*)} f - \mathbb{E}_k^{(*)} f)(x_*, \cdot), \chi_{I_n} \rangle \chi_{I_n}(x_n) + \langle (\mathbb{E}_{k-1}^{(*)} f)(x_*, \cdot), h_{I_n} \rangle h_{I_n}(x_n)$$

Now we write $\mathbb{E}_{k-1}^{(*)} = \Delta_k^{(*)} + \mathbb{E}_k^{(*)}$. Applying the induction hypothesis we have

$$\begin{aligned}
&\sum_{I_n \in \mathcal{I}_k} \langle (\Delta_k^{(*)} f)(x_*, \cdot), \chi_{I_n} \rangle \chi_{I_n}(x_n) + \langle (\Delta_k^{(*)} f + \mathbb{E}_k^{(*)} f)(x_*, \cdot), h_{I_n} \rangle h_{I_n}(x_n) \\
= &\sum_{I_n \in \mathcal{I}_k} \Big( \langle \sum_{Q_* \in \mathcal{D}_k(\mathbb{R}^{n-1})} \sum_{j \in \widetilde{V}_{n-1}} \langle f, h_{Q_*}^j \rangle_{(*)} h_{Q_*}^j(x_*), \chi_{I_n} \rangle_{(n)} \chi_{I_n}(x_n) \\
&+ \langle \sum_{Q_* \in \mathcal{D}_k(\mathbb{R}^{n-1})} \sum_{j \in \widetilde{V}_{n-1}} \langle f, h_{Q_*}^j \rangle_{(*)} h_{Q_*}^j(x_*), h_{I_n} \rangle_{(n)} h_{I_n}(x_n) \\
&+ \langle \sum_{Q_* \in \mathcal{D}_k(\mathbb{R}^{n-1})} \langle f, h_{Q_*}^0 \rangle_{(*)} h_{Q_*}^0(x_*), h_{I_n} \rangle_{(n)} h_{I_n}(x_n) \Big)
\end{aligned}$$

where $h_{Q_*}^0 := \otimes_{k=1}^{n-1} h_I^0$. But this is the same as

$$\sum_{Q \in \mathcal{D}_k} \sum_{j \in \widetilde{V}_n} \langle f, h_Q^j \rangle h_Q^j(x)$$

(2) Let $x \in \mathbb{R}^n$ and let $Q_{x,k}$ be the unique cube in $\mathcal{D}_k$ containing $x$. Then

$$(\mathbb{E}_k f)(x) = \sum_{Q \in \mathcal{D}_k} \frac{1}{|Q|} \int_Q f \, \mathbf{1}_Q(x) = \frac{1}{|Q_{x,k}|} \int_{Q_{x,k}} f$$

By the Lebesgue differentiation theorem we have that for almost all $x \in \mathbb{R}^n$, $\mathbb{E}_k f \to f$ as $k \to -\infty$. Since $|\mathbb{E}_k f| \leq Mf$ and $Mf \in L^2(\mathbb{R}^n)$, by the dominated convergence theorem $\mathbb{E}_k f \to f$ in $L^2$ as $k \to -\infty$. Now we consider $k \to \infty$. With one application of the Cauchy-Schwarz inequality we have

$$|(\mathbb{E}_k f)(x)| = \left| \frac{1}{|Q_{x,k}|} \int_{Q_{x,k}} f \right| \leq \left( \frac{1}{|Q_{x,k}|} \int_{Q_{x,k}} |f|^2 \right)^{1/2} \leq 2^{-\frac{kn}{2}} \|f\|_{L^2(\mathbb{R}^n)}$$

which tends to 0 as $k \to \infty$. Convergence in $L^2$ follows again from $|\mathbb{E}_k f| \leq Mf$ and the dominated convergence theorem. To prove the claim it remains to observe that

$$\sum_{k=N}^{M} \Delta_k f = \mathbb{E}_{N-1} f - \mathbb{E}_M f \ \to f$$

a.e. and in $L^2$ as $N \to -\infty$, $M \to \infty$.

□

**Corollary 2.9.** *For every function $f$ in $L^2(\mathbb{R}^n)$ we have the representation*

$$f = \sum_{Q \in \mathcal{D}} \sum_{j \in V_n} \langle f, h_Q^j \rangle h_Q^j$$

*Thus, the Haar system is complete.*

We call the family of functions $(h_Q^j)$ the *Haar basis* of $L^2(\mathbb{R}^n)$.

# 3 Proof of Theorem 1.1

This is the central section of this thesis. We discuss the structure of $\Lambda$, determine the numbers $d_j$ and prove the main theorem.

First let us say a few words on the notation. Let $A \subseteq N := \{1, \ldots, n\}$. For $|A|-$tuples $(x_a^0)_{a \in A}, (x_a^1)_{a \in A} \in \mathbb{R}^{|A|}$ and $j_A := (j_a)_{a \in A} \in \{0,1\}^{|A|}$ we denote $x_A^{(j_A)} := (x_a^{j_a})_{a \in A}$. For brevity we denote $(0) := (0, \ldots, 0)$, $(1) := (1, \ldots, 1)$ and also $x_A^{(0)} := x_A^{((0))}$, $x_A^{(1)} := x_A^{((1))}$. Moreover, if $A = N$ we omit the writing of this set in the subscript, for example, $x^{(0)} := x_N^{(0)} = (x_1^0, \ldots, x_n^0)$ and $x^{(j)} := x_N^{(j_N)} = (x_1^{j_1}, \ldots, x_n^{j_n})$. For an $|A|-$tuple of dyadic intervals $(I_a)_{a \in A}$ we write $I_A := \prod_{a \in A} I_a$ for their Cartesian product. If $A = N$ we simply denote $Q = I_N$.

We will also use the following convention. Whenever in concrete applications $j \in V_n$ appears as an index, we replace $j$ by $\sum_{a=1}^n 2^{a-1} j_a$. In other words, then we do not distinguish between a number and its binary decomposition. For instance, we shortly write $F_3$ instead of $F_{(1,1,0)}$.

Part of the notation which we will use throughout the text is taken from [7] and [8]. In particular, for a function $F$ on a dyadic interval $I$ we denote

$$[F]_I = [F(x)]_{x \in I} := \frac{1}{|I|} \int_I F(x) dx$$
$$\langle F \rangle_I = \langle F(x) \rangle_{x \in I} := \frac{1}{|I|} \left( \int_{I_l} F(x) dx - \int_{I_r} F(x) dx \right)$$

Since we will be dealing with functions on dyadic cubes $Q$ of more than one variable we emphasize the variable in which the averages are taken, for example $[F(x_1, \ldots, x_n)]_{x_1 \in I_1}$

or $\langle F(x_1,\dots,x_n)\rangle_{x_1\in I_1}$. We omit writing the interval if it is understood. We also use shortcuts such as

$$[F(x_1,\dots,x_n)]_{x_1,\dots,x_n} := [\dots[[F(x_1,\dots,x_n)]_{x_1}]_{x_2}\dots]_{x_n}$$

and

$$[F]_Q := [F(x_1,\dots,x_n)]_{x_1\dots,x_n}$$

In all of the following the functions $F_j$ are assumed to be real valued dyadic step functions. That is, finite linear combinations of characteristic functions of dyadic cuboids. We also assume that all functions are non-negative. This is enabled by splitting the functions into real and imaginary, positive and negative parts, using multilinearity and density arguments.

With the notation defined above, $\Lambda$ can be written as

$$\Lambda((F_j)_{j\in V}) = \sum_{Q\in\mathcal{D}} |Q|\,\Phi_Q((F_j)_{j\in V})$$

where

$$\Phi_Q((F_j)_{j\in V}) := \Big[\Big\langle \prod_{j\in V} F_j(x^{(j)})\Big\rangle_{x_n^0,\,x_n^1\in I_n}\Big]_{x_k^0,\,x_k^1\in I_k \text{ for } k=1,\dots,n-1}$$

The proof of Theorem 1.1 will consist of two parts. First we define for any finite convex tree $\mathcal{T}$ the (absolute) local version of $\Lambda$ which sums only over the cubes in $\mathcal{T}$,

$$\Lambda_{\mathcal{T}}((F_j)_{j\in V}) := \sum_{Q\in\mathcal{T}} |Q|\,|\Phi_Q((F_j)_{j\in V})|$$

In the first part of the proof we will prove an estimate for $\Lambda_{\mathcal{T}}$. In the second step we will "integrate" it into a global estimate by decomposing the collection of dyadic cubes into trees using a stopping time argument.

Let us also mention that $\Lambda$ is invariant under simultaneous dyadic scaling,

$$\Lambda((\mathrm{D}_{2^k}F_j)_{j\in V}) = 2^{nk}\Lambda((F_j)_{j\in V}) \tag{3.1}$$

where $\mathrm{D}_{2^k}F(x^{(j)}) := F(2^{-k}x^{(j)})$ and $k\in\mathbb{Z}$. A scaling argument implies the necessity of $\sum_{j\in V}\frac{1}{p_j}=1$. Indeed, assuming the conclusion of Theorem 1.1 and applying it to $\Lambda((\mathrm{D}_{2^k}F_j)_{j\in V})$, using (3.1) and $\|\mathrm{D}_{2^k}F_j\|_{L^{p_j}(\mathbb{R}^n)} = 2^{nk/p}\|F_j\|_{L^{p_j}(\mathbb{R}^n)}$ yields

$$|\Lambda((F_j)_{j\in V})| \lesssim_{n,(p_j)} 2^{nk(\sum_{j\in V}\frac{1}{p_j}-1)} \prod_{j\in V}\|F_j\|_{L^{p_j}(\mathbb{R}^n)}$$

Since this should hold for any $k\in\mathbb{Z}$, $\sum_{j\in V}\frac{1}{p_j}\neq 1$ would imply $\Lambda((F_j)_{j\in V})=0$.

Now we give a few comments on the structure of $\Lambda$. The vertex set of the unit cube in $\mathbb{R}^n$ can be represented by the set $V_n=\{0,1\}^n$. We say that two vertices are connected by an edge whenever the corresponding points differ in exactly one coordinate. A face of

the cube is determined by the data $(k, b)$ where $0 \leq k \leq n-1,\ b \in \{0,1\}$. The vertices of the corresponding face are given by

$$\{0,1\}^k \times \{b\} \times \{0,1\}^{n-k-1}$$

The unit cube in $\mathbb{R}^n$ has $2^n$ vertices and $2n$ faces.

The structure of $\Lambda$ can now be described as follows. The subset $V$ of the vertex set determines which functions appear in the multilinear form. To each face we assign a variable $x_i^{j_i}$ and to each vertex a function of $n$ variables, corresponding to the $n$ adjacent faces. The variables which correspond to the opposite faces $x_i^0$ and $x_i^1$ take values in the same dyadic interval $I_i$. One pair of these opposite faces is assigned a Haar function according to the corresponding interval, while all other faces are assigned appropriate characteristic functions. By symmetry we choose the Haar function to be assigned to the faces belonging to the last two variables $x_n^0$, $x_n^1$.

**Example 3.1** (Classical paraproduct)**.** If $n = 2$ and $V = \{(0,0),(1,1)\}$, the form $\Lambda$ corresponds to the the classical two-dimensional dyadic paraproduct with $|S| = 2$. In the multilinear form are involved $F_0(x^{(0,0)})$ and $F_3(x^{(1,1)})$.

**Example 3.2** (Twisted paraproduct)**.** If $n = 2$ and $V = \{(0,0),(1,0),(0,1)\}$ we are in the case of the twisted paraproduct. Involved in $\Lambda_{\mathrm{d}}$ are

$$F_0(x^{(0,0)}),\ F_1(x^{(1,0)}),\ F_2(x^{(0,1)})$$

These two examples are schematically represented in Figure 1. For more examples consult Section 4 where we state some three- and four-dimensional instances.

Figure 1: A two-dimensional unit cube representation of the classical (left) and twisted (right) paraproduct. The Haar function is assigned to the emphasized edges.

### 3.1 Determining $d_j$

In this section we determine the range of exponents from Theorem 1.1.

For $j \in V_n$ and $A \subseteq N$ recall the notation $j_A = (j_a)_{a \in A}$. In words, $j_A$ is the projection of $j$ onto the components indexed by $A$. For $U \subseteq V_n$ we denote by $\pi_A(U)$ the set of images under this projection, $\pi_A(U) := \{j_A : j \in U\}$. For $\mathcal{A} \subseteq \mathcal{P}(N)$ let us denote $\cup\mathcal{A} := \cup_{A \in \mathcal{A}} A$.

**Definition 3.1.** We call a pair $(\mathcal{A}, U)$ *eligible*, if

- $\mathcal{A}$ is a subset of a partition of $N$, i.e. $\mathcal{A} \subseteq \mathcal{P}(N)$ and $A \cap B = \emptyset$ if $A \neq B$ and $A, B \in \mathcal{A}$
- $U \subseteq V_n$ and if $\cup\mathcal{A} \neq \emptyset$, then $|\pi_{\cup\mathcal{A}}(U)| = 1$

We call a triple $(\mathcal{A}, U, i)$ *eligible*, if $(\mathcal{A}, U)$ is eligible and $i \in N \setminus \cup\mathcal{A}$.

**Example 3.3.** For any $U \subseteq V_n$, $(\emptyset, U)$ is eligible.

**Example 3.4.** Let $U$ consist of vertices

$$\begin{gathered}(0,0,0,1,1,0,1)\\(1,0,0,1,1,0,1)\\(0,1,0,1,1,0,1)\\(1,1,0,1,1,0,1)\end{gathered}$$

and let $\mathcal{A} = \{\{3,4\},\{5,6\},\{7\}\}$. Then $(\mathcal{A}, U)$ is eligible.

For $j_1, j_2 \in \{0,1\}$ let us denote by $j_1 \oplus j_2$ their sum in $\mathbb{Z}/2\mathbb{Z}$. For two vertices $j, k$ we denote $j \oplus k := (j_i \oplus k_i)_{i=1}^n$ and for a vertex $j$ we call $j \oplus (1)$ the vertex *antipodal* to $j$.

Let $(\mathcal{A}, U, i)$ be eligible. For each $j \in U$ we define

$$\lambda_j(\mathcal{A}, U, i) := (\mathcal{A}', A, U_0, U_1)$$

where $\mathcal{A}', A, U_0, U_1$ are defined as follows. Let $A \subseteq N \setminus \cup\mathcal{A}$ be the set of maximal cardinality such that $i \in A$ and such that we can split $U = U_0 \cup U_1$, where for $b = 0, 1$, $U_b = U_{b,j}$ is defined as

$$U_b := \{u \in U : u_a = j_a \oplus b \text{ for all } a \in A\}$$

In words, we split the set $U$ into two parts, $U_0$ and $U_1$. In $U_0$ we take the vertices which agree with $j$ on the components indexed by $A$, and in $U_1$ those, which agree on these components with its antipodal point. A set $A$, such that the above decomposition is possible, certainly exists. Since $A = \{i\}$ already satisfies the above properties except possibly maximality, we simply add a maximal number of vertices such that the required conditions are still fulfilled. In the best case we can take $A = \{1, \ldots, n\}$. Also, the set $A$ is unique. If we had two maximal sets $A \neq B$ satisfying the above conditions, then $A \cup B$ contains both of them and satisfies the same conditions as well, contradicting the maximality of $A$ and $B$.

Now we define $\mathcal{A}' := \mathcal{A} \cup \{A\}$. Note that the pairs $(\mathcal{A}', U_b)$ are eligible and that $j \in U_0$. Let us also remark that in the above splitting we allow one of the sets $U_b$ to be empty.

**Definition 3.2.** Let $(\mathcal{A}, U, i)$ be eligible, $j \in U$ and $(\mathcal{A}', A, U_0, U_1) = \lambda_j(\mathcal{A}, U, i)$. If one of the sets $U_0$ or $U_1$, is empty, we call the triple $(\mathcal{A}, U, i)$ *degenerate*. We call a set $U \subseteq V_n$ *degenerate*, if there exist $i \in N$ such that the triple $(\emptyset, U, i)$ is degenerate. If a set or a triple is not degenerate, we call it *non-degenerate*.

Note that for two different vertices $j, k \in U$ either $U_{b,j} = U_{b,k}$ or $U_{b,j} = U_{b\oplus 1,k}$, $b = 0, 1$. Therefore, in Definition 3.2 it does not matter which vertex $j$ we pick. In other words, with a different selection would be the roles of $U_0$ and $U_1$ possibly interchanged.

**Example 3.5.** Let $n = 3$. The vertex set $V_3$ is non-degenerate. The set $U = \{(0, 0, 0), (0, 0, 1)\}$ is degenerate, since $(\emptyset, U, 1)$ and $(\emptyset, U, 2)$ are degenerate.

Now we can define $d_j$ using the following recursive procedure. Let $\sigma \in S_n$ be a permutation. Let $(\mathcal{A}, U)$ be an eligible pair and let $j \in U$. If it exists, pick the smallest $k \in N$ such that $(\mathcal{A}, U, \sigma(k))$ is eligible and, if possible, non-degenerate. If a $k$ such that $(\mathcal{A}, U, \sigma(k))$ is eligible does not exist, the algorithm is finished. Let $(\mathcal{A}', A, U_0, U_1) = \lambda_j(\mathcal{A}, U, \sigma(k))$. We repeat the procedure on $(\mathcal{A}', U_0)$. The algorithm must come to an end since $\cup\mathcal{A}$ increases in size by at least one every step but at the same time must be a subset of $N$. Once $\cup\mathcal{A} = N$, the algorithm will end because there is no more eligible triple $(\mathcal{A}, U, i)$. By the definition of eligibility we then have $U = \{j\}$. The corresponding partition $\mathcal{A}$ of $N$ we then denote by $\mathcal{A}_j$. Let us write $\mu_j((\mathcal{A}, U), \sigma) := \mathcal{A}_j$.

Applying this procedure to $(\emptyset, V)$, for $j \in V$ we define $d_j$ as

$$d_j := \max_{\sigma:\sigma(1)=n} 2^{|\mu_j((\emptyset,V),\sigma)|-1} \tag{3.2}$$

if $V$ is non-degenerate. If $(\emptyset, V, i)$ is degenerate for some $i \neq n$, let $A$ be given by $\lambda_j(\emptyset, V, i)$ and denote $\widetilde{V} := \pi_{N\setminus A}(V)$. Note that $\widetilde{V}$ is non-degenerate. We define

$$d_j := d_{\tilde{j}} \tag{3.3}$$

where $\tilde{j} := j_{N\setminus A} \in \widetilde{V}$. If the triple $(\emptyset, V, n)$ is degenerate, we set $d_j := 0$.

Let us compute $d_j$ for several concrete examples. We will elaborate on the last two examples in Section 4. It will always be assumed that $\sigma = (n, \ldots, 1)$. In the case of these vertex sets, the claim for any other $\sigma$ follows by symmetry.

**Example 3.6.** If $V$ equals $\{(0, \ldots, 0)\}$ or $\{(0, \ldots, 0), (1, \ldots, 1, 0)\}$, then $(\emptyset, V, n)$ is degenerate and $d_j = 0$ for $j \in V$.

**Example 3.7.** For $V = \{(0, \ldots, 0), (1, \ldots, 1)\}$ we have $d_0 = d_{2^{n-1}} = 1$.

Indeed, starting with $(\emptyset, V)$ it is easy to see that for any $j \in V$, there is only one iteration of the above algorithm. We have $\mu_j((\emptyset, V), \sigma) = \{\{1, \ldots, n\}\}$.

For $V \subseteq V_n$, $i \in N$ and $b = 0, 1$ we denote

$$V^{b,i} := \{v \in V : v_i = b\} \tag{3.4}$$

Thus for any $i \in N$ we have the decomposition $V = V^{0,i} \cup V^{1,i}$.

**Example 3.8.** For $V = V_n$ we have $d_j = 2^{n-1}$ for all $j \in V_n$.

Before proceeding let us observe that in this case for any $i \in N$ we have $|V^{b,i}| = 2^{n-1}$ and that $\pi_{N\setminus\{i\}}(V^{b,i}) = V_{n-1}$.

We prove $d_j = 2^{n-1}$ by induction on $n$. We claim that for every $j \in V$,

$$\mu_j((\emptyset, V), \sigma) = \{\{1\}, \ldots, \{n\}\}$$

If $n = 1$, $V_1 = \{0\}$ and the claim holds. If $n > 1$, by symmetry it suffices to show this for $j = (0)$. Consider the eligible triple $(\emptyset, V, n)$ and let $(\mathcal{A}', A, U_0, U_1) = \lambda_j(\emptyset, V, n)$. We claim that $A = \{n\}$. Indeed, since $\pi_{N\setminus\{n\}}(V^{0,n}) = V_{n-1}$, for any $k \in N \setminus \{n\}$, half of the vertices of $\pi_{N\setminus\{n\}}(V^{0,n})$ lie on the face determined by the data $(k, 0)$ and half on the face determined by $(k, 1)$. Hence there can be no other elements in $A$ but $n$. Also, $U_0 = V^{0,n}$ and $\mathcal{A}' = \{A\}$. By induction $\mu_j((\mathcal{A}', U_0), \sigma) = \mathcal{A}' \cup \{\{1\}, \ldots, \{n-1\}\}$. From $\mu_j((\mathcal{A}', U_0), \sigma) = \mu_j((\emptyset, V), \sigma)$ follows the claim.

**Example 3.9.** Let $n \geq 2$ and let $V$ equal to $V_n^0$ or $V_n^1$, where for $\beta = 0, 1$,

$$V_n^\beta := \{v \in V_n : \oplus_{i=1}^n v_i = \beta\}$$

that is, $V_n^0$ (or $V_n^1$) is the sets of all vertices which have an even (or odd) number of components equal to 1. Then, $d_j = 2^{n-2}$ for all $j \in V$.

To see that, first note that since $V$ contains exactly half of the vertices from $V_n$, then $|V| = 2^{n-1}$ and that if $n = 2$,

$$V_2^0 = \{(0,0), (1,1)\}$$
$$V_2^1 = \{(0,1), (1,0)\}$$

which consist of two antipodal points.

Now, by induction we can see that for each $i \in N$ we have $|V^{b,i}| = 2^{n-2}$ and $\pi_{N\setminus\{i\}}(V^{b,i}) = V_{n-1}^{\beta\oplus b}$. Indeed, for $n = 2$ this is clear. For $n > 2$ note that $\pi_{N\setminus\{i\}}(V^{b,i}) \subseteq V_{n-1}^{\beta\oplus b}$ and hence $|V^{b,i}| \leq |V_{n-1}^{\beta\oplus b}|$. By the induction hypothesis, $|V^{b,i}| \leq 2^{n-2}$. Since the union (3.4) is disjoint, $|V^{b,i}| = 2^{n-2}$.

To determine $d_j$ we proceed in the same way as in Example 3.8. We claim that for every $j \in V$,

$$\mu_j((\emptyset, V), \sigma) = \{\{1, 2\}, \{3\}, \ldots, \{n\}\}$$

If $n = 2$ it is clear that $\mu_j((\emptyset, V), \sigma) = \{\{1, 2\}\}$. If $n \geq 2$ assume for simplicity that $V = V_n^0$ and $j = (0)$, the general case being similar. Let $(\mathcal{A}', A, U_0, U_1) = \lambda_j(\emptyset, V, n)$. We have $A = \{n\}$, which can be seen as before using $\pi_{N\setminus\{n\}}(V^{0,n}) = V_{n-1}^0$. Also, $U_0 = V^{0,n}$ and $\mathcal{A}' = \{A\}$. By the induction hypothesis, $\mu_j((\mathcal{A}', U_0), \sigma) = \mathcal{A}' \cup \{\{1, 2\}, \{3\}, \ldots, \{n-1\}\}$. The assertion follows.

## 3.2 A single tree estimate

The estimate for a single tree which we prove in this section is

**Proposition 3.1.** *Let $V \subseteq V_n$ be non-degenerate. For any finite convex tree $\mathcal{T}$ with root $Q_{\mathcal{T}}$ and leaves $\mathcal{L}(\mathcal{T})$ we have*

$$\Lambda_{\mathcal{T}}((F_j)_{j\in V}) \lesssim_n |Q_{\mathcal{T}}| \prod_{j\in V} \max_{Q\in\mathcal{L}(\mathcal{T})} \big[F_j^{d_j}\big]_Q^{d_j^{-1}} \tag{3.5}$$

Define the maps $\varphi : \mathcal{P}(N) \to V_n$ as $\varphi(A) := (\mathbf{1}_A(i))_{i\in N}$ and $s : \mathcal{P}(V_n) \to V_n$ as $s(U) := \bigoplus_{u\in U} u$. To every eligible pair $(\mathcal{A}, U)$ we associate an averaging term $[(F_j)_{j\in U}]_{\mathcal{A},U} = [(F_j)_{j\in U}]_{\mathcal{A},U,Q}$ given by

$$[(F_j)_{j\in U}]_{\mathcal{A},U} := \Big[ \prod_{j\in U} \prod_{\widetilde{A}\subseteq\varphi(\mathcal{A})} F_j(x^{(j\oplus s(\widetilde{A}))}) \Big]_{x^{(0)},x^{(1)}\in Q} \tag{3.6}$$

where $\varphi(\mathcal{A}) := \{\varphi(A) : A \in \mathcal{A}\}$. Note that this term involves $2^{|\mathcal{A}|}$ copies of each function $F_j$. Now let us turn our attention to some familiar instances.

**Example 3.10.** If $U = V_n$ and $\mathcal{A} = \emptyset$, denote $[(F_j)_{j\in V_n}]_{\square^n(Q)} := [(F_j)_{j\in V_n}]_{\emptyset,V_n}$. This expression

$$[(F_j)_{j\in V_n}]_{\square^n(Q)} = \Big[ \prod_{j\in V_n} F_j(x^{(j)}) \Big]_{x^{(0)},x^{(1)}\in Q}$$

is called the *Gowers box inner product* of $2^n$ functions. Its two-dimensional version plays the same role in the proof of the boundedness of the twisted paraproduct [7] as (3.6) will play later on.

**Example 3.11.** If $U = \{(0)\}$ and $\mathcal{A} = \{\{1\}, \dots, \{n\}\}$, denote $\|F\|_{\square^n(Q)} := [F]_{\mathcal{A},U}^{2^{-n}}$. This expression

$$\|F\|_{\square^n(Q)} = \Big[ \prod_{j\in V_n} F(x^{(j)}) \Big]_{x^{(0)},x^{(1)}\in Q}^{2^{-n}}$$

is called the *Gowers box norm* of $F$. Its two-dimensional version was used in [7]. This paper contains more reference to its first appearance in additive combinatorics as well. Gowers box norms were also more closely studied in [6].

The following two lemmas generalize the box Cauchy-Schwarz inequality and the estimate of the Gowers box norm by the "2nd power" average from [7]. They are crucial for the restricted range of exponents in Theorem 1.1.

**Lemma 3.2.** *Assume that $U \subseteq V_n$ consists of $d$ pairs of antipodal points. For a function $F$ on a dyadic cube $Q$ we have*

$$\Big[ \prod_{j\in U} F(x^{(j)}) \Big]_{x^{(0)},x^{(1)}\in Q} \le \Big[ |F|^d \Big]_Q^2$$

*Proof.* The estimate we have to prove is a special case of the strong $(\frac{1}{d})_{j\in U}$ estimate for the multilinear form

$$(F_j)_{j\in U} \mapsto |Q|^2 \Big[ \prod_{j\in U} F_j(x^{(j)}) \Big]_{x^{(0)},x^{(1)}\in Q}$$

when for every $j \in U$, $F_j = F$. Since $(\frac{1}{d})_{j\in U}$ is in the convex hull of points in

$$\big\{(\delta_{k,j} + \delta_{k',j})_{j\in U} : k \in U\big\}$$

where

$$\delta_{k,j} = \begin{cases} 1 & : k = j \\ 0 & : k \neq j \end{cases}$$

and $k' := k \oplus (1)$, it suffices to show the estimate for the latter points and then interpolate using the multilinear complex interpolation. But for these points it is clear, since for $(\delta_{k,j} + \delta_{k',j})_{j\in U}$ we can estimate

$$\begin{aligned} \Big[\prod_{j\in U} F_j(x^{(j)})\Big]_{x^{(0)},x^{(1)}} &\leq \prod_{j\notin\{k,k'\}} \|F_j\|_{L^\infty(Q)} \left[F_k(x^{(k)})F_{k'}(x^{(k)\prime})\right]_{x^{(k)},\,x^{(k)\prime}} \\ &= \prod_{j\notin\{k,k'\}} \|F_j\|_{L^\infty(Q)} \, |Q|^{-2} \, \|F_k\|_{L^1(Q)} \, \|F_{k'}\|_{L^1(Q)} \end{aligned}$$

□

An alternative way to establish this estimate would be by inductive applications of the Cauchy-Schwarz inequality. Some examples of sets in $\mathbb{R}^3$ to which Lemma 3.2 applies are $V_3$, $\{(0,0,0),(1,1,1)\}$ and $\{(0,0,0),(1,1,0),(0,0,1),(1,1,1)\}$.

**Lemma 3.3.** *Let $(\mathcal{A}, U)$ be eligible and $\sigma \in S_n$. For each $j \in U$ let $\mathcal{A}_j = \mu_j((\mathcal{A}, U), \sigma)$. Then*

$$\big[(F_j)_{j\in U}\big]_{\mathcal{A},U,Q} \leq \prod_{j\in U} \Big[|F_j|^{2^{|\mathcal{A}_j|-1}}\Big]_Q^{2^{-|\mathcal{A}_j\setminus\mathcal{A}|+1}}$$

*Proof.* We induct on $k := n - |\cup\mathcal{A}|$. If $k = 0$, then $\cup\mathcal{A} = N$. By the definition of eligibility we have $|U| = 1$, i.e. $U = \{j\}$ and $\mathcal{A}_j = \mathcal{A}$. Note that

$$\{j \oplus s(A) : A \subseteq \varphi(\mathcal{A})\}$$

consists of $2^{|\mathcal{A}|-1}$ pairs of antipodal points. The claim then follows by Lemma 3.2. For the induction step, pick a vertex $l \in U$. Let $m \in N$ be the smallest such that $(\mathcal{A}, U, \sigma(m))$ is eligible and, if possible, non-degenerate. Let $(\mathcal{A}', A, U_0, U_1) = \lambda_l(\mathcal{A}, U, \sigma(m))$ and note that $|\cup\mathcal{A}'| > |\cup\mathcal{A}|$. Using Fubini rewrite

$$[(F_j)_{j\in U}]_{\mathcal{A},U} = \Big[\prod_{b\in\{0,1\}} \Big[\prod_{j\in U_b} \prod_{\widetilde{A}\subseteq\varphi(\mathcal{A})} F_j(x^{(j\oplus s(\widetilde{A}))})\Big]_{x_A^{(j_A\oplus(b))}}\Big]_{x^{(0)}_{N\setminus A},\,x^{(1)}_{N\setminus A}}$$

Since for two different $j, l \in U$ either $U_{b,j} = U_{b,l}$ or $U_{b,j} = U_{b\oplus 1,l}$, $b = 0, 1$, it does not matter which vertex $l$ we pick for the above decomposition.

Now apply the Cauchy-Schwarz inequality in the variables $x^{(0)}_{N\setminus A}, x^{(1)}_{N\setminus A} \in I_{N\setminus A}$ to bound this term by

$$[(F_j)_{j\in U_0}]^{1/2}_{\mathcal{A}',U_0} \, [(F_j)_{j\in U_1}]^{1/2}_{\mathcal{A}',U_1}$$

Applying the induction hypothesis to each of these two factors gives the claim. □

To every eligible pair $(\mathcal{A}, U)$ we associate $n$ local forms $\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}} = \Theta^{(i)}_{\mathcal{A},U,\mathcal{T}}((F_j)_{j\in U})$, $i = 1, \dots, n$, given by

$$\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}} := \sum_{Q\in\mathcal{T}} \sum_{\alpha_{i+1},\dots,\alpha_n\in\{l,r\}} |Q|\, \Phi^{(i)}_{\mathcal{A},U}$$

with $\Phi^{(i)}_{\mathcal{A},U} = \Phi^{(i)}_{\mathcal{A},U,Q,\alpha_{i+1},\dots,\alpha_n}((F_j)_{j\in U})$ defined as

$$\Phi^{(i)}_{\mathcal{A},U} := \Big[\Big\langle \prod_{j\in U} \prod_{\widetilde{A}\subseteq\varphi(\mathcal{A})} F_j(x^{(j\oplus s(\widetilde{A}))})\Big\rangle_{x_i^0,x_i^1\in I_i}\Big]_{\substack{x_k^0,\, x_k^1\in I_k & \text{for } k=1,\dots,i-1\\ x_k^0,\, x_k^1\in I_{k,\,\alpha_k} & \text{for } k=i+1,\dots,n}}$$

We also define their absolute versions $\widetilde{\Theta}^{(i)}_{\mathcal{A},U,\mathcal{T}} := \sum_{Q\in\mathcal{T}} \sum_{\alpha_{i+1},\dots,\alpha_n\in\{l,r\}} |Q|\, |\Phi^{(i)}_{\mathcal{A},U}|$. Note that $|\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}}| \le \widetilde{\Theta}^{(i)}_{\mathcal{A},U,\mathcal{T}}$ and that $\widetilde{\Theta}^{(n)}_{\emptyset,V,\mathcal{T}} = \Lambda_{V,\mathcal{T}}$. Also, if $i \neq n$, these forms sum over certain dyadic cuboids. If the triple $(\mathcal{A}, U, i)$ is degenerate, none of the functions depends on the variables $x_i^0$ or $x_i^1$. By the mean zero property of the Haar functions then $\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}} \equiv 0$. Observe also that if $i \in \cup\mathcal{A}$ and $(\mathcal{A}, U)$ is eligible, then $\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}} \ge 0$.

Finally, for any collection of dyadic cubes $\mathcal{C}$ define $\Xi_{\mathcal{A},U,\mathcal{C}} = \Xi_{\mathcal{A},U,\mathcal{C}}((F_j)_{j\in U})$ by

$$\Xi_{\mathcal{A},U,\mathcal{C}} := \sum_{Q\in\mathcal{C}} \big[(F_j)_{j\in U}\big]_{\mathcal{A},U,Q}$$

The following lemma generalizes the telescoping identity from [7] to $\mathbb{R}^n$ and is one of the main ingredients in our proof.

**Lemma 3.4** (Telescoping identity)**.** *For any finite convex tree $\mathcal{T}$ with root $Q_{\mathcal{T}}$ and leaves $\mathcal{L}(\mathcal{T})$ we have*

$$\sum_{i=1}^{n} \Theta^{(i)}_{\mathcal{A},U,\mathcal{T}} = \Xi_{\mathcal{A},U,\mathcal{L}(\mathcal{T})} - \Xi_{\mathcal{A},U,\{Q_{\mathcal{T}}\}}$$

*Proof.* Convexity of $\mathcal{T}$ means that each cube in $\mathcal{T} \setminus \{Q_{\mathcal{T}}\}$ has all $2^n$ children and the parent in $\mathcal{T} \cup \mathcal{L}(\mathcal{T})$. Hence the right hand side can be expanded into a telescoping sum

$$\sum_{Q\in\mathcal{T}} \Big( \sum_{\widetilde{Q}\text{ is a child of } Q} \Xi_{\{\widetilde{Q}\}} - \Xi_{\{Q\}} \Big)$$

Thus it is enough to verify the identity when $\mathcal{T}$ consists of only one cube $Q = I_1 \times \cdots \times I_n$. By definition of $h_I$ and $\chi_I$ we have for any dyadic interval $I$

$$\chi_{I_l}(x)\chi_{I_l}(y) + \chi_{I_r}(x)\chi_{I_r}(y) - \chi_I(x)\chi_I(y) = h_I(x)h_I(y) \tag{3.7}$$

We also have the identity

$$\sum_{i=1}^{n}\Big[\Big(\prod_{j=1}^{i-1}\chi_{I_j}(x_j^0)\chi_{I_j}(x_j^1)\Big)\Big(\sum_{\alpha_i\in\{l,r\}}\chi_{I_{i,\alpha_i}}(x_i^0)\chi_{I_{i,\alpha_i}}(x_i^1)-\chi_{I_i}(x_i^0)\chi_{I_i}(x_i^1)\Big)$$
$$\Big(\prod_{j=i+1}^{n}\sum_{\alpha_j\in\{l,r\}}\chi_{I_{j,\alpha_j}}(x_j^0)\chi_{I_{j,\alpha_j}}(x_j^1)\Big)\Big]$$
$$=\Big(\prod_{j=1}^{n}\sum_{\alpha_j\in\{l,r\}}\chi_{I_{j,\alpha_j}}(x_j^0)\chi_{I_{j,\alpha_j}}(x_j^1)\Big)-\Big(\prod_{j=1}^{n}\chi_{I_j}(x_j^0)\chi_{I_j}(x_j^0)\Big)$$

Combining this identity with (3.7) we establish the claim. □

The importance of the telescoping identity is that it reduces the controlling of a multi-scaled quantity $\Theta^{(i)}_{\mathcal{A},U\mathcal{T}}$ to the controlling of the other $n-1$ "$\Theta-$terms" and two single scaled expressions: one of the finest scale $\mathcal{L}(\mathcal{T})$ and one of the roughest $Q_{\mathcal{T}}$. Note that if $F_j\geq 0$ for all $j\in U$, then

$$\sum_{i=1}^{n}\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}}\leq \Xi_{\mathcal{A},U,\mathcal{L}(\mathcal{T})}$$

We will use this fact without further mention.

Another important part of the proof of the main theorem is the following lemma. It will be used to gradually control the local forms $\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}}$.

**Lemma 3.5** (Reduction inequality)**.** *Let $(\mathcal{A},U,i)$ be eligible and $l\in U$. Let $(\mathcal{A}',A,U_0,U_1)=\lambda_l(\mathcal{A},U,i)$. Then*

$$\widetilde{\Theta}^{(i)}_{\mathcal{A},U,\mathcal{T}}\leq\big(\Theta^{(i)}_{\mathcal{A}',U_0,\mathcal{T}}\big)^{1/2}\,\big(\Theta^{(i)}_{\mathcal{A}',U_1,\mathcal{T}}\big)^{1/2}$$

*Proof.* Assume that $i=n$. Rewrite

$$\widetilde{\Theta}^{(n)}_{\mathcal{A},U,\mathcal{T}}=\sum_{Q\in\mathcal{T}}|Q|\,|\Phi^{(n)}_{\mathcal{A},U}|$$

with

$$\Phi^{(n)}_{\mathcal{A},U}=\Big[\prod_{b\in\{0,1\}}\Big[\Big\langle\prod_{j\in U_b}\prod_{\widetilde{A}\subseteq\varphi(\mathcal{A})}F_j(x^{(j\oplus s(\widetilde{A}))})\Big\rangle_{x_n^{jn\oplus b}}\Big]_{x_{A\setminus\{n\}}^{(j_{A\setminus\{n\}}\oplus(b))}\in I_{A\setminus\{n\}}}\Big]_{x^{(0)}_{N\setminus A},x^{(1)}_{N\setminus A}\in I_{N\setminus A}}$$

Now apply the Cauchy-Schwarz inequality twice, first in the variables $x^{(0)}_{N\setminus A},x^{(1)}_{N\setminus A}$ and then over $I_{N\setminus A}\in\mathcal{T}$. If $i\neq n$ we proceed similarly. The only difference is that we need to rewrite the expression with respect to appropriate dyadic cuboids. □

Now we have everything prepared to prove Proposition 3.1. The essence of the proof are Lemma 3.4 and Lemma 3.5. To prove (3.5) we first try to dominate $\widetilde{\Theta}^{(n)}_{\emptyset,V,\mathcal{T}}((F_j)_{j\in V})$

by simpler terms which involve a smaller number of different functions, using the reduction inequalities. If this is not possible, we use the telescoping identity which reduces its bounding to the bounding of the other $n-1$ local forms and try to bound the latter. Any positive terms coming from applications of the telescoping identity can be ignored, while any terms involving $\Xi_{\mathcal{A},U,\mathcal{C}}$ are estimated using Lemma 3.3.

*Proof of Proposition 3.1.* We prove a stronger statement: Let $(\mathcal{A}, U, i)$ be eligible and non-degenerate. For each $j \in U$ we choose a permutation $\sigma_j$ such that $\sigma_j(1) = i$ and $|\mathcal{A}_j|$ is maximal, where $\mathcal{A}_j = \mu_j((\mathcal{A}, U), \sigma_j)$. We claim that

$$\widetilde{\Theta}^{(i)}_{\mathcal{A},U,\mathcal{T}}((F_j)_{j\in U}) \lesssim_n |Q_{\mathcal{T}}| \prod_{j\in U} \max_{Q\in\mathcal{L}(\mathcal{T})} [F_j^{2^{|\mathcal{A}_j|-1}}]^{2^{-|\mathcal{A}_j\setminus\mathcal{A}|+1}} \tag{3.8}$$

The proposition will then follow from the case $\mathcal{A} = \emptyset$, $U = V$ and $i = n$.

We induct on $k := n - |\cup\mathcal{A}|$. Suppose that $k = 0$, i.e. $\cup\mathcal{A} = N$. As there does not exist any eligible triple $(\mathcal{A}, U, i)$, the claim trivially holds.

Now consider $\widetilde{\Theta}^{(i)}_{\mathcal{A},U,\mathcal{T}}$ where $(\mathcal{A}, U, i)$ is eligible and non-degenerate with $n-|\cup\mathcal{A}| = k$ and assume that we have proved (3.8) for all eligible and non-degenerate $(\mathcal{A}', U', m)$ with $n - |\cup\mathcal{A}'| < k$. By the reduction inequality,

$$\widetilde{\Theta}^{(i)}_{\mathcal{A},U,\mathcal{T}} \leq \left(\Theta^{(i)}_{\mathcal{A}',U_0,\mathcal{T}}\right)^{1/2} \left(\Theta^{(i)}_{\mathcal{A}',U_1,\mathcal{T}}\right)^{1/2} \tag{3.9}$$

where $\mathcal{A}'$, $U_0$ and $U_1$ are given by $\lambda_l(\mathcal{A}, U, i)$ for some $l \in U$. By the same reasoning as in Lemma 3.3 it does not matter which vertex $l$ we pick. Now, $(\mathcal{A}', U_0)$ and $(\mathcal{A}', U_1)$ are eligible, $i \in \cup\mathcal{A}'$ and $|\cup\mathcal{A}'| > |\cup\mathcal{A}|$. For $b = 0, 1$ we use the telescoping identity to estimate

$$\Theta^{(i)}_{\mathcal{A}',U_b,\mathcal{T}} \leq \Xi_{\mathcal{A}',U_b,\mathcal{L}(\mathcal{T})} - \sum_{m\neq i} \Theta^{(m)}_{\mathcal{A}',U_b,\mathcal{T}}$$

For all $m \in \cup\mathcal{A}'$ we have $\Theta^{(m)}_{\mathcal{A}',U_b,\mathcal{T}} \geq 0$. Thus,

$$\Theta^{(i)}_{\mathcal{A}',U_b,\mathcal{T}} \leq \Xi_{\mathcal{A}',U_b,\mathcal{L}(\mathcal{T})} + \sum_{m\notin\cup\mathcal{A}'} |\,\Theta^{(m)}_{\mathcal{A}',U_b,\mathcal{T}}\,| \tag{3.10}$$

Let $\sigma \in S_n$ be such that $\sigma(1) = i$. For $j \in U_b$ let $\mathcal{A}'_j := \mu_j((\mathcal{A}', U_b), \sigma)$. Since $|\mathcal{A}'_j| \leq |\mathcal{A}_j|$, by monotonicity of $p \mapsto [F^p]_Q^{1/p}$ we have

$$[\,F_j^{2^{|\mathcal{A}'_j|-1}}\,]_Q^{2^{-|\mathcal{A}'_j\setminus\mathcal{A}'|+1}} \leq [\,F_j^{2^{|\mathcal{A}_j|-1}}\,]_Q^{2^{-|\mathcal{A}_j\setminus\mathcal{A}'|+1}} \tag{3.11}$$

By Lemma 3.3 together with (3.11) and the fact that $\mathcal{L}(\mathcal{T})$ partitions $Q_{\mathcal{T}}$ it follows

$$\begin{aligned}\Xi_{\mathcal{A}',U_b,\mathcal{L}(\mathcal{T})} &\leq |Q_{\mathcal{T}}| \prod_{j\in U_b} \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_j^{2^{|\mathcal{A}_j|-1}}\right]_Q^{2^{-|\mathcal{A}_j\setminus\mathcal{A}'|+1}} \\ &= |Q_{\mathcal{T}}| \prod_{j\in U_b} \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_j^{2^{|\mathcal{A}_j|-1}}\right]_Q^{2^{-|\mathcal{A}_j\setminus\mathcal{A}|+2}}\end{aligned} \tag{3.12}$$

Let $m \notin \cup\mathcal{A}'$. If $(\mathcal{A}', U_b, m)$ is degenerate, then $\Theta^{(m)}_{\mathcal{A}',U_b,\mathcal{T}} = 0$. So from now on assume that this is not the case. For $j \in U_b$ we choose a permutation $\sigma'_j \in S_n$ such that $\sigma'_j(1) = m$ and $|\mathcal{A}''_j|$ is maximal, where $\mathcal{A}''_j := \mu_j((\mathcal{A}', U_b), \sigma'_j)$. We have $|\mathcal{A}''_j| \leq |\mathcal{A}_j|$. We estimate $|\Theta^{(m)}_{\mathcal{A}',U_b,\mathcal{T}}| \leq \widetilde{\Theta}^{(m)}_{\mathcal{A}',U_b,\mathcal{T}}$ and since the triple $(\mathcal{A}', U_b, m)$ is eligible, we can apply the induction hypothesis. Together with an inequality analogous to (3.11) this yields

$$|\,\Theta^{(m)}_{\mathcal{A}',U_b,\mathcal{T}}\,| \lesssim_n |Q_{\mathcal{T}}| \prod_{j\in U_b} \max_{Q\in\mathcal{L}(\mathcal{T})} \big[F_j^{2^{|\mathcal{A}_j|-1}}\big]_Q^{2^{-|\mathcal{A}_j\setminus\mathcal{A}|+2}} \tag{3.13}$$

To finish the proof it remains to plug (3.12) and (3.13) into (3.10) and use (3.9). □

### 3.3 Completing the proof

This section is devoted to the completion of the proof of Theorem 1.1. To establish the global estimate we decompose the collection of dyadic cubes into trees, to which we apply Proposition 3.1. This closely follows the procedure in [8].

We are set to show the bound

$$|\Lambda((F_j)_{j\in V})| \lesssim_{n,(p_j)} \prod_{j\in V} \|F_j\|_{L^{p_j}(\mathbb{R}^n)}$$

whenever $\sum_{j\in V} \frac{1}{p_j} = 1$ and $d_j < p_j < \infty$ for every $j \in V$.

*Proof of Theorem 1.1.* First we consider the case when $V$ is non-degenerate. In the following we assume that all dyadic cubes are contained in

$$\mathcal{D}^M := \{Q \in \mathcal{D} :\ \ell(Q) \geq 2^{-M}\}$$

for some fixed positive integer $M$. Since we will prove

$$\sum_{Q\in\mathcal{D}^M} |Q||\Phi_Q((F_j)_{j\in V})| \lesssim_{n,(p_j)} \prod_{j\in V} \|F_j\|_{L^{p_j}(\mathbb{R}^n)}$$

with the implicit constant independent of $M$, letting $M \to \infty$ will imply the result. The inequality we want to prove is homogeneous, so we may normalize

$$\|F_j\|_{L^{p_j}(\mathbb{R}^n)} = 1 \tag{3.14}$$

for every $j \in V$.

For each $|V|$-tuple of integers $\mathbf{k} = (k_j)_{j\in V} \in \mathbb{Z}^{|V|}$ we define the collection of dyadic cubes

$$\mathcal{P}_{\mathbf{k}} := \{Q : 2^{k_j} \leq \sup_{Q'\supseteq Q} \big[F_j^{d_j}\big]_{Q'}^{d_j^{-1}} < 2^{k_j+1} \text{ for every } j \in V\}$$

The side lengths of the cubes in $P_{\mathbf{k}}$ are bounded from above. Indeed, take $Q \in \mathcal{P}_{\mathbf{k}}$ and choose $Q' \supseteq Q$ with $\big[F_j^{d_j}\big]_{Q'}^{d_j^{-1}} > 2^{k_j-1}$. Then

$$2^{k_j-1} < \big[F_j^{d_j}\big]_{Q'}^{d_j^{-1}} \leq \big[F_j^{p_j}\big]_{Q'}^{p_j^{-1}} = |Q'|^{-p_j^{-1}} \|F_j\|_{L^{p_j}(Q')} \leq |Q'|^{-p_j^{-1}}$$

and hence $|Q| \le |Q'| \le 2^{-p_j(k_j-1)}$.

Let $\mathcal{M}_{\mathbf{k}}$ be the the collection of maximal cubes in $\mathcal{P}_{\mathbf{k}}$ with respect to the set inclusion. For every $Q \in \mathcal{M}_{\mathbf{k}}$ define

$$\mathcal{T}_Q := \{\widetilde{Q} \in \mathcal{P}_{\mathbf{k}} : \widetilde{Q} \subseteq Q\}$$

By monotonicity of $Q \mapsto \sup_{Q' \supseteq Q}$ it follows that $\mathcal{T}_Q$ is a finite convex tree with root $Q$. Note that for different $Q \in \mathcal{M}_{\mathbf{k}}$ the corresponding trees $\mathcal{T}_Q$ cover disjoint areas in $\mathbb{R}^n$ and that these trees partition the collection $\mathcal{P}_{\mathbf{k}}$ for any choice of $\mathbf{k}$.

If $\widetilde{Q} \in \mathcal{L}(\mathcal{T}_Q)$ and $\overline{Q}$ denotes its parent we have $\big[F_j^{d_j}\big]_{\widetilde{Q}}^{d_j^{-1}} \le 2^n \big[F_j^{d_j}\big]_{\overline{Q}}^{d_j^{-1}}$ and by definition of $\mathcal{P}_{\mathbf{k}}$ also $\big[F_j^{d_j}\big]_{\overline{Q}}^{d_j^{-1}} < 2^{k_j+1}$. Thus for all $j \in V$

$$\big[F_j^{d_j}\big]_{\widetilde{Q}}^{d_j^{-1}} \le 2^{k_j+n+1} \lesssim_n 2^{k_j}$$

Proposition 3.1 now yields

$$\Lambda_{\mathcal{T}_Q}((F_j)_{j\in V}) \lesssim_n |Q|\, 2^{\sum_{j\in V} k_j}$$

Since for a cube $\widetilde{Q}$ not contained in any $\mathcal{P}_{\mathbf{k}}$ at least one of the functions $F_j$ vanishes on $\widetilde{Q}$ and therefore does not contribute to $\Lambda$, we decompose and estimate

$$\begin{aligned}\sum_{Q\in\mathcal{D}^M} |Q||\Phi_Q((F_j)_{j\in V})| &= \sum_{\mathbf{k}\in\mathbb{Z}^{|V|}} \sum_{Q\in\mathcal{M}_{\mathbf{k}}} \Lambda_{\mathcal{T}_Q}((F_j)_{j\in V}) \\ &\lesssim_n \sum_{\mathbf{k}\in\mathbb{Z}^{|V|}} 2^{\sum_{j\in V} k_j} \Big(\sum_{Q\in\mathcal{M}_{\mathbf{k}}} |Q|\Big) \qquad (3.15)\end{aligned}$$

Consider the maximal function

$$\mathrm{M}_d F(x) = \sup_{x\in Q\in\mathcal{D}} \Big(\frac{1}{|Q|}\int_Q |F|^d\Big)^{d^{-1}}$$

which is by the maximal theorem (Proposition (2.5)) bounded on $\mathrm{L}^p$ for $d < p \le \infty$. We have $\mathbb{Z}^{|V|} = \cup_{j_0\in V}\mathcal{K}_{j_0}$ where we define

$$\mathcal{K}_{j_0} := \{\mathbf{k} = (k_j)_{j\in V} : p_{j_0}k_{j_0} \ge p_j k_j \ \text{ for every } j \in V\}$$

For any $x \in Q \in \mathcal{P}_{\mathbf{k}}$ we have that

$$\mathrm{M}_{d_{j_0}} F_{j_0}(x) \ge \sup_{Q'\supseteq Q} \big[F_{j_0}^{d_{j_0}}\big]_{Q'}^{d_{j_0}^{-1}} > 2^{k_{j_0}}$$

Hence for $Q \in \mathcal{M}_{\mathbf{k}}$,

$$Q \subseteq \{\mathrm{M}_{d_{j_0}} F_{j_0} \ge 2^{k_{j_0}}\}$$

Since the cubes in $\mathcal{M}_{\mathbf{k}}$ are by maximality disjoint,

$$\sum_{Q\in\mathcal{M}_{\mathbf{k}}} |Q| = \Big|\bigcup_{Q\in\mathcal{M}_{\mathbf{k}}} Q\Big| \le |\{\mathrm{M}_{d_{j_0}} F_{j_0} \ge 2^{k_{j_0}}\}| \qquad (3.16)$$

for any choice of $j_0 \in V$.

Lemma 2.2 yields that for any $1 \le p < \infty$,

$$\sum_{k\in\mathbb{Z}} 2^{pk} |\{\mathrm{M}_d F \ge 2^k\}| \lesssim_p \|\mathrm{M}_d F\|_{L^p}^p$$

Finally combine (3.15) and (3.16) to obtain

$$\begin{aligned}
&\sum_{Q\in\mathcal{D}^M} |Q||\Phi_Q((F_j)_{j\in V})| \lesssim_n \sum_{j_0\in V} \sum_{\mathbf{k}\in\mathcal{K}_{j_0}} 2^{\sum_{j\in V} k_j} |\{\mathrm{M}_{d_{j_0}} F_{j_0} \ge 2^{k_{j_0}}\}| \\
&= \sum_{j_0\in V} \sum_{k_{j_0}\in\mathbb{Z}} 2^{p_{j_0}k_{j_0}} |\{\mathrm{M}_{d_{j_0}} F_{j_0} \ge 2^{k_{j_0}}\}| \prod_{j\in V;\, j\neq j_0} \sum_{\substack{k_j\in\mathbb{Z} \\ p_j k_j \le p_{j_0} k_{j_0}}} 2^{k_j - \frac{p_{j_0}k_{j_0}}{p_j}} \qquad (3.17) \\
&\lesssim_{|V|} \sum_{j_0\in V} \sum_{k_{j_0}\in\mathbb{Z}} 2^{p_{j_0}k_{j_0}} |\{\mathrm{M}_{d_{j_0}} F_{j_0} \ge 2^{k_{j_0}}\}| \\
&\lesssim_{(p_j)} \sum_{j_0\in V} \left\|\mathrm{M}_{d_{j_0}} F_{j_0}\right\|_{L^{p_{j_0}}}^{p_{j_0}} \lesssim_{(p_j)} \sum_{j_0\in V} \|F_{j_0}\|_{L^{p_{j_0}}}^{p_{j_0}} \lesssim_{|V|} 1
\end{aligned}$$

which is the desired estimate, in view of the normalization (3.14). In (3.17) we use $\sum_{j\in V} \frac{1}{p_j} = 1$ and then sum $|V|-1$ convergent geometric series with initial terms in $(1/2, 1]$ and common ratios $1/2$.

It remains to prove the estimate for degenerate sets. Let $V$ be degenerate and $j \in V$. Then, for some $i \in N$ we have that one of the sets $U_0$, $U_1$ given by $(\mathcal{A}', A, U_0, U_1) = \lambda_j(\emptyset, V, i)$ is empty. If $i = n$, then $\Lambda = 0$ and $d_j = 0$.

If $i \neq n$, let us assume without loss of generality that $i = 1$ and $A = \{1, \dots, \alpha\}$ for some $\alpha \in N$. Assume also that $U_1 = \emptyset$ and $\pi_{\cup\mathcal{A}}(U_0) = \{(0)\}$. In other words, we assume that the functions do not depend on the variables $x_1^1, \dots, x_\alpha^1$, which have assigned the normalized characteristic functions. Then, $\Lambda$ can be rewritten as

$$\sum_{Q\in\mathcal{D}(\mathbb{R}^n)} |I_{N\setminus A}| \int_{I_A} \Big[\Big\langle \prod_{j\in V} F_j(x^{(j)}) \Big\rangle_{x_n^0, x_n^1 \in I_n}\Big]_{x^{(0)}_{N\setminus(A\cup\{n\})}, x^{(1)}_{N\setminus(A\cup\{n\})}} dx_A^{(0)}$$

Let us split the sum and consider

$$\sum_{I_{N\setminus A}\in\mathcal{D}(\mathbb{R}^{|N\setminus A|})} |I_{N\setminus A}| \sum_{\substack{I_A\in\mathcal{D}(\mathbb{R}^{|A|}) \\ \ell(I_A)=\ell(I_{N\setminus A})}} \int_{I_A} \Big|\Big[\Big\langle \prod_{j\in V} F_j(x^{(j)}) \Big\rangle_{x_n^0, x_n^1 \in I_n}\Big]_{x^{(0)}_{N\setminus(A\cup\{n\})}, x^{(1)}_{N\setminus(A\cup\{n\})}}\Big| dx_A^{(0)} \qquad (3.18)$$

Denote $\widetilde{V} := \pi_{N\setminus A}(V)$ and $\tilde{j} := j_{N\setminus A}$. Let $(d_{\tilde{j}})_{\tilde{j}}$ be the numbers (3.2) determining the range for $\Lambda_{\widetilde{V}}$. Since $\widetilde{V}$ is non-degenerate,

$$|\Lambda_{\widetilde{V}}((G_{\tilde{j}})_{\tilde{j}\in\widetilde{V}})| \lesssim_{(p_{\tilde{j}})} \prod_{\tilde{j}\in\widetilde{V}} \|G_{\tilde{j}}\|_{L^{p_{\tilde{j}}}(\mathbb{R}^{|N\setminus A|})} \qquad (3.19)$$

whenever $\sum_{\tilde{\jmath}\in\widetilde{V}} \frac{1}{p_{\tilde{\jmath}}} = 1$ and $d_{\tilde{\jmath}} < p_{\tilde{\jmath}} < \infty$ for all $\tilde{\jmath} \in \widetilde{V}$. By definition we have $d_j = d_{\tilde{\jmath}}$ for all $j \in V$. Using the fact that the set of dyadic cubes of scale $k$ partitions $\mathbb{R}^n$ for any $k \in \mathbb{Z}$, in (3.18) interchanging the summation and integration and using (3.19), we can bound (3.18) by

$$\int_{\mathbb{R}^{|A|}} \prod_{j\in V} \left\| F_j(x_A^{(0)}, \cdot) \right\|_{L^{p_j}(\mathbb{R}^{|N\setminus A|})} dx_A^{(0)}$$

whenever $\sum_{j\in V} \frac{1}{p_j} = 1$ and $d_j < p_j < \infty$. To finish the proof it remains to apply Hölder's inequality in the first $|A|$ variables. $\square$

# 4 Examples

In this section we summarize the main steps of the proof on three concrete examples. We mainly work in $\mathbb{R}^3$ and briefly discuss an example in $\mathbb{R}^4$.

We have $V_3 = \{0,1\}^3$, which is depicted in Figure 4. We write $\mathcal{D} = \mathcal{D}(\mathbb{R}^3)$ for the collection of all dyadic cubes $Q = I_1 \times I_2 \times I_3$ in $\mathbb{R}^3$. We denote vertices by $j = (j_1, j_2, j_3) \in V_3$ and we denote $x^{(j)} = (x_1^{j_1}, x_2^{j_2}, x_3^{j_3}) \in \mathbb{R}^3$.

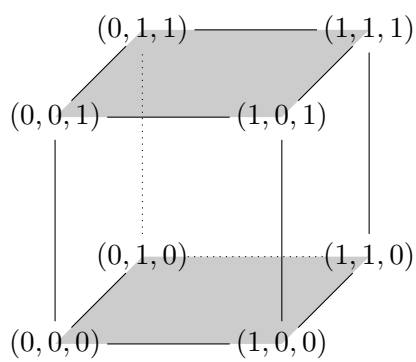


Figure 2: The unit cube in $\mathbb{R}^3$. The Haar function is assigned to the variables corresponding to the shaded faces.

For any $V \subseteq V_3$ the $|V|-$linear form $\Lambda$ now takes the form

$$\begin{aligned}\Lambda((F_j)_{j\in V}) &= \sum_{Q\in\mathcal{D}} \int_{\mathbb{R}^6} \prod_{j\in V} F_j(x^{(j)}) \chi_{I_1}(x_1^0)\chi_{I_1}(x_1^1)\chi_{I_2}(x_2^0)\chi_{I_2}(x_2^1) \\ &\qquad h_{I_3}(x_3^0)h_{I_3}(x_3^1)dx_1^0dx_1^1dx_2^0dx_2^1dx_3^0dx_3^1 \\ &= \sum_{Q\in\mathcal{D}} |Q| \left[ \left\langle \prod_{j\in V} F_j(x^{(j)}) \right\rangle_{x_3^0, x_3^1 \in I_3} \right]_{\substack{x_1^0, x_1^1 \in I_1 \\ x_2^0, x_2^1 \in I_2}}\end{aligned}$$

For any finite convex tree $\mathcal{T}$ and $U \subseteq V_3$ we define the local forms $\Theta^{(i)}_{U,\mathcal{T}}$

$$\Theta^{(1)}_{U,\mathcal{T}}((F_j)_{j\in U}) := \sum_{Q\in\mathcal{T}} \sum_{\alpha,\beta\in\{l,r\}} |Q| \Big[ \Big\langle \prod_{j\in U} F_j(x^{(j)}) \Big\rangle_{x^0_1,x^1_1\in I_1} \Big]_{\substack{x^0_2,x^1_2\in I_{2,\alpha}\\ x^0_3,x^1_3\in I_{3,\beta}}}$$
$$\Theta^{(2)}_{U,\mathcal{T}}((F_j)_{j\in U}) := \sum_{Q\in\mathcal{T}} \sum_{\beta\in\{l,r\}} |Q| \Big[ \Big\langle \prod_{j\in U} F_j(x^{(j)}) \Big\rangle_{x^0_2,x^1_2\in I_2} \Big]_{\substack{x^0_1,x^1_1\in I_1\\ x^0_3,x^1_3\in I_{3,\beta}}}$$
$$\Theta^{(3)}_{U,\mathcal{T}}((F_j)_{j\in U}) := \sum_{Q\in\mathcal{T}} |Q| \Big[ \Big\langle \prod_{j\in U} F_j(x^{(j)}) \Big\rangle_{x^0_3,x^1_3\in I_3} \Big]_{\substack{x^0_1,x^1_1\in I_1\\ x^0_2,x^1_2\in I_2}}$$

We also define

$$\Xi_{U,\mathcal{C}}((F_j)_{j\in U}) := \sum_{Q\in\mathcal{C}} |Q| \Big[ \prod_{j\in U} F_j(x^{(j)}) \Big]_{x^0_1,x^1_1,x^0_2,x^1_2,x^0_3,x^1_3,}$$

for any collection of dyadic cubes $\mathcal{C}$ in $\mathbb{R}^3$. Then we have the telescoping identity

$$\Theta^{(1)}_{U,\mathcal{T}} + \Theta^{(2)}_{U,\mathcal{T}} + \Theta^{(3)}_{U,\mathcal{T}} = \Xi_{\mathcal{L}(\mathcal{T})} - \Xi_{\{Q_{\mathcal{T}}\}} \tag{4.1}$$

(1) $\mathbf{V} = \mathbf{V_3}$

By the discussion in Section 3 we have $d_j = 4$ for all $j \in V_3$. Theorem 1.1 claims the bound

$$|\Lambda((F_j)_{j\in V})| \lesssim_{(p_j)} \prod_{j\in V} \|F_j\|_{L^{p_j}(\mathbb{R}^n)} \tag{4.2}$$

whenever the exponents $p_j$ are such that $\sum_{j\in V} \frac{1}{p_j} = 1$ and $4 < p_j < \infty$.

Let us state the main ingredients of the proof. First note that in this case all forms $\Theta^{(i)}_{\mathcal{A},U,\mathcal{T}}$ and $\Xi_{\mathcal{A},U,\mathcal{T}}$ associated to eligible pairs $(\mathcal{A}, U)$ which appear in the proof of Proposition 3.1, are associated to the vertex set $V = V_3$. In the case of $\Xi_{\mathcal{A},U,\mathcal{T}}$, eligible pairs thus always give rise to the averaging terms involving the Gowers box-inner product of eight functions, which is defined as

$$[F_0, \dots, F_7]_{\square^3(Q)} = \Big[ \prod_{j\in V_3} F_j(x^{(j)}) \Big]_{x^0_1,x^1_1\in I_1,\, x^0_2,x^1_2\in I_2,\, x^0_3,x^1_3\in I_3}$$

The corresponding Gowers box norm of a function $F$ is

$$\|F\|_{\square^3(Q)} = [F, \dots, F]^{1/4}_{\square^3(Q)}$$

In the same way as in the proof of Lemma (3.3) one can see by subsequent applications of the Cauchy-Schwarz inequality that

$$[F_0, \dots, F_7]_{\square^3(Q)} \le \prod_{j\in V_3} \|F_j\|_{\square^3(Q)} \quad \text{and} \quad \|F\|_{\square^3(Q)} \le \Big( \frac{1}{|Q|} \int_Q |F|^4 \Big)^{1/4}$$

Then, for any collection of dyadic cubes $\mathcal{C}$ in $\mathbb{R}^3$ we have

$$\Xi_{V,\mathcal{C}}(F_0,\dots,F_7) = \sum_{Q\in\mathcal{C}} |Q|[F_0,\dots,F_7]_{\square^3(Q)} \le \sum_{Q\in\mathcal{C}} |Q| \prod_{j\in V_3} \left[F_j^4\right]_Q^{1/4}$$

For any finite convex tree we define the local variant $\Lambda_{\mathcal{T}}$ which sums only over the cubes in $\mathcal{T}$. The reduction inequalities which are proven by an application of the Cauchy-Schwarz inequality are[4]

$$\Lambda_{\mathcal{T}}(F_0,\dots,F_7) \le \Theta^{(3)}_{V,\mathcal{T}}(F_{0,1,2,3,0,1,2,3})^{1/2}\Theta^{(3)}_{V,\mathcal{T}}(F_{4,5,6,7,4,5,6,7})^{1/2}$$

and

$$|\Theta^{(1)}_{V,\mathcal{T}}(F_0,\dots,F_7)| \le \Theta^{(1)}_{V,\mathcal{T}}(F_{0,0,2,2,4,4,6,6})^{1/2}\Theta^{(1)}_{V,\mathcal{T}}(F_{1,1,3,3,5,5,7,7})^{1/2}$$
$$|\Theta^{(2)}_{V,\mathcal{T}}(F_0,\dots,F_7)| \le \Theta^{(2)}_{V,\mathcal{T}}(F_{0,1,0,1,4,5,4,5})^{1/2}\Theta^{(2)}_{V,\mathcal{T}}(F_{2,3,2,3,6,7,6,7})^{1/2}$$

For any finite convex tree $\mathcal{T}$ we will prove the estimate

$$\Lambda_{\mathcal{T}}((F_j)_{j\in V}) \lesssim |Q_{\mathcal{T}}| \prod_{j\in V} \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_j^4\right]_Q^{1/4} \tag{4.3}$$

We perform the steps of the proof which consist of alternating applications of reduction inequalities and the telescoping identity. We start with eight positive functions $F_j$ and normalize $\max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_j^4\right]_Q^{1/4} = 1$, since the inequality (4.3) is homogeneous. By scaling invariance we may also assume $|Q_{\mathcal{T}}| = 1$. With these assumptions we need to prove

$$\Lambda_{\mathcal{T}}(F_0,\dots,F_7) \lesssim 1$$

Using the fact that $\mathcal{L}(\mathcal{T})$ partitions $Q_{\mathcal{T}}$ we have

$$\Xi_{\mathcal{L}(\mathcal{T})}(F_{k_0},\dots,F_{k_7}) \le \prod_{i=0}^{7} \max_{Q\in\mathcal{L}(\mathcal{T})} [F_{k_i}^4]_Q^{1/4} \sum_{Q\in\mathcal{L}(\mathcal{T})} |Q| \le 1$$

for any $k_0,\dots,k_7 \in V_3$. Note that $\Theta^{(3)}_{V,\mathcal{T}}(F,\dots,F) \ge 0$, since we can write it as

$$\sum_{Q\in\mathcal{T}} |Q| \Big[ \Big\langle F(x^{(0,0,0)}), F(x^{(1,0,0)}), F(x^{(0,1,0)}), F(x^{(1,1,0)}) \Big\rangle^2_{x_3^0} \Big]_{x_1^0,x_1^1,x_2^0,x_2^1}$$

Analoguosly one sees that $\Theta^{(1)}_{V,\mathcal{T}}(F,\dots,F)$, $\Theta^{(2)}_{V,\mathcal{T}}(F,\dots,F) \ge 0$. Hence (4.1) yields

$$\Theta^{(2)}_{V,\mathcal{T}}(F_j,\dots,F_j) \le \Xi_{\mathcal{L}(\mathcal{T})}(F_j,\dots,F_j) \le 1$$

for any $j \in V_3$. The reduction inequality for $\Theta^{(2)}_{V,\mathcal{T}}$ then gives

$$|\Theta^{(2)}_{V,\mathcal{T}}(F_{0,0,2,2,0,0,2,2})| \le 1$$

[4]For $k_1,\dots,k_m \in \mathbb{N}$ we shortly write $F_{k_1,\dots,k_m} := F_{k_1},\dots,F_{k_m}$.

Since

$$\Theta^{(3)}_{V,\mathcal{T}}(F_{0,0,2,2,0,0,2,2}) \geq 0$$

using the telescoping identity once more gives

$$\Theta^{(1)}_{V,\mathcal{T}}(F_{0,0,2,2,0,0,2,2}) \leq \Xi_{\mathcal{L}(\mathcal{T})}(F_{0,0,2,2,0,0,2,2}) - \Theta^{(2)}_{V,\mathcal{T}}(F_{0,0,2,2,0,0,2,2})$$

and so

$$\Theta^{(1)}_{V,\mathcal{T}}(F_{0,0,2,2,0,0,2,2}) \leq 2 \tag{4.4}$$

Analoguosly we get

$$\Theta^{(1)}_{V,\mathcal{T}}(F_{1,1,3,3,1,1,3,3}) \leq 2 \tag{4.5}$$

and by another application of the reduction inequality, (4.4) and (4.5)

$$\begin{tikzcd}
\Lambda_{\mathcal{T}}(F_{0,1,2,3,4,5,6,7}) \arrow[d] \arrow[dr] & \\
\Theta^{(3)}_V(F_{0,1,2,3,0,1,2,3}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(3)}_V(F_{4,5,6,7,4,5,6,7}) \\
\Theta^{(1)}_V(F_{0,1,2,3,0,1,2,3}) \arrow[d] \arrow[dr] & \Theta^{(2)}_V(F_{0,1,2,3,0,1,2,3}) \\
\Theta^{(1)}_V(F_{0,0,2,2,0,0,2,2}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(1)}_V(F_{1,1,3,3,1,1,3,3}) \\
\Theta^{(2)}_V(F_{0,0,2,2,0,0,2,2}) \arrow[d] \arrow[dr] & \Theta^{(3)}_V(F_{0,0,2,2,0,0,2,2}) \geq 0 \\
\Theta^{(2)}_V(F_{0,0,0,0,0,0,0,0}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(2)}_V(F_{2,2,2,2,2,2,2,2}) \\
\Theta^{(1)}_V(F_{0,0,0,0,0,0,0,0}) \geq 0 & \Theta^{(3)}_V(F_{0,0,0,0,0,0,0,0}) \geq 0
\end{tikzcd}$$

Figure 3: Proof of Proposition 3.1 in form of a tree diagram for the case $n = 3$, $V = V_3$. Induction starts at the bottom. A solid arrow denotes an application of the reduction inequality, while a broken arrow denotes an application of the telescoping identity. We omit the branches which can be deduced by analogy.

$$|\Theta^{(1)}_{V,\mathcal{T}}(F_{0,1,2,3,0,1,2,3})| \leq 2$$

By similar arguments we arrive at

$$|\Theta^{(2)}_{V,\mathcal{T}}(F_{0,1,2,3,0,1,2,3})| \leq 2$$

Using Lemma 3.4 again, together with the two previous estimates we obtain

$$\Theta^{(3)}_{V,\mathcal{T}}(F_{0,1,2,3,0,1,2,3}) \leq 5 \tag{4.6}$$

In exactly the same manner we establish

$$\Theta^{(3)}_{V,\mathcal{T}}(F_{4,5,6,7,4,5,6,7}) \leq 5 \tag{4.7}$$

and finally the reduction inequality and (4.6), (4.7) give

$$\Lambda_{\mathcal{T}}(F_{0,1,2,3,4,5,6,7}) \leq 5$$

which establishes (4.3). To obtain (4.2) one needs to proceed as in Section 3.3.

The proof of the single tree estimate can be represented in the form of a tree diagram as depicted in Figure 4. We inductively bound the terms from the bottom to the top using the reduction inequalities and the telescoping identity. In the diagram we omit the writing of the terms $\Xi_{\mathcal{L}(\mathcal{T})}$, as well as the dependence of the forms $\Theta^{(i)}_{V,\mathcal{T}}$ on the tree $\mathcal{T}$.

(2) $\mathbf{V} = \{(\mathbf{0},\mathbf{0},\mathbf{0}), (\mathbf{1},\mathbf{1},\mathbf{0}), (\mathbf{0},\mathbf{1},\mathbf{1}), (\mathbf{1},\mathbf{0},\mathbf{1})\}$

The vertex set is represented in Figure 4. This is a three-dimensional version of Example

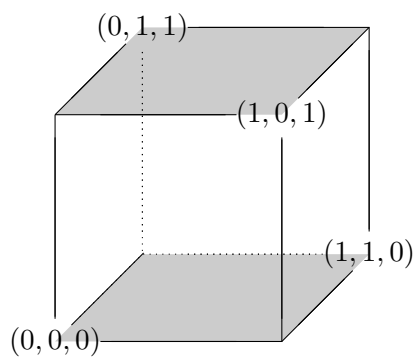


Figure 4: The vertex set $\{(0,0,0), (1,1,0), (0,1,1), (1,0,1)\}$.

3.9. The set $V$ contains all those vertices from $V_3$ with an even number of components equal to 1. The estimate (1.4) holds whenever $\sum_{j\in V} \frac{1}{p_j}$ and $2 < p_j < \infty$ for all $j \in V$.

Let us define the following sets forming eligible pairs $(\mathcal{A}_i, U_j)$. Together with $\mathcal{A}_1 = \{\{3\}\}$ we form eligible pairs with

$$U_0 = \{(0,0,0), (1,1,0)\}$$
$$U_1 = \{(0,1,1), (1,0,1)\}$$

and with $\mathcal{A}_2 = \{\{1,2\},\{3\}\}$ we consider

$$\begin{aligned} U_{00} &= \{(0,0,0)\} \\ U_{01} &= \{(1,1,0)\} \\ U_{10} &= \{(0,1,1)\} \\ U_{11} &= \{(1,0,1)\} \end{aligned}$$

An eligible pair $(\mathcal{A}, U)$ gives rise to the vertex set $\{j \oplus s(A) : A \subseteq \varphi(\mathcal{A}),\ j \in U\} \subseteq V_3$. From Section 3 recall that this set determines the variables, on which the functions in the local forms $\Theta^{(k)}_{\mathcal{A},U}$ depend. We will denote this set by $W_{\mathcal{A},U}$. For short we write $W_{i,k} = W_{\mathcal{A}_i,U_k}$. In our case we have

$$\begin{aligned} W_{1,0} &= \{(0,0,z),(1,1,z)\} \\ W_{1,1} &= \{(0,1,z),(1,0,z)\} \end{aligned}$$

and

$$\begin{aligned} W_{2,00} &= W_{2,01} = \{(y,y,z)\} \\ W_{2,10} &= W_{2,11} = \{(y, y \oplus 1, z)\} \end{aligned}$$

Here, $\{(y,y,z)\}$ is a shorthand for $\{(y,y,z) : y,z \in \{0,1\}\}$ and correspondingly for other examples. Note that $W_{1,0} = W_{2,00}$ and $W_{1,1} = W_{2,10}$, but as opposed to the previous example, these sets are not the same as $V$.

$$\begin{tikzcd}
\Lambda_{\mathcal{T}}(F_{0,3,5,6}) \arrow[d] \arrow[dr] & \\
\Theta^{(3)}_{W_{1,0}}(F_{0,3,0,3}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(3)}_{W_{1,1}}(F_{5,6,5,6}) \\
\Theta^{(1)}_{W_{1,0}}(F_{0,3,0,3}) \arrow[d] \arrow[dr] & \Theta^{(2)}_{W_{1,0}}(F_{0,3,0,3}) \\
\Theta^{(1)}_{W_{2,00}}(F_{0,0,0,0}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(1)}_{W_{2,01}}(F_{3,3,3,3}) \\
\Theta^{(2)}_{W_{2,00}}(F_{0,0,0,0}) \geq 0 & \Theta^{(3)}_{W_{2,00}}(F_{0,0,0,0}) \geq 0
\end{tikzcd}$$

Figure 5: The proof of the single tree estimate in form of a tree diagram for the case $V = \{(0,0,0),(1,1,0),(0,1,1),(1,0,1)\}$. We omit the branches which can be deduced by analogy and involve the forms associated to $W_{2,10}$ and $W_{2,11}$.

The proof of the single tree estimate proceeds inductively as in the previous example and starts with forms involving only one function. The terms $\Xi_{W_{i,k},\mathcal{L}(\mathcal{T})}$ are controlled

$$\Xi_{W_{i,k},\mathcal{L}(\mathcal{T})}((F_j)_{j\in W_{i,k}}) \leq |Q_{\mathcal{T}}| \prod_{j\in W_{i,k}} \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_j^2\right]_Q^{1/2}$$

which is the reason for the restriction of the range of exponents. The reduction inequalities can be read off the tree diagram which represents the proof of the single tree estimate and is depicted in Figure 5.

(3) **A four-dimensional variant**

Consider

$$V = U \cup \{(0,0,0,0),(1,1,1,1)\}$$

where

$$U = \{(1,1,0,0),(1,0,1,0),(0,1,1,0),(1,0,0,1),(0,1,0,1),(0,0,1,1)\}$$

The set $V$ contains all the vertices with an even number of components equal to 1 and is a four-dimensional variant of Example 3.9. The set $U$ can be seen as the vertex set of an octahedron in $\mathbb{R}^3$, if we interpret it in the way that any two vertices are connected by an edge if and only if they differ in exactly two components. See Figure 6.

Assigning a multilinear form to $U$ acting on 6 functions $F_j : \mathbb{R}^4 \to \mathbb{C}$ can be seen as assigning a multilinear form to the octahedron.

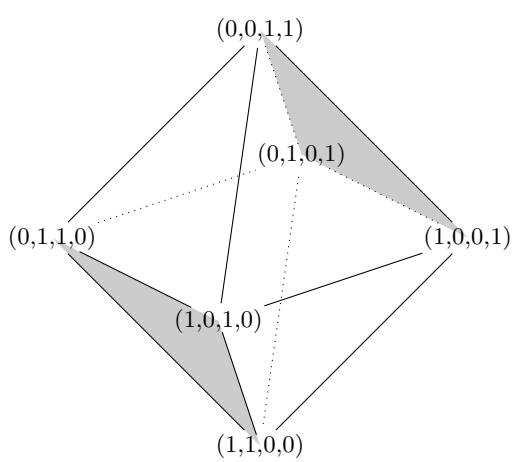


Figure 6: The set $U$ which can be seen as the vertex set of an octahedron. The Haar function in $\Lambda_U$ is assigned to the shaded faces.

We do this in the same way as for the cube. Each vertex of the octahedron corresponds to a function, which depends on the variables corresponding to the 4 adjacent

faces. A face of the octahedron in $\mathbb{R}^3$ is determined by the data $(k, b)$, $k \in \{0, 1, 2, 3\}$, $b \in \{0, 1\}$ and consists of vertices

$$U \cap \left( \{0,1\}^k \times \{b\} \times \{0,1\}^{4-k-1} \right)$$

As discussed in Section 3, Theorem 1.1 yields the estimate for $\Lambda_V$ in the range $4 < p_j < \infty$. In the same range we obtain the estimate (1.4) for $\Lambda_U$. This can be seen by directly determining the $d_j$ or by setting the functions $F_{(0,0,0,0)} = F_{(1,1,1,1)} \equiv 1$ in $\Lambda_{V,\mathcal{T}}$ and continuing the proof as usually.

(4) $\mathbf{V} = \{(\mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{1}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{1}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{1})\}$

The vertex set of this example is depicted in Figure 7. We have $d_0 = d_1 = d_2 = 4$ and $d_4 = 2$. The estimate (1.4) thus holds whenever $\sum_{j \in V} \frac{1}{p_j} = 1$ and $4 < p_0, p_1, p_2 < \infty$

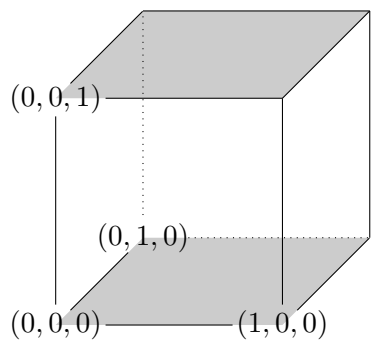


Figure 7: The vertex set $\{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}$.

and $2 < p_4 < \infty$.
The sets which form eligible pairs with $\mathcal{A}_1 = \{\{3\}\}$ are

$$U_0 = \{(0, 0, 0), (1, 0, 0), (0, 1, 0)\}$$
$$U_1 = \{(0, 0, 1)\}$$

With $\mathcal{A}_2 = \{\{1\}, \{3\}\}$ we take

$$U_{00} = \{(0, 0, 0), (0, 1, 0)\}$$
$$U_{01} = \{(1, 0, 0)\}$$

Together with $\mathcal{A}_3 = \{\{1\}, \{2\}, \{3\}\}$ we form eligible pairs with the sets

$$U_{000} = \{(0, 0, 0)\}$$
$$U_{001} = \{(0, 1, 0)\}$$

The first two pairs give rise to the following vertex sets, which determine the variables in the local forms

$$W_{1,0} = \{(0, 0, z), (1, 0, z), (0, 1, z)\}$$
$$W_{1,1} = \{(0, 0, z)\}$$

The second two pairs give

$$W_{2,00} = \{(x,0,z),(x,1,z)\}$$
$$W_{2,01} = \{(x,0,z)\}$$

and the last two

$$W_{3,000} = W_{3,001} = \{(x,y,z)\}$$

Observe that $W_{2,00} = W_{3,000} = V_3$.

$$\begin{tikzcd}
\Lambda_{\mathcal{T}}(F_{0,1,2,4}) \arrow[d] \arrow[dr] & & & \\
\Theta^{(3)}_{W_{1,0}}(F_{0,1,2,0,1,2}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(3)}_{W_{1,1}}(F_{4,4}) \arrow[dr, dashed] \arrow[drr, dashed] \arrow[drrr, dashed] & & \\
\Theta^{(1)}_{W_{1,0}}(F_{0,1,2,0,1,2}) \arrow[d] \arrow[dr] & \Theta^{(2)}_{W_{1,0}}(F_{0,1,2,0,1,2}) & \Theta^{(1)}_{W_{1,1}}(F_{4,4}) \equiv 0 & \Theta^{(2)}_{W_{1,1}}(F_{4,4}) \equiv 0 \\
\Theta^{(1)}_{W_{2,00}}(F_{0,0,2,2,0,0,2,2}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(1)}_{W_{2,01}}(F_{1,1,1,1}) \arrow[dr, dashed] \arrow[drr, dashed] \arrow[drrr, dashed] & & \\
\Theta^{(2)}_{W_{2,00}}(F_{0,0,2,2,0,0,2,2}) \arrow[d] \arrow[dr] & \Theta^{(3)}_{W_{2,00}}(F_{0,0,2,2,0,0,2,2}) & \Theta^{(3)}_{W_{2,01}}(F_{1,1,1,1}) \geq 0 & \Theta^{(2)}_{W_{2,01}}(F_{1,1,1,1}) \equiv 0 \\
\Theta^{(2)}_{W_{3,000}}(F_{0,0,0,0,0,0,0,0}) \arrow[d, dashed] \arrow[dr, dashed] & \Theta^{(2)}_{W_{3,001}}(F_{2,2,2,2,2,2,2,2}) & & \\
\Theta^{(1)}_{W_{3,000}}(F_{0,0,0,0,0,0,0,0}) \geq 0 & \Theta^{(3)}_{W_{3,000}}(F_{0,0,0,0,0,0,0,0}) \geq 0 & &
\end{tikzcd}$$

Figure 8: A schematic presentation of the proof of the single tree estimate when the from is associated to $V = \{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\}$.

The tree diagram representing the single tree estimate is depicted in Figure 8. We draw only the branches corresponding to the $W_i$ which are defined above. The reduction inequalities can be read off the diagram. The terms $\Xi_{W_i,\mathcal{L}(\mathcal{T})}$ arising from applications

of the telescoping identity are controlled as

$$\Xi_{W_{i,k},\mathcal{L}(\mathcal{T})}((F_j)_{j\in W_{i,k}}) \leq |Q_{\mathcal{T}}| \prod_{j\in W_{i,k}} \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_j^4\right]_Q^{1/4}$$

if $k \neq 1$, and

$$\Xi_{W_{1,1},\mathcal{L}(\mathcal{T})}(F_0, F_5) \leq |Q_{\mathcal{T}}| \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_0^2\right]_Q^{1/2} \max_{Q\in\mathcal{L}(\mathcal{T})} \left[F_5^2\right]_Q^{1/2}$$

otherwise. Note that in this case the forms which are associated to degenerate triples $(\mathcal{A}_1, U_1, 1)$, $(\mathcal{A}_1, U_1, 2)$ and $(\mathcal{A}_2, U_{01}, 2)$ equal 0, which immediately terminates the corresponding branch of the tree.

# 5 Final remarks

## 5.1 A simpler proof if $V = V_n,\ p_j = 2^n$

Let us look at the case of $V = V_n$. If we are only interested in proving

$$|\Lambda((F_j)_{j\in V_n})| \lesssim_n \prod_{j\in V_n} \|F_j\|_{L^{2^n}(\mathbb{R}^n)}$$

a simpler proof can be given. For $n = 2$ this is remarked in [7].

If we do not decompose the collection of dyadic cubes into trees and let $Q \in \mathcal{D}$, the telescoping identity becomes

$$\sum_{i=1}^{n} \Theta_{\mathcal{D}}^{(i)}(F_0, \ldots, F_{2^n-1}) = \int_{\mathbb{R}^n} F_0 \cdots F_{2^n-1} \tag{5.1}$$

where we define for $i = 1, \ldots, n$ and any collection of dyadic cubes $\mathcal{C}$ in $\mathbb{R}^n$

$$\Theta_{\mathcal{C}}^{(i)}((F_j)_{j\in V}) = \sum_{Q\in\mathcal{C}} \sum_{\alpha_{i+1},\ldots,\alpha_n\in\{l,r\}} |Q| \left[\left\langle \prod_{j\in V} F_j(x^{(j)}) \right\rangle_{x_i^0, x_i^1\in I_i}\right]_{\substack{x_k^0,\, x_k^1\in I_k \quad \text{for } k=1,\ldots,i-1 \\ x_k^0,\, x_k^1\in I_{k,\,\alpha_k} \quad \text{for } k=i+1,\ldots,n}}$$

To see that, decompose the collection of all dyadic cubes $\mathcal{D}$ into $2^n$ subcollections depending on in which "$n-$th quadrant" they lie, i.e.

$$C_1 := \{Q \in \mathcal{D} : Q \subseteq (\mathbb{R}_+)^n\}$$
$$C_2 := \{Q \in \mathcal{D} : Q \subseteq \mathbb{R}_- \times (\mathbb{R}_+)^{n-1}\}$$

etc. For a fixed positive integer $M$ define $2^n$ finite convex trees

$$\mathcal{C}_1^M := \{Q \in \mathcal{C}_1 : Q \subseteq [0, 2^M)^n,\ \ell(Q) > 2^{-M}\}$$
$$\mathcal{C}_2^M := \{Q \in \mathcal{C}_2 : Q \subseteq [-2^M, 0) \times [0, 2^M)^{n-1},\ \ell(Q) > 2^{-M}\}$$

etc. As $M \to \infty$, the collection $C_i^M$ exhausts $C_i$. If we set

$$\Xi_{\mathcal{C}}((F_j)_{j\in V_n}) := \sum_{Q\in\mathcal{C}} |Q| \big[(F_j)_{j\in V_n}\big]_{\square^n(Q)}$$

we have by Lemma 3.4 applied to $\mathcal{T} = \mathcal{C}_1^M$ that

$$\sum_{i=1}^n \Theta_{\mathcal{T}}^{(i)} = \Xi_{\mathcal{L}(\mathcal{T})} - \Xi_{\{Q_{\mathcal{T}}\}}$$

Define $G(x^{(0)}, x^{(1)}) := \prod_{j\in V_n} F_j(x^{(j)})$ and rewrite

$$\begin{aligned}\Xi_{\mathcal{L}(\mathcal{T})}((F_j)_{j\in V_n}) &= \int_{[0,2^M)^n} \sum_{Q\in\mathcal{L}(\mathcal{T})} \Big[G(x^{(0)},\cdot)\Big]_Q \mathbf{1}_Q(x^{(0)})\, dx^{(0)} \\ &= \int_{[0,2^M)^n} (\mathbb{E}_M G(x^{(0)},\cdot))(x^{(0)})\, dx^{(0)} \qquad (5.2)\end{aligned}$$

where $\mathbb{E}_M$ is the averaging operator of scale $M$ on $\mathbb{R}^n$. If $G(x^{(0)},\cdot)$ is a function constant on dyadic cubes of side length $2^{-k}$ for some $k \in \mathbb{Z}$, then $\mathbb{E}_M G(x^{(0)},\cdot) = G(x^{(0)},\cdot)$ if $M \geq k$. Since we are assuming that the functions $F_j$ are finite dyadic step functions, this is indeed the case and (5.2) equals

$$\int_{[0,2^M)^n} G(x^{(0)}, x^{(0)})\, dx^{(0)}$$

if $M$ is sufficiently large. It remains to consider

$$\Xi_{\{Q_{\mathcal{T}}\}}((F_j)_{j\in V_n}) = \frac{1}{|Q_{\mathcal{T}}|} \int_{Q_{\mathcal{T}}\times Q_{\mathcal{T}}} G(x^{(0)}, x^{(1)})\, dx^{(0)}\, dx^{(1)}$$

By subsequent applications of the Cauchy-Schwarz inequality (in the same way as in the proof of Lemma 3.3) we can estimate this term by

$$|Q_{\mathcal{T}}| \prod_{j\in V_n} \Big(\frac{1}{|Q_{\mathcal{T}}|}\int_{Q_{\mathcal{T}}} F_j^{2^{n-1}}\Big)^{1/2^{n-1}} \leq \frac{1}{2^{Mn}} \prod_{j\in V_n} \|F_j\|_{L^{2^{n-1}}(\mathbb{R}^n)}$$

Now we let $M \to \infty$ which establishes (5.1) for the collection $\mathcal{C}_1$. Analogously we proceed on the other $2^n - 1$ collections and obtain the final identity.

We begin with bounding $\Theta_{\mathcal{D}}^{(n)}$ using the reduction inequality for $\Theta_{\mathcal{D}}^{(n)}$

$$\begin{aligned}\Theta_{\mathcal{D}}^{(n)}(F_0,\dots,F_{2^n-1}) \leq &\Theta_{\mathcal{D}}^{(n)}(F_0,\dots,F_{2^{n-1}-1},F_0,\dots,F_{2^{n-1}-1})^{1/2} \\ &\Theta_{\mathcal{D}}^{(n)}(F_{2^{n-1}},\dots,F_{2^n-1},F_{2^{n-1}},\dots,F_{2^n-1})^{1/2}\end{aligned}$$

Now we apply the telescoping identity (5.1) and obtain

$$\begin{aligned}\Theta_{\mathcal{D}}^{(n)}(F_{0,\dots,2^{n-1}-1,0,\dots,2^{n-1}-1}) &= \|F_0\cdots F_{2^{n-1}-1}\|_{L^2}^2 \\ &\quad - \sum_{i\neq n}\Theta_{\mathcal{D}}^{(i)}(F_{0,\dots,2^{n-1}-1,0,\dots,2^{n-1}-1}) \\ \Theta_{\mathcal{D}}^{(n)}(F_{2^{n-1},\dots,2^n-1,2^{n-1},\dots,2^n-1}) &= \|F_{2^{n-1}}\cdots F_{2^n-1}\|_{L^2}^2 \\ &\quad - \sum_{i\neq n}\Theta_{\mathcal{D}}^{(i)}(F_{2^{n-1},\dots,2^n-1,2^{n-1},\dots,2^n-1})\end{aligned}$$

We proceed by inductively estimating the remaining terms on the right hand side by alternating applications of the reduction inequalities for various $\Theta_{\mathcal{D}}^{(i)}$ (in other words, we keep applying the Cauchy-Schwarz inequality) and the telescoping identity (5.1), arguing in the same way as in the proof of Proposition 3.1. At each step this increases the $L^p-$norm for a power of two until we end up with $\|F_j\|_{L^{2^n}}^{2^n}$. In the end it remains to apply Hölder's inequality in the terms involving products of different functions which arise from the telescoping identity (5.1), such as $\|F_0\cdots F_{2^{n-1}-1}\|_{L^2}^2$.

## 5.2 Discussion on the exponent range

The range of exponents in Theorem 1.1 is not optimal, as can already be seen from the example of the twisted paraproduct. In that case it is possible to use the one-dimensional Calderón-Zygmund decomposition in each fiber $F(\cdot,y)$ or $G(x,\cdot)$, as for all $(x,y)\in\mathbb{R}^2$ we have

$$T_{\mathrm{d}}(F,G)(x,y) = T_{\mathrm{d}}(F(\cdot,y),G(x,\cdot))(x,y)$$

In other words, to compute $T_{\mathrm{d}}$ it is not required to know the information on the whole $F$ and $G$, but only on the one fiber, respectively.

In three dimensions, one could use the same technique on the forms associated to $\{(0,0,0),(1,0,0),(0,1,0),(0,0,1)\}$ which appeared in the Section 4, by considering an operator $T$ satisfying

$$\Lambda(F_0,F_1,F_2,F_4) = \langle F_0, T(F_1,F_2,F_4)\rangle_{L^2(\mathbb{R}^3)}$$

To compute $T(F_1,F_2,F_4)(x^{(0,0,0)})$ we need to know the information only on parts of the functions $F_1(\cdot,x_2^0,x_3^0)$, $F_2(x_1^0,\cdot,x_3^0)$ and $F_4(x_1^0,x_2^0,\cdot)$ With one-dimensional Calderón-Zygmund decomposition performed on $F_4(x_1^0,x_2^0,\cdot)$ as described in [2] or [7], one can show the weak type estimates

$$\|T(F_1,F_2,F_4)\|_{L^{p,\infty}(\mathbb{R}^3)} \lesssim_{p_1,p_2,p_4} \|F_1\|_{L^{p_1}(\mathbb{R}^3)}\|F_2\|_{L^{p_2}(\mathbb{R}^3)}\|F_4\|_{L^1(\mathbb{R}^3)}$$

where $p=(p_1p_2)/(p_1+p_2+p_1p_2)$. Then we use the multilinear interpolation to cover the part $p_4\le 2$, $p_0,p_1,p_2>4$. We proceed further performing the Calderón-Zygmund decomposition on $F_1(\cdot,x_2^0,x_3^0)$ and $F_2(x_1^0,\cdot,x_3^0)$ and interpolating. By duality we obtain the bound for $\Lambda$ in the range $1<d_1,d_2,d_4<\infty$, $4<p_0\le\infty$. However, in the general case we cannot always make use of any similar arguments.

Let us comment on another three-dimensional example. As briefly mentioned in the introduction, if $V$ equals

$$\{(0,0,0),(1,0,0),(0,0,1),(0,1,1)\}$$

Theorem 1.1 yields $d_j = 4$ for all $j \in V$ and thus an empty range of exponents. However, one can establish boundedness in the point $(4,4,4,4)$ by avoiding the decomposition of the dyadic cubes into trees and proceeding as in Section 5.1. Another option is to apply the Cauchy-Schwarz inequality, which bounds $\Lambda$ by two multilinear forms associated to

$$U_1 = \{(0,0,0),(1,0,0),(0,0,1),(1,0,1)\}$$

and

$$U_2 = \{(0,0,0),(0,1,0),(0,0,1),(0,1,1)\}$$

respectively, i.e.

$$|\Lambda_V(F_0,F_1,F_5,F_7)| \leq \Lambda_{U_1}(F_0,F_1,F_0,F_1)^{1/2}\Lambda_{U_2}(F_5,F_7,F_5,F_7)^{1/2}$$

Using the known estimates for the latter forms one establishes

$$|\Lambda_V((F_j)_{j\in V})| \lesssim_{(p_j)} \prod_{j\in V} \|F_j\|_{L^{p_j}(\mathbb{R}^n)}$$

whenever $\frac{1}{p_1}+\frac{1}{p_2}=\frac{1}{p_5}+\frac{1}{p_7}=\frac{1}{2}$ and $2<p_1,p_2,p_5,p_7<\infty$.

Let us conclude with a table of the numbers $d_j$ for all essentially different (up to symmetries) subsets of the vertex set of the unit cube in $\mathbb{R}^3$.

| $\lvert V\rvert$ | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|---|
| **8** | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| **7** | 4 | 4 | 4 | 4 | 4 | 4 | 4 | |
| **6** | 4 | 4 | 4 | 4 | 4 | 4 | | |
| | 4 | 4 | 4 | 4 | 2 | | | 2 |
| | 4 | 4 | 4 | | 4 | 4 | 4 | |
| | 4 | 4 | 4 | | 4 | 4 | | 4 |
| | 4 | 4 | 4 | | | 4 | 4 | 4 |
| **5** | 4 | 4 | 4 | 4 | 2 | | | |
| | 4 | 4 | 4 | | | 2 | 2 | |
| | 4 | 4 | 4 | | 2 | | | 2 |
| | 4 | 4 | 4 | | 4 | 4 | | |
| | 4 | 4 | 4 | | | 4 | | 4 |
| **4** | 0 | 0 | 0 | 0 | | | | |
| | 2 | 2 | | | 2 | 2 | | |
| | 4 | 4 | | | 4 | | 4 | |
| | 4 | 4 | | | 2 | | | 2 |
| | 2 | | | 2 | 2 | | | 2 |
| | 2 | | | 2 | | 2 | 2 | |
| | 2 | 2 | | | | | 2 | 2 |
| | 4 | 4 \| 1 | 4 \| 1 | | 2 \| 1 | | | |
| | 4 | 4 | 4 | | | 2 | | |
| | 4 | 4 | 4 | | | | | 2 |
| **3** | 0 | 0 | 0 | | | | | |
| | 2 \| 2 | 2 \| 1 | | | 2 \| 1 | | | |
| | 2 | 2 | | | | | 2 | |
| | 2 | | | 2 | 2 | | | |
| | 2 | | | 2 | | 2 | | |
| **2** | 0 | 0 | | | | | | |
| | 0 | | | 0 | | | | |
| | 1 | | | | 1 | | | |
| | 1 | | | | | 1 | | |
| | 1 | | | | | | | 1 |
| **1** | 0 | | | | | | | |

Table 1: The numbers $d_j$ for $V \subseteq V_3$. The columns correspond to vertices $(j_1, j_2, j_3)$ represented by the numbers $j_1 + 2j_2 + 4j_3$. The rows are grouped by the cardinalities of the corresponding vertex sets which they represent. Lightly shaded are the rows corresponding to degenerate sets, while the dark row corresponds to $V$ for which Theorem 1.1 yields an empty range. Whenever it is possible to extend the range via Calderón Zygmund decomposition, the lower numbers $d_j$ are listed nearby.

## References


[1] J. Bergh and J. Löfström, *Interpolation spaces, An Introduction*, Springer.-Verlag, New York, 1976.

[2] F. Bernicot, *Fiber-wise Calderón-Zygmund decoposition and application to a bi-dimensional paraproduct*, Illinois J. Math. (2012)

[3] C. Demeter and C. Thiele, *On the two-dimensional bilinear Hilbert transform*, Amer. J. Math., **132** (2010), no. 1, 201-256.

[4] J. Duoandikoetxea, *Fourier Analysis*, Vol. 29 of Graduate Studies in Mathematics, American Math. Society, Providence, 2001.

[5] L. Grafakos, *Classical Fourier Analysis*, Vol. 249 of Graduate Texts in Mathematics. Springer, New York, second edition, 2008.

[6] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2), **171** (2010), no. 3, 1753-1850.

[7] V. Kovač, *Boundedness of the twisted paraproduct*, Revista Matemática Iberoamericana, **28** (2012), no. 4, 1143-1164.

[8] V. Kovač, *Bellman function technique for multilinear estimates and an application to generalized paraproducts*, Indiana University Mathematics Journal, **60** (2011), no. 3, 813-846.

[9] C. Thiele, *Wave Packet Analysis*, CBMS Reg. Conf. Ser. Math., **105**, AMS, Providence, RI, 2006.