University of California
Los Angeles

Applications of the Bellman Function Technique in Multilinear and Nonlinear Harmonic Analysis

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Vjekoslav Kovac

2011

The dissertation of Vjekoslav Kovac is approved.

John Garnett

Stefano Soatto

Terence Tao

Christoph Thiele, Committee Chair

University of California, Los Angeles

2011

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am most indebted to my advisor Professor Christoph Thiele for his constant guidance and invaluable support throughout my studies at UCLA. His word of advice was crucial in many situations, especially while performing the research presented here. Furthermore, I wish to thank the committee members, Professor John Garnett, Professor Stefano Soatto, and Professor Terence Tao, for spending their valuable time reading the manuscript. I also use this opportunity to express my gratitude to Professor Hrvoje Šikić for introducing me to the field of harmonic analysis several years ago and mentoring me for my undergraduate and Master's theses. Last but not least, I have to mention the unconditional support and relentless patience of my parents, Ljubica and Slavko, and my grandmother Terezija.

Part of the material presented in Chapters 1 and 3 is accepted for publication in the form of the paper [23] in *Revista Matemática Iberoamericana* and its inclusion is possible by the kind permission of the editors. Part of the material from Chapters 1 and 5 is first published as [24] in *Proceedings of the AMS* published by the *American Mathematical Society*; currently posted online and to appear in print.

Vita

1979	Born, Zagreb, Croatia
2002	B.S., Mathematics, University of Zagreb, Croatia
2002–2006	Teaching Assistant, Department of Mathematics, University of Zagreb, Croatia
2006	M.S., Mathematics, University of Zagreb, Croatia
2006–2011	Research and Teaching Assistant, Department of Mathematics, University of California, Los Angeles
2007	*Horn-Moez Prize* for Excellence in First Year Graduate Studies

Publications

Kovac, V. (2011) *Boundedness of the twisted paraproduct.* To appear in Revista Matemática Iberoamericana.

Kovac, V. (2011) *Uniform constants in Hausdorff-Young inequalities for the Cantor group model of the scattering transform.* To appear in Proceedings of the American Mathematical Society.

Abstract of the Dissertation

Applications of the Bellman Function Technique in Multilinear and Nonlinear Harmonic Analysis

by

Vjekoslav Kovac

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2011

Professor Christoph Thiele, Chair

Large part of this work is motivated by a question raised by Demeter and Thiele in [14] on establishing L^p estimates for a two-dimensional bilinear operator of paraproduct type, called the *twisted paraproduct*. It is given by

$$T(F,G)(x,y) \;:=\; \sum_{k\in\mathbb{Z}} 2^{2k}\Big(\int_{\mathbb{R}} F(x-s,y)\,\varphi(2^k s)\,ds\Big)\Big(\int_{\mathbb{R}} G(x,y-t)\,\psi(2^k t)\,dt\Big)\,,$$

where φ,ψ are Schwartz functions and $\hat{\psi}(\xi)$ is supported "near" $|\xi|=1$. We confirm this conjecture by proving

$$\|T(F,G)\|_{\mathrm{L}^r(\mathbb{R}^2)} \;\leq\; C_{p,q}\,\|F\|_{\mathrm{L}^p(\mathbb{R}^2)}\|G\|_{\mathrm{L}^q(\mathbb{R}^2)}$$

whenever $1<p,q<\infty$ and $\frac{1}{r}=\frac{1}{p}+\frac{1}{q}>\frac{1}{2}$. As a byproduct of the approach we develop a rather general technique for verifying multilinear estimates. This method is subsequently further applied to show L^p bounds for a class of two-dimensional multilinear forms that generalize (dyadic variants of) both classical paraproducts and the twisted paraproduct.

The remaining material is related to the one-dimensional *Dirac scattering*

transform $f \mapsto \left[\begin{smallmatrix} a(\infty,\xi) & \overline{b(\infty,\xi)} \\ b(\infty,\xi) & \overline{a(\infty,\xi)} \end{smallmatrix}\right]$, defined by the initial value problem

$$\frac{\partial}{\partial x}\begin{bmatrix} a(x,\xi) & \overline{b(x,\xi)} \\ b(x,\xi) & \overline{a(x,\xi)} \end{bmatrix} = \begin{bmatrix} a(x,\xi) & \overline{b(x,\xi)} \\ b(x,\xi) & \overline{a(x,\xi)} \end{bmatrix}\begin{bmatrix} 0 & \overline{f(x)}e^{-2\pi i x\xi} \\ f(x)e^{2\pi i x\xi} & 0 \end{bmatrix},$$
$$a(-\infty,\xi) = 1, \quad b(-\infty,\xi) = 0\,.$$

Muscalu, Tao, and Thiele asked in [33] if the analogues of Hausdorff-Young inequalities are valid with constants independent of p,

$$\|(\ln|a(\infty,\xi)|)^{1/2}\|_{\mathrm{L}^q_\xi(\mathbb{R})} \leq C\,\|f\|_{\mathrm{L}^p(\mathbb{R})}, \qquad \text{for } 1 \leq p \leq 2,\ \tfrac{1}{p}+\tfrac{1}{q}=1\,.$$

We provide positive answer to this question in the case where the exponentials are replaced by the character function of the "d-adic model" of the real line.

Our main tool for all of the attempted problems, both multilinear and nonlinear in nature, is the *Bellman function technique*, briefly described as "systematic induction over scales".

CHAPTER 1

Introduction and overview of the results

1.1 Some problems in multilinear harmonic analysis

An object that motivated much of the modern multilinear time-frequency analysis is the *bilinear Hilbert transform*,

$$\mathrm{BH}_\beta(f,g)(x) := \mathrm{p.v.}\int_{\mathbb{R}} f(x-t)\, g(x-\beta t)\, \frac{dt}{t}\,,$$

where β is a real parameter and $\beta \neq 0, 1$ to avoid degeneracy. This operator was introduced by Calderón in the 1960s, and was motivated by the study of Cauchy integral operator on Lipschitz curves Γ,

$$(\mathrm{C}_\Gamma f)(z) := \lim_{\eta\to 0+}\int_\Gamma \frac{f(\zeta)}{\zeta-(z+i\eta)} d\zeta\,, \qquad z\in\Gamma,\quad f\in \mathrm{L}^2(\Gamma)\,.$$

Calderón conjectured that BH_β maps continuously from $\mathrm{L}^2 \times \mathrm{L}^\infty$ to L^2, as this would have implied L^2-boundedness of C_Γ; see [21]. Although the latter result was established by different methods, the former question became interesting in its own right, and remained unsolved until the late 1990s. Lacey and Thiele proved its boundedness[1] in a pair of breakthrough papers [25], [26]:

$$\|\mathrm{BH}_\beta(f,g)\|_{\mathrm{L}^r(\mathbb{R})} \lesssim_{\beta,p,q} \|f\|_{\mathrm{L}^p(\mathbb{R})}\|g\|_{\mathrm{L}^q(\mathbb{R})}$$

[1]For two nonnegative quantities A and B, we write $A \lesssim B$ if there exists an absolute constant $C \geq 0$ such that $A \leq CB$, and we write $A \lesssim_P B$ if $A \leq C_P B$ holds for some constant $C_P \geq 0$ depending on a parameter P. Finally, we write $A \sim_P B$ if both $A \lesssim_P B$ and $B \lesssim_P A$.

for $\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, $1<p,q\leq\infty$, $\frac{2}{3}<r<\infty$. Their methods have been successfully adapted to many similar problems.

Demeter and Thiele investigated a two-dimensional analogue of the bilinear Hilbert transform in [14]. For any two linear maps $A,B\colon\mathbb{R}^2\to\mathbb{R}^2$ they considered

$$\mathrm{BH}_{A,B}(F,G)(x,y) := \text{p.v.}\int_{\mathbb{R}^2} F\big((x,y)-A(s,t)\big)\, G\big((x,y)-B(s,t)\big)\, K(s,t)\, ds\, dt,$$

where $K\colon\mathbb{R}^2\setminus\{0,0\}\to\mathbb{C}$ is a Calderón-Zygmund kernel, i.e. $\hat{K}$ is a symbol satisfying

$$|\partial^\alpha \hat{K}(\xi,\eta)| \lesssim_\alpha (\xi^2+\eta^2)^{-|\alpha|/2}$$

for all derivatives ∂^α up to some large (unspecified) order. In [14], the bound

$$\|\mathrm{BH}_{A,B}(F,G)\|_{\mathrm{L}^r(\mathbb{R}^2)} \lesssim_{A,B,p,q} \|F\|_{\mathrm{L}^p(\mathbb{R}^2)}\|G\|_{\mathrm{L}^q(\mathbb{R}^2)}$$

is proven in the range $2<p,q<\infty$, $\frac{1}{r}=\frac{1}{p}+\frac{1}{q}>\frac{1}{2}$, and for most cases depending on A and B.

Some instances of A,B were handled by an adaptation of the approach from [25], [26], while some cases lead the authors of [14] to invent a “one-and-a-half-dimensional” time-frequency analysis. On the other extreme, some instances of A,B degenerate to the one-dimensional bilinear Hilbert transform or the pointwise product. Up to the symmetry obtained by considering the adjoints, the only case of A,B that was left unresolved in [14] is

$$T(F,G)(x,y) := \text{p.v.}\int_{\mathbb{R}^2} F(x-s,y)\, G(x,y-t)\, K(s,t)\, ds\, dt\,. \tag{1.1}$$

This case was denoted “Case 6” and, as remarked there, it is largely degenerate but still nontrivial, so the usual wave-packet decompositions showed to be ineffective. This operator can be viewed as the simplest example disclosing certain higher-dimensional phenomena, or more precisely, complications not visible from the perspective of one-dimensional multilinear analysis arising in [25], [26], even in quite general framework such as the one in [13] or [37].

In Chapter 3 we establish bounds on the bilinear multiplier (1.1). It will be enough to discuss the special case of the symbol

$$\hat{K}(\xi,\eta) = \sum_{k\in\mathbb{Z}} \hat{\varphi}(2^{-k}\xi)\,\hat{\psi}(2^{-k}\eta)\,,$$

i.e. the kernel

$$K(s,t) = \sum_{k\in\mathbb{Z}} 2^{2k}\varphi(2^k s)\,\psi(2^k t)\,,$$

with φ and ψ having absolutely bounded first several Schwartz norms. A standard technique of "cone decomposition" (see [50]) then addresses general kernels K. This special case is a two-dimensional bilinear operator of paraproduct[2] type that does not fall into the realm of Calderón-Zygmund theory. We call it the *twisted paraproduct*; the name was suggested by Camil Muscalu. In Chapter 3 we prove boundedness of this operator in a certain range of L^p spaces:

Theorem 3.1 (restated; also see Section 3.1)**.** *Suppose that $\varphi,\psi\in\mathrm{C}^1(\mathbb{R})$ satisfy*

$$\begin{aligned}
&|\varphi(x)|, |\tfrac{d}{dx}\varphi(x)|, |\psi(x)|, |\tfrac{d}{dx}\psi(x)| \;\lesssim\; (1+|x|)^{-3}\,,\\
&\mathrm{supp}(\hat{\psi}) \subseteq \{\xi\in\mathbb{R} \,:\, \tfrac{1}{2}\leq|\xi|\leq 2\}\,.
\end{aligned}$$

Bilinear operator

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

satisfies the estimate

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

whenever $1<p,q<\infty,\ \frac{1}{r}=\frac{1}{p}+\frac{1}{q}>\frac{1}{2}$.

[2]The general notion of a *paraproduct* will be discussed later in this section.

It is worth mentioning that our main method proves the bound in the region $2 < p, q < \infty$ only and then we use the result of Bernicot from [1] to extend it to the case when either $p \leq 2$ or $q \leq 2$.

Another source of motivation for studying multilinear singular operators (such as the one above) is from ergodic theory, i.e. the study of general measure preserving systems. The most fundamental ergodic theorem is *Birkhoff's a.e. convergence theorem* [3]: If $(X, \mathcal{F}, \mu)$ is a probability space, $T\colon X \to X$ is a measure μ preserving transformation, and $f \in \mathrm{L}^1(\mu)$, then the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$$

exists for μ-a.e. $x \in X$. Of the similar spirit is a result by Cotlar [11] with averages replaces by the (weighted) series:

$$\lim_{N\to\infty} \sum_{-N\leq n\leq N,\ n\neq 0} \frac{f(T^n x)}{n}$$

exists for μ-a.e. $x \in X$.

There are many ways to prove Birkhoff's and Cotlar's theorems above, but one possible way is to transfer them from general X to $\mathbb{R}$ using $\mathbb{Z}$ as a mediator. This is the simplest instance of transference principles between ergodic theory and the theory of (linear or multilinear) singular integral operators, called *Calderón's transference principle*; see [7]. If we introduce the *variational* V^r*-norm* of a function $\theta\colon (0,\infty) \to \mathbb{R}$ by

$$\|\theta\|_{\mathrm{V}^r} := \sup_{\varepsilon} |\theta(\varepsilon)| + \sup_{\varepsilon_0<\varepsilon_1<\ldots<\varepsilon_M} \Big(\sum_{j=1}^{M} |\theta(\varepsilon_j) - \theta(\varepsilon_{j-1})|^r \Big)^{\frac{1}{r}},$$

then it is possible to prove the following variational inequalities for $1 < p < \infty$, $f \in \mathrm{L}^p(\mathbb{R})$, and $r > 2$:

$$\Big\| \Big\| \frac{1}{\varepsilon} \int_{-\varepsilon}^{\varepsilon} f(x-t)dt \Big\|_{\mathrm{V}^r_\varepsilon} \Big\|_{\mathrm{L}^p_x(\mathbb{R})} \lesssim_p \|f\|_{\mathrm{L}^p(\mathbb{R})} \,,$$

$$\Big\|\Big\| \int_{|t|>\varepsilon} f(x-t)\frac{dt}{t}\Big\|_{\mathrm{V}^r_\varepsilon}\Big\|_{\mathrm{L}^p_x(\mathbb{R})} \lesssim_p \|f\|_{\mathrm{L}^p(\mathbb{R})} .$$

Such estimates are first discussed by Lépingle [27]. We might recognize the operators on the left as the Hardy-Littlewood averages and the truncated (linear) Hilbert transform respectively. These inequalities imply pointwise convergence theorems, but they are actually more quantitative than "soft convergence results", and so harder to prove.

Following this analogy, the objects corresponding to the bilinear Hilbert transform (and to the bilinear Hardy-Littlewood averages) would be ergodic averages and series of the form

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)g(T^{mn}x)\,, \qquad \lim_{N\to\infty} \sum_{-N\le n\le N,\ n\neq 0} \frac{f(T^n x)g(T^{mn}x)}{n}\,,$$

for a fixed positive integer m. Their a.e. convergence was established by Bourgain [5] and Demeter [12] respectively. Replacing T^m by another transformation S leads to a famous conjecture in ergodic theory.

Conjecture 1.1. *If $(X,\mathcal{F},\mu)$ is a probability space, $T,S\colon X\to X$ are two commuting (i.e. $ST=TS$) measure μ preserving transformations, and $f,g\in\mathrm{L}^\infty(\mu)$, then the limit*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)g(S^n x)$$

exists for μ-a.e. $x\in X$.

This conjecture is still unsolved and seems to be slightly out of reach of the current techniques. Convergence of the same averages in the L^2-norm sense was first proved by Conze and Lesigne in [10]. This was further generalized to the case of several commuting transformations by Tao [46].

When we transfer results from X to $\mathbb{R}$, this leads us to studying "singular bilinear averages"

$$\mathrm{p.v.}\int_{\mathbb{R}} F(x-t,y)\,G(x,y-t)\,\frac{dt}{t} \tag{1.2}$$

and variational inequalities for them. Unfortunately, even the ordinary L^p bounds are not known at this moment, leaving another interesting open problem that has attracted much attention recently, this time in harmonic analysis.

Conjecture 1.2. *Bilinear operator* (1.2) *satisfies*

$$\Big\|\,\mathrm{p.v.}\int_{\mathbb{R}} F(x-t,y)\,G(x,y-t)\,\frac{dt}{t}\,\Big\|_{\mathrm{L}^r_{(x,y)}(\mathbb{R}^2)} \lesssim_{p,q} \|F\|_{\mathrm{L}^p(\mathbb{R}^2)}\|G\|_{\mathrm{L}^q(\mathbb{R}^2)}$$

for some choice of exponents satisfying $\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, $1\leq p,q,r\leq\infty$.

It is easy to notice that this conjectured bound implies boundedness of the bilinear Hilbert transform, both one-dimensional and two-dimensional, even uniformly over the parameters β, A, B and including the case (1.1). It also implies boundedness of the *Carleson operator*

$$(\mathrm{C}f)(x) := \sup_{\eta}\Big|\int_{\mathbb{R}} f(x-t)e^{i\eta t}\frac{dt}{t}\Big|$$

related to a.e. convergence of the Fourier series.

Operators in [14] can be viewed as harmonic analysis analogues of ergodic double averages like

$$\frac{1}{N^2}\sum_{m,n=0}^{N-1} f(S^mT^nx)g(S^{-m}T^nx) \quad \text{or} \quad \frac{1}{N^2}\sum_{m,n=0}^{N-1} f(S^mT^nx)g(S^nx)\,,$$

while the twisted paraproduct operator from Theorem 3.1 is relevant for the averages

$$\frac{1}{N^2}\sum_{m,n=0}^{N-1} f(T^mx)g(S^nx) = \Big(\frac{1}{N}\sum_{m=0}^{N-1} f(T^mx)\Big)\Big(\frac{1}{N}\sum_{n=0}^{N-1} g(S^nx)\Big)\,.$$

However, the latter are immediately seen to converge a.e. by two applications of Birkhoff's ergodic theorem. Here we see that corresponding problems for singular integrals are typically harder as they require summability over scales, but sometimes other techniques do not provide any better understanding of convergence.

The *Demeter-Thiele program* (proposed in [14]) suggests to approach Conjecture 1.1 via harmonic analysis methods, by first examining Conjecture 1.2. For the latter one it was required to begin by establishing bounds on (1.1). Therefore, Theorem 3.1 accomplishes the first step of that program.

In Chapter 4 we define a class of two-dimensional multilinear forms that naturally generalize both classical paraproducts and the twisted paraproduct from Chapter 3. *Paraproducts* first appeared in the work of Pommerenke [43], who showed that $h' = f'g$ and $h(0) = 0$ imply $\|h\|_{\mathrm{H}^2} \lesssim \|f\|_{\mathrm{BMOA}} \|g\|_{\mathrm{H}^2}$ for analytic functions on the unit disk.[3] They were named and used extensively by Bony [4] in the context of his theory of paradifferential operators. Since then, many variants have been studied and they have proven to be a useful concept in various mathematical disciplines. We choose the formulation appearing in [50]. A multilinear form Λ is called a *(classical) model paraproduct* if it is given by

$$\begin{aligned}\Lambda(f_1,\ldots,f_n) &= \sum_{I \text{ dyadic interval}} |I|^{1-\frac{n}{2}} \prod_{i=1}^{n} \big\langle f_i, \varphi_I^{(i)} \big\rangle_{\mathrm{L}^2(\mathbb{R})} \\ &= \sum_{I \text{ dyadic interval}} |I|^{1-\frac{n}{2}} \int_{\mathbb{R}^n} \Big(\prod_{i=1}^{n} f_i(x_i) \varphi_I^{(i)}(x_i) \Big)\, dx_1 \ldots dx_n\,,\end{aligned}$$

where each $\varphi_I^{(i)}$ is a smooth bump function adapted to the interval I. We also assume that for each I at least two of the functions $\varphi_I^{(1)}, \ldots, \varphi_I^{(n)}$ have mean zero. Classical Calderón-Zygmund theory establishes the estimate

$$|\Lambda(f_1,\ldots,f_n)| \lesssim_{n,(p_i)} \prod_{i=1}^{n} \|f_i\|_{\mathrm{L}^{p_i}(\mathbb{R})}$$

for the exponents satisfying $\sum_{i=1}^{n} \frac{1}{p_i} = 1$ and $1 < p_i < \infty$; see [50]. One can consult [45] for applications and [2], [38] for more recent results and references.

[3]The author would like to thank Professor John Garnett for correcting the historical reference on paraproducts.

Significant conceptual complications that do not seem to have been extensively studied arise as soon as one proceeds to higher dimensions. Dyadic variant of the (dualized) twisted paraproduct (trilinear) form attains a similar, "paraproduct-like" shape:

$$\Lambda_{\mathrm{d}}(F,G,H) = \sum_{I\times J \text{ dyadic square}} \int_{\mathbb{R}^4} F(u,y)G(x,v)H(x,y) \\ \varphi_I^{\mathrm{d}}(u)\varphi_I^{\mathrm{d}}(x)\psi_J^{\mathrm{d}}(v)\psi_J^{\mathrm{d}}(y)\, dudxdvdy\,,$$

where $\varphi_I^{\mathrm{d}} := |I|^{-1/2}\mathbf{1}_I$ and $\psi_I^{\mathrm{d}} := |I|^{-1/2}(\mathbf{1}_{I_{\text{left half}}} - \mathbf{1}_{I_{\text{right half}}})$ can be thought of as dyadic versions of bump functions.[4]

In the effort to find a common generalization of the previous two objects, it turns out convenient to associate the "paraproduct-type forms" to finite bipartite undirected graphs. Theorem 4.1 is the main result here, proving their L^p estimates in certain ranges of exponents depending on the structure of a particular multilinear form. This time we prefer to work in the dyadic setting only and do not discuss continuous analogues. Also, we only deal with functions on $\mathbb{R}^2$, for expositional and notational simplicity.

Theorem 4.1 (restated; also see Section 4.1)**.** *Let m,n be positive integers, $E \subseteq \{1,\ldots,m\}\times\{1,\ldots,n\}$, $S\subseteq\{1,\ldots,m\}$, $T\subseteq\{1,\ldots,n\}$, and assume that $|S|\geq 2$ or $|T|\geq 2$. We can interpret E as a bipartite undirected graph, so that each $(i,j)\in E$ determines an edge, and let $\mathrm{d}_{i,j}$ be larger size of the two bipartition classes of the connected component containing that edge. We define a multilinear form acting on $|E|$ functions by*

$$\Lambda\big((F_{i,j})_{(i,j)\in E}\big) \;:=\; \sum_{I\times J \text{ dyadic square}} |I|^{2-\frac{m+n}{2}} \int_{\mathbb{R}^{m+n}} \Big(\prod_{(i,j)\in E} F_{i,j}(x_i,y_j)\Big) \\ \Big(\prod_{i\in S}\psi_I^{\mathrm{d}}(x_i)\Big)\Big(\prod_{i\in S^c}\varphi_I^{\mathrm{d}}(x_i)\Big)\Big(\prod_{j\in T}\psi_J^{\mathrm{d}}(y_j)\Big)\Big(\prod_{j\in T^c}\varphi_J^{\mathrm{d}}(y_j)\Big)\, dx_1\ldots dx_m\, dy_1\ldots dy_n\,.$$

[4] We use the notation $\mathbf{1}_A$ for the characteristic function of a set $A\subseteq\mathbb{R}$.

Then Λ *satisfies the estimate*

$$\big|\Lambda\big((F_{i,j})_{(i,j)\in E}\big)\big| \ \lesssim_{m,n,(p_{i,j})} \prod_{(i,j)\in E} \|F_{i,j}\|_{\mathrm{L}^{p_{i,j}}(\mathbb{R}^2)}$$

for any choice of exponents such that $\sum_{(i,j)\in E} \frac{1}{p_{i,j}} = 1$ *and* $d_{i,j} < p_{i,j} < \infty$ *for each* $(i,j) \in E$.

There seems to be many other higher-dimensional phenomena worth studying. Already "singular bilinear averages" (1.2) are not well understood at the time of writing, and are an object of current research. There is also a reason for caution because (even more singular) *bi-parameter bilinear Hilbert transform*

$$\text{p.v.} \int_{\mathbb{R}^2} F(x-s, y-t)\, G(x+s, y+t)\, \frac{ds}{s} \frac{dt}{t}$$

does not satisfy any L^p estimates, as shown in [31].

1.2 A problem in nonlinear scattering theory

Yet another motivation for research in multilinear analysis comes from the study of nonlinear differential equations. The following context is taken from [33] or [48]. Let $f\colon \mathbb{R} \to \mathbb{C}$ be a compactly supported integrable function. Consider the generalized eigenproblem for the matrix-valued *Dirac operator*[5]:

$$\begin{bmatrix} \frac{d}{dx} & -\overline{f(x)} \\ f(x) & -\frac{d}{dx} \end{bmatrix} \begin{bmatrix} u(x,\xi) \\ v(x,\xi) \end{bmatrix} = \pi i \xi \begin{bmatrix} u(x,\xi) \\ v(x,\xi) \end{bmatrix}, \qquad \xi \in \mathbb{R}\,.$$

Since the special case $f(x) \equiv 0$ has the solution

$$\begin{bmatrix} u_0(x,\xi) \\ v_0(x,\xi) \end{bmatrix} = \begin{bmatrix} Ae^{\pi i x\xi} \\ Be^{-\pi i x\xi} \end{bmatrix},$$

[5]Since the operator is skew-adjoint, we consider only imaginary eigenvalues and find convenient to write them as $\pi i\xi$.

in general it is natural to make the ansatz

$$u(x,\xi) = \overline{a(x,\xi)}e^{\pi i x\xi}\,, \quad v(x,\xi) = b(x,\xi)e^{-\pi i x\xi}\,.$$

This leads to the following initial value problem, written conveniently in the matrix form:

$$\frac{\partial}{\partial x}G(x,\xi) = G(x,\xi)W(x,\xi), \qquad G(-\infty,\xi) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{1.3}$$

where

$$G(x,\xi) = \begin{bmatrix} a(x,\xi) & \overline{b(x,\xi)} \\ b(x,\xi) & \overline{a(x,\xi)} \end{bmatrix}, \qquad W(x,\xi) = \begin{bmatrix} 0 & \overline{f(x)}e^{-2\pi i x\xi} \\ f(x)e^{2\pi i x\xi} & 0 \end{bmatrix}.$$

The problem (1.3) has a unique solution with absolutely continuous functions $a(\cdot,\xi)$ and $b(\cdot,\xi)$ that satisfy the differential equation for a.e. $x\in\mathbb{R}$ and eventually become constant as $x\to-\infty$ or $x\to\infty$. The limit

$$G(\infty,\xi) = \begin{bmatrix} a(\infty,\xi) & \overline{b(\infty,\xi)} \\ b(\infty,\xi) & \overline{a(\infty,\xi)} \end{bmatrix} = \lim_{x\to\infty} \begin{bmatrix} a(x,\xi) & \overline{b(x,\xi)} \\ b(x,\xi) & \overline{a(x,\xi)} \end{bmatrix} \tag{1.4}$$

is a function in $\xi\in\mathbb{R}$, called the *Dirac scattering transform* of f. It is easy to see that all matrices $G(x,\xi)$ must belong to the classical Lie group

$$\mathrm{SU}(1,1) := \left\{ \begin{bmatrix} a & \overline{b} \\ b & \overline{a} \end{bmatrix} \ :\ a,b\in\mathbb{C},\ |a|^2-|b|^2=1 \right\},$$

and so $\xi\mapsto G(\infty,\xi)$ is indeed a function from $\mathbb{R}$ to $\mathrm{SU}(1,1)$. In analogy with the (linear) Fourier transform on $\mathbb{R}$, we also call it the $\mathrm{SU}(1,1)$ *nonlinear Fourier transform* of f, the term originating in [48]. We simply write $G(\xi)$, $a(\xi)$, $b(\xi)$ in place of $G(\infty,\xi)$, $a(\infty,\xi)$, $b(\infty,\xi)$.

Rewriting the system as

$$\begin{aligned} a(x,\xi) &= 1 + \int_{-\infty}^{x} f(t)e^{2\pi i t\xi}\,\overline{b(t,\xi)}dt \\ b(x,\xi) &= \int_{-\infty}^{x} f(t)e^{2\pi i t\xi}\,\overline{a(t,\xi)}dt \end{aligned}$$

and using Picard's iteration, we arrive at the following *multilinear expansions*:

$$
\begin{aligned}
a(\xi) = 1 + \sum_{n=1}^{\infty} \int_{\{x_1>x_2>\ldots>x_{2n-1}>x_{2n}\}} & f(x_1)\overline{f(x_2)}\ldots f(x_{2n-1})\overline{f(x_{2n})} \\
& e^{2\pi i(x_1-x_2+\ldots+x_{2n-1}-x_{2n})\xi}\, dx_1 dx_2 \ldots dx_{2n-1} dx_{2n} \\
= 1 + \frac{1}{2}|\hat{f}(\xi)|^2 + \mathrm{O}\big(\|f\|_{\mathrm{L}^1(\mathbb{R})}^4\big), & \qquad \text{for } \|f\|_{\mathrm{L}^1(\mathbb{R})} \text{ ``small''}, \\
b(\xi) = \sum_{n=1}^{\infty} \int_{\{x_1>x_2>\ldots>x_{2n-2}>x_{2n-1}\}} & f(x_1)\overline{f(x_2)}\ldots \overline{f(x_{2n-2})} f(x_{2n-1}) \\
& e^{2\pi i(x_1-x_2+\ldots-x_{2n-2}+x_{2n-1})\xi}\, dx_1 dx_2 \ldots dx_{2n-2} dx_{2n-1} \\
= \hat{f}(\xi) + \mathrm{O}\big(\|f\|_{\mathrm{L}^1(\mathbb{R})}^3\big), & \qquad \text{for } \|f\|_{\mathrm{L}^1(\mathbb{R})} \text{ ``small''},
\end{aligned}
$$

where

$$\hat{f}(\xi) := \int_{\mathbb{R}} f(x) e^{2\pi i x\xi} dx \,.$$

This motivates the heuristic approximation

$$G(\xi) \approx \begin{bmatrix} 1 + \frac{1}{2}|\hat{f}(\xi)|^2 & \overline{\hat{f}(\xi)} \\ \hat{f}(\xi) & 1 + \frac{1}{2}|\hat{f}(\xi)|^2 \end{bmatrix} \approx \exp \begin{bmatrix} 0 & \overline{\hat{f}(\xi)} \\ \hat{f}(\xi) & 0 \end{bmatrix}$$

and we can think of $\xi \mapsto G(\infty, \xi)$ as a nonlinear version of the (linear) Fourier transform $\hat{f}$. The above expansions proved to be extremely useful when $f \in \mathrm{L}^p(\mathbb{R})$, $1 \leq p < 2$, but there is a reason for caution since even individual summands can be unbounded for $f \in \mathrm{L}^2(\mathbb{R})$; see [34].

Using elementary contour integration one can show a "nonlinear analogue" of the Plancherel theorem:

$$\|(2 \ln |a(\xi)|)^{1/2}\|_{\mathrm{L}^2_\xi(\mathbb{R})} = \|f\|_{\mathrm{L}^2(\mathbb{R})} \,.$$

The first appearance of this identity (although in discrete setting) dates back to [51], [52]. From this equality it seems that $(\ln |a|)^{1/2}$ is the appropriate measure of size for matrices in $\mathrm{SU}(1,1)$, so in the spirit of classical Fourier analysis one can consider nonlinear analogues of Hausdorff-Young inequalities for $1 \leq p < 2$:

$$\|(\ln|a(\xi)|)^{1/2}\|_{\mathrm{L}^q_\xi(\mathbb{R})} \leq C_p \, \|f\|_{\mathrm{L}^p(\mathbb{R})}, \tag{1.5}$$

where p and q are conjugated exponents, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Besides the trivial Riemann-Lebesgue type estimate for $p = 1$, one can show (1.5) for $1 < p < 2$ using the above multilinear expansions, which is first done implicitly in [8], [9] and formulated explicitly in [48]. These papers also prove the maximal version of (1.5), i.e. Menshov-Paley-Zygmund type inequality. Even stronger, variational estimates for $1 \leq p < 2$ are shown recently in [42]. Table 1.1 compares classical estimates for the Fourier transform with their "nonlinear versions".

	Fourier transform	Scattering transform
Riemann-Lebesgue est.	$\big\|\widehat{f}\big\|_{\mathrm{L}^\infty} \leq \|f\|_{\mathrm{L}^1}$	$\big\|(2\ln\|a(\xi)\|)^{\frac{1}{2}}\big\|_{\mathrm{L}^\infty_\xi} \leq \|f\|_{\mathrm{L}^1}$
Hausdorff-Young ineq. $1<p<2,\ 1/p+1/q=1$	$\big\|\widehat{f}\big\|_{\mathrm{L}^q} \leq \|f\|_{\mathrm{L}^p}$	$\big\|(\ln\|a(\xi)\|)^{\frac{1}{2}}\big\|_{\mathrm{L}^q_\xi} \lesssim_p \|f\|_{\mathrm{L}^p}$
Plancherel identity	$\big\|\widehat{f}\big\|_{\mathrm{L}^2} = \|f\|_{\mathrm{L}^2}$	$\big\|(2\ln\|a(\xi)\|)^{\frac{1}{2}}\big\|_{\mathrm{L}^2_\xi} = \|f\|_{\mathrm{L}^2}$
Menshov-Paley-Zygmund $1\leq p<2,\ 1/p+1/q=1$	$\Big\|\sup_x\Big\|\int_{-\infty}^{x} f(t)e^{2\pi it\xi}dt\Big\|\Big\|_{\mathrm{L}^q_\xi} \lesssim \|f\|_{\mathrm{L}^p}$	$\Big\|\sup_x\big(\ln\|a(x,\xi)\|\big)^{\frac{1}{2}}\Big\|_{\mathrm{L}^q_\xi} \lesssim_p \|f\|_{\mathrm{L}^p}$
Strong Carleson est.	$\Big\|\sup_x\Big\|\int_{-\infty}^{x} f(t)e^{2\pi it\xi}dt\Big\|\Big\|_{\mathrm{L}^2_\xi} \lesssim \|f\|_{\mathrm{L}^2}$	$\Big\|\sup_x\big(\ln\|a(x,\xi)\|\big)^{\frac{1}{2}}\Big\|_{\mathrm{L}^2_\xi} \lesssim \|f\|_{\mathrm{L}^2}$ (conjectured)

Table 1.1: Analogy between the linear Fourier transform and the Dirac scattering transform.

However, the truncation method from [8], [9] or [42] gives constants C_p in (1.5) that blow up as $p \to 2-$. For that reason Muscalu, Tao, and Thiele raised the following conjecture in [33].

Conjecture 1.3. *There exists a universal constant $C > 0$ such that for any pair of conjugated exponents $1 \leq p \leq 2$ and $2 \leq q \leq \infty$ and every function f as above* *one has*

$$\|(\ln|a(\xi)|)^{1/2}\|_{\mathrm{L}^q_\xi(\mathbb{R})} \leq C\,\|f\|_{\mathrm{L}^p(\mathbb{R})}.$$

It is interesting to notice that, although we have (1.5) in the endpoint case $p = 2$, we still cannot conclude uniformity of C_p for neighboring values of p. Such anomalies are not possible for linear operators due to the Riesz-Thorin interpolation theorem. However, our transformation $f \mapsto (\ln|a(\cdot)|)^{1/2}$ is not linear, and no standard interpolation result can be applied directly to prove the conjecture.

We support Conjecture 1.3 by instead proving the case when the exponentials are replaced by the character function of the d-adic model of the real line. A rigorous statement is the following theorem, the main result of Chapter 5.

Theorem 5.1 (restated; also see Section 5.1)**.** *Suppose that in the definition of* $W(x,\xi)$ *the exponentials* $e^{2\pi i x\xi}$ *are replaced with*

$$E_d(x,\xi) := e^{(2\pi i/d)\sum_{n\in\mathbb{Z}} x_n \xi_{-1-n}}$$

for base d expansions $x = \sum_{n\in\mathbb{Z}} x_n d^n$, $\xi = \sum_{n\in\mathbb{Z}} \xi_n d^n$. *Then*

$$\|(\ln|a(\xi)|)^{1/2}\|_{\mathrm{L}^q_\xi(\mathbb{R})} \lesssim_d \|f\|_{\mathrm{L}^p(\mathbb{R})}$$

whenever $1 \leq p \leq 2$, $\frac{1}{p} + \frac{1}{q} = 1$.

It is important to emphasize that the implicit constant depends only on d and not on p, q.

Let us remark that our qualitative assumption on f is crucial in order to be able to define the scattering transform properly. If f is merely in $\mathrm{L}^p(\mathbb{R})$ for $1 \leq p < 2$ (but without compact support), then from maximal inequalities in [8], [9] it follows that the limit in (1.4) exists for a.e. $\xi \in \mathbb{R}$, but this is a rather nontrivial result. However, for $f \in \mathrm{L}^2(\mathbb{R})$ that is still an open problem, commonly known as the *nonlinear Carleson theorem*; see Table 1.1 again. Its Cantor group model variant is proven in [33]. One can still extend the definition of the scattering transform to $\mathrm{L}^2(\mathbb{R})$ using density arguments, as in [48].

1.3 A few words on the techniques

After discussing the problems and the results, let us say something about the methods we develop and apply. Exposition of these tools is quite general and somewhat informal in this section. One should refer to the following chapters for a rigorous treatment.

The most important feature of all problems we attempt in this work is the *finite group structure*, either explicit in the problem statement, or implicit in the approach. This structure can be set up by replacing the real line with its *dyadic* (a.k.a. *Walsh*) *model* or more general *d-adic* (a.k.a. *Cantor group*) *model*, having different binary operation, different topology, different character function, etc; see Section 5.1. However, we often only utilize the fact that in these models it is most convenient to work with dyadic step functions instead of smooth bump functions; that the Haar system is the "ideal" wavelet basis and there is no need for more subtle wavelet constructions as in [28], [29]; etc. In all of the treated topics these finite group models are variants of the corresponding better-known continuous models. Sometimes the result can be easily transferred from the finite group version to the continuous version (as in Chapter 3), but sometimes we cannot count on that. In the latter cases we hope that the proof in the dyadic or d-adic model lights the way on how to approach the (typically harder) continuous analogue. It has been a very fruitful practice in time-frequency analysis to first prove the corresponding (carefully formulated) problem in a finite characteristic, most often in the dyadic case, and then work on the technicalities required for the continuous case; compare [49] with [25] and [26], or [35] with [36].

Our main tool is the *Bellman function technique*, but we understand the term quite loosely, so that it can be used interchangeably with "telescoping over scales" or even simply "induction over scales". The very basic idea is that, if we want to control a multiscale quantity $\sum_{n=0}^{N-1}\mathcal{A}_n$, it is enough to construct a controllable quantity $\mathcal{B}_n$ that satisfies $|\mathcal{A}_n| \leq \mathcal{B}_{n+1} - \mathcal{B}_n$, since then we have

$$\Big|\sum_{n=0}^{N-1}\mathcal{A}_n\Big| \leq \mathcal{B}_N - \mathcal{B}_0\,.$$

Sometimes even the special case $\mathcal{A}_n = 0$ is interesting if already the monotonicity $\mathcal{B}_0 \leq \mathcal{B}_1 \leq \ldots \leq \mathcal{B}_{N-1} \leq \mathcal{B}_N$ yields worthwhile information. Furthermore, the above quantities are often sums over d-adic intervals (or squares),

$$\mathcal{A}_n = d^{-n}\sum_{|I|=d^{-n}}\mathcal{A}_I\,, \qquad \mathcal{B}_n = d^{-n}\sum_{|I|=d^{-n}}\mathcal{B}_I\,,$$

which reduces the construction to a single interval (or a square) and its d (or d^2) children, and thus hopefully just to a finitary computation.

Bellman function methods are first successfully applied in probability and functional analysis by Burkholder [6], who used them to compute L^p norms of martingale transforms. The first application to a problem in harmonic analysis is due to Nazarov, Treil, and Volberg in [41]. They also popularized the term "Bellman function", developed the idea into a systematic theory in a series of papers beginning with [39], and explained connections with Bellman's original work in optimal control theory, [40]. Very often Bellman functions provide a simpler alternative to "stopping time arguments", especially when reproving classical results. We make a compromise between the two perspectives and benefit from both of them, as our approach reveals its full power when applied locally, i.e. to a single tree of dyadic intervals or squares.

In Chapter 2 we develop a particular instance of the technique that is suited for proving L^p estimates for two-dimensional multilinear forms having a certain paraproduct-type flavor. Dyadic variant of the twisted paraproduct also falls into that category, allowing its treatment in Chapter 3. Then it turns out very convenient to use the square function introduced by Calderón and generalized by Jones, Seeger, and Wright in [22]. That way we transfer the result to the continuous case, thus finally settling boundedness of (1.1). Chapter 4 is a rather general realization of the technique from Chapter 2.

Chapter 5 deals with d-adic version of Conjecture 1.3. The method of the proof is simply a monotonicity argument over scales, which is typically a privilege of finite group models. The main idea is taken from the "local proof" of the Cantor group model Plancherel theorem given in [33]. A new contribution is the construction of the modified "swapping function" β_d that satisfies a certain $L^p \to L^q$ estimate uniformly in $1 \leq p \leq 2$, $\frac{1}{p} + \frac{1}{q} = 1$ and thus allows us to construct a Bellman function $\mathcal{B}$ independent of p. In the proof we use linear Hausdorff-Young inequalities on $\mathbb{Z}/d\mathbb{Z}$, as a substitute for some cancellation identities from [33].

The scheme of dependencies between chapters is given in Figure 1.1.

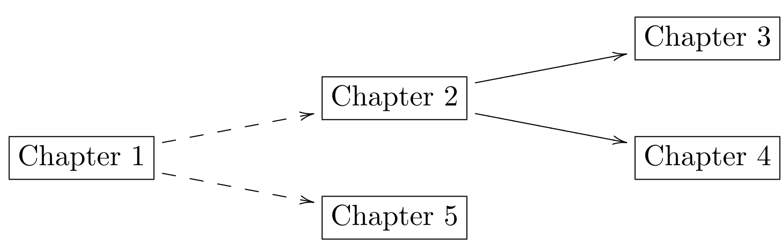

Figure 1.1: Chapter dependencies.

CHAPTER 2

Bellman functions and multilinear estimates

2.1 The general multilinear setting

In this chapter we set up the Bellman function machinery appropriate for proving certain estimates for multilinear operators acting on higher-dimensional functions. Our primary motivation is the twisted paraproduct operator introduced in Chapter 3, but as we will see, the same technique can be used for establishing boundedness of a larger class of multilinear operators.

The main difference between our setup and the Bellman function theory developed in [39], [40], [41] and the subsequent papers by the same authors and their collaborators is that we do not insist on optimality conditions and our approach can only be used to establish positive results about boundedness of operators. This simplifies the theory, since otherwise the corresponding Bellman functions in the sense of optimal control theory would necessarily have to be "infinite-dimensional", i.e. would not depend on finitely many scalar parameters.

For conceptual and notational simplicity we will only work with functions on $\mathbb{R}^2$, but all results and examples easily generalize to higher dimensions. On the other hand, dimension 2 already leads to certain complications that do not seem to have been extensively studied prior to this work. Informally, we say that functions appear in a certain "twisted", "entwined", or "entangled" way in definitions of some multilinear operators — again the simplest and the most representative example being the twisted paraproduct.

In all of the following, functions are assumed to be nonnegative, measurable, bounded, and compactly supported. The general case can always be deduced by splitting into positive/negative, real/imaginary parts, and invoking density arguments.

Let $\mathcal{D}$ denote the family of all *dyadic intervals*, i.e.

$$\mathcal{D} := \left\{ \left[2^k l, 2^k (l+1)\right) \subseteq \mathbb{R} \,:\, k, l \in \mathbb{Z} \right\}.$$

For each dyadic interval I, we denote its left and right halves by I_{left} and I_{right} respectively. Let $\mathcal{C}$ denote the collection of all *dyadic squares*, i.e.

$$\mathcal{C} := \left\{ I \times J \subseteq \mathbb{R}^2 \,:\, I, J \in \mathcal{D},\ |I| = |J| \right\}.$$

We write $|Q|$ for the Lebesgue measure of $Q \in \mathcal{C}$. Every dyadic square Q partitions into four congruent dyadic squares that are called *children* of Q, and conversely, Q is said to be their *parent*.

$$Q = I \times J = \begin{array}{|c|c|} \hline I_{\mathrm{left}} \times J_{\mathrm{right}} & I_{\mathrm{right}} \times J_{\mathrm{right}} \\ \hline I_{\mathrm{left}} \times J_{\mathrm{left}} & I_{\mathrm{right}} \times J_{\mathrm{left}} \\ \hline \end{array}$$

Figure 2.1: A "parent square" divided into four "children squares".

Consequently, each function F defined on a dyadic square Q can be decomposed into four restrictions of F to children of Q.

Our method requires working with structured families of dyadic squares called *convex trees*. A *tree* is a collection $\mathcal{T}$ of dyadic squares in $\mathbb{R}^2$ such that there exists $Q_{\mathcal{T}} \in \mathcal{T}$, called the *root* of $\mathcal{T}$, satisfying $Q \subseteq Q_{\mathcal{T}}$ for every $Q \in \mathcal{T}$. A tree $\mathcal{T}$ is said to be *convex* if whenever $Q_1 \subseteq Q_2 \subseteq Q_3$, and $Q_1, Q_3 \in \mathcal{T}$, then also $Q_2 \in \mathcal{T}$. We will only be working with finite convex trees. Informally, convex trees "do not skip any scales". A *leaf* of $\mathcal{T}$ is a square that is not contained in $\mathcal{T}$, but its parent is. The family of leaves of $\mathcal{T}$ will be denoted $\mathcal{L}(\mathcal{T})$. Notice that for every finite convex tree $\mathcal{T}$ squares in $\mathcal{L}(\mathcal{T})$ partition the root $Q_{\mathcal{T}}$.

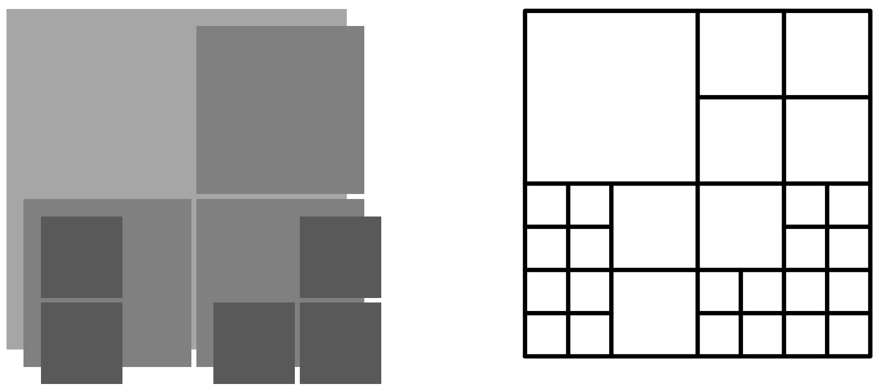

Figure 2.2: A finite convex tree (left) and the partition of its root into leaves (right).

For a function f on a dyadic interval I we denote

$$[f]_I = [f(x)]_{x\in I} := \frac{1}{|I|}\int_I f(x)\,dx\,,$$
$$\langle f\rangle_I = \langle f(x)\rangle_{x\in I} := \frac{1}{|I|}\Big(\int_{I_{\text{left}}} f(x)\,dx - \int_{I_{\text{right}}} f(x)\,dx\Big)\,.$$

We prefer to emphasize the variable in which the average is taken, as for instance in $[F(x,y)]_{x\in I}$ and $\langle F(x,y)\rangle_{x\in I}$ we deal with functions of more than one variable. We simply write $[F(x,y)]_x$ and $\langle F(x,y)\rangle_x$ if the interval I is a generic one or is understood. Notational shortcuts such as

$$\langle\Phi(x,x',y)\rangle_{x,x',y} = \Big\langle\Big\langle\big\langle\Phi(x,x',y)\big\rangle_{x\in I}\Big\rangle_{x'\in I}\Big\rangle_{y\in J}\,,$$

$$\big[\Phi(x_1,x_2,\ldots)\big]_{x_i\in I \text{ for } 1\leq i\leq n} = \Big[\ldots\Big[\big[\Phi(x_1,x_2,\ldots)\big]_{x_1\in I}\Big]_{x_2\in I}\ldots\Big]_{x_n\in I}$$

are also allowed. Finally, for a dyadic square $Q = I\times J$ and a function F on it we denote

$$[F]_Q := \frac{1}{|Q|}\int_Q F(x,y)\,dxdy = \big[F(x,y)\big]_{x\in I,\,y\in J}\,.$$

Let us now turn to multilinear (and multi-sublinear) operators we want to study. A broad class of interesting objects can be reduced to

$$\Lambda_{\mathcal{T}}(F_1,\ldots,F_l) := \sum_{Q\in\mathcal{T}} |Q|\;\mathcal{A}_Q(F_1,\ldots,F_l)\,, \tag{2.1}$$

where $\mathcal{T}$ is a finite convex tree of dyadic squares and $\mathcal{A} = \mathcal{A}_Q(F_1,\ldots,F_l)$ is a "scale-invariant" quantity depending on several two-dimensional functions $F_1,\ldots,F_l$ and a square $Q\in\mathcal{T}$. (A rigorous definition of *paraproduct-type terms* will be given in Section 2.2.) Several illustrative examples (used later in the text) are

$$\mathcal{A}_{I\times J}(F) := 3\big[[F(x,y)]_{x\in I}\langle F(x,y)\rangle_{x\in I}^2\big]_{y\in J}\,,$$

$$\mathcal{A}_{I\times J}(F,G) := \big\langle[F(x,y)]_{x\in I}^2\big\rangle_{y\in J}\big\langle[G(x,y)]_{x\in I}\big\rangle_{y\in J}\,, \tag{2.2}$$

$$\mathcal{A}_{I\times J}(F,G,H) := \Big\langle\big[F(u,y)G(x,v)H(x,y)\big]_{x,u\in I}\Big\rangle_{y,v\in J}\,. \tag{2.3}$$

Sometimes we want to deal with sums over infinite collections of squares Q, but this extension is immediate if we do not allow any constants to depend on $\mathcal{T}$. A special instance is when $\mathcal{T}$ is replaced by the collection of all dyadic squares $\mathcal{C}$, but in certain applications this specialization is not general enough.

Let $\mathcal{B} = \mathcal{B}_Q(F_1,\ldots,F_l)$ be another "scale-invariant" quantity depending on functions $F_1,\ldots,F_l$ and a dyadic square Q. We define the *first order difference* of $\mathcal{B}$, denoted $\Box\mathcal{B}$, as the following quantity:

$$\begin{aligned}\Box\mathcal{B}_{I\times J}(F_1,\ldots,F_l) := {} & \frac{1}{4}\mathcal{B}_{I_{\text{left}}\times J_{\text{left}}}(F_1,\ldots,F_l) + \frac{1}{4}\mathcal{B}_{I_{\text{left}}\times J_{\text{right}}}(F_1,\ldots,F_l) \\ & + \frac{1}{4}\mathcal{B}_{I_{\text{right}}\times J_{\text{left}}}(F_1,\ldots,F_l) + \frac{1}{4}\mathcal{B}_{I_{\text{right}}\times J_{\text{right}}}(F_1,\ldots,F_l) \\ & - \mathcal{B}_{I\times J}(F_1,\ldots,F_l)\,.\end{aligned}$$

For instance,

$$\mathcal{B}_{I\times J}(F) := \big[[F(x,y)]^3_{x\in I}\big]_{y\in J} \tag{2.4}$$

leads to

$$\begin{aligned}\Box\mathcal{B}_{I\times J}(F) &= \frac{1}{4}\big[[F(x,y)]^3_{x\in I_{\text{left}}}\big]_{y\in J_{\text{left}}} + \frac{1}{4}\big[[F(x,y)]^3_{x\in I_{\text{left}}}\big]_{y\in J_{\text{right}}} \\ &\quad + \frac{1}{4}\big[[F(x,y)]^3_{x\in I_{\text{right}}}\big]_{y\in J_{\text{left}}} + \frac{1}{4}\big[[F(x,y)]^3_{x\in I_{\text{right}}}\big]_{y\in J_{\text{right}}} \\ &\quad - \big[[F(x,y)]^3_{x\in I}\big]_{y\in J} \\ &= \frac{1}{2}\big[[F(x,y)]^3_{x\in I_{\text{left}}}\big]_{y\in J} + \frac{1}{2}\big[[F(x,y)]^3_{x\in I_{\text{right}}}\big]_{y\in J} - \big[[F(x,y)]^3_{x\in I}\big]_{y\in J} \\ &= 3\big[[F(x,y)]_{x\in I}\langle F(x,y)\rangle^2_{x\in I}\big]_{y\in J}\,. \end{aligned}\tag{2.5}$$

Above we used obvious identities

$$\left.\begin{aligned}[f(x)]_{x\in I} &= \tfrac{1}{2}\big([f(x)]_{x\in I_{\text{left}}} + [f(x)]_{x\in I_{\text{right}}}\big), \\ \langle f(x)\rangle_{x\in I} &= \tfrac{1}{2}\big([f(x)]_{x\in I_{\text{left}}} - [f(x)]_{x\in I_{\text{right}}}\big), \\ [f(x)]_{x\in I_{\text{left}}} &= [f(x)]_{x\in I} + \langle f(x)\rangle_{x\in I}, \\ [f(x)]_{x\in I_{\text{right}}} &= [f(x)]_{x\in I} - \langle f(x)\rangle_{x\in I},\end{aligned}\right\} \tag{2.6}$$

and $\frac{1}{2}(A+B)^3 + \frac{1}{2}(A-B)^3 - A^3 = 3AB^2$.

We provide more examples in the following two tables. Table 2.1 contains examples where all functions have (x,y) as their argument and in those cases we omit writing the variables x,y. Therefore, F stands for $F(x,y)$, etc. In Table 2.2 we give a couple of examples with several different variables involved, and thus those variables have to be explicitly denoted. Some entries in these tables would require rather lengthy computation. Much more convenient way of obtaining such identities is using Theorem 2.2 from the next section.

$\mathcal{B}$	$\Box\mathcal{B}$
$\big[[F]_x^2\big]_y$	$\big[\langle F\rangle_x^2\big]_y$
$\big[[F]_x^3\big]_y$	$3\,\big[[F]_x\langle F\rangle_x^2\big]_y$
$[F]_{x,y}^3$	$3\,[F]_{x,y}\big([\langle F\rangle_x]_y^2 + \langle [F]_x\rangle_y^2 + \langle F\rangle_{x,y}^2\big) + 6\,\big[\langle F\rangle_x\big]_y\,\big\langle [F]_x\big\rangle_y\,\langle F\rangle_{x,y}$
$\big[[F]_x^2\big]_y\,[G]_{x,y}$	$\big[\langle F\rangle_x^2\big]_y\,[G]_{x,y} + \big\langle [F]_x^2\big\rangle_y\,\big\langle [G]_x\big\rangle_y + \big\langle \langle F\rangle_x^2\big\rangle_y\,\big\langle [G]_x\big\rangle_y$ $+2\,\big[[F]_x\langle F\rangle_x\big]_y\,\big[\langle G\rangle_x\big]_y + 2\,\big\langle [F]_x\langle F\rangle_x\big\rangle_y\,\langle G\rangle_{x,y}$
$[F]_{x,y}^2[G]_{x,y}$	$\big([\langle F\rangle_x]_y^2 + \langle [F]_x\rangle_y^2 + \langle F\rangle_{x,y}^2\big)\;[G]_{x,y}$ $+\,2\,\big([F]_{x,y}\,\big[\langle F\rangle_x\big]_y + \langle F\rangle_{x,y}\,\big\langle [F]_x\big\rangle_y\big)\,[\langle G\rangle_x]_y$ $+\,2\,\big([F]_{x,y}\,\big\langle [F]_x\big\rangle_y + \langle F\rangle_{x,y}\,\big[\langle F\rangle_x\big]_y\big)\,\langle [G]_x\rangle_y$ $+\,2\,\big([F]_{x,y}\langle F\rangle_{x,y} + \big\langle [F]_x\big\rangle_y\,\big[\langle F\rangle_x\big]_y\big)\,\langle G\rangle_{x,y}$

Table 2.1: A sample table of first order differences.

$\mathcal{B}$	$\Box\mathcal{B}$
$\big[[F(u,y)H(x,y)]_y^2\big]_{x,u}$	$\big[\langle F(u,y)H(x,y)\rangle_y^2\big]_{x,u} + \big\langle [F(u,y)H(x,y)]_y^2\big\rangle_{x,u}$ $+\big\langle \langle F(u,y)H(x,y)\rangle_y^2\big\rangle_{x,u}$
$\big[[F(u,y)F(u,v)]_u^2\big]_{y,v}$	$\big[\langle F(u,y)F(u,v)\rangle_u^2\big]_{y,v} + \big\langle [F(u,y)F(u,v)]_u^2\big\rangle_{y,v}$ $+\big\langle \langle F(u,y)F(u,v)\rangle_u^2\big\rangle_{y,v}$

Table 2.2: More complicated examples of first order differences.

Let us now suppose that we have found a quantity $\mathcal{B}$ such that $\mathcal{A} \leq \Box\mathcal{B}$, i.e. more precisely

$$\mathcal{A}_Q(F_1,\ldots,F_l) \leq \Box\mathcal{B}_Q(F_1,\ldots,F_l)\,, \tag{2.7}$$

for all squares $Q \in \mathcal{T}$ and all nonnegative functions $F_1, \ldots, F_l$. By fixing an l-tuple of functions and applying (2.7) to an arbitrary $Q \in \mathcal{T}$, we get

$$\mathcal{A}_Q(F_1, \ldots, F_l) \;\leq\; \frac{1}{4} \sum_{\widetilde{Q} \text{ is a child of } Q} \mathcal{B}_{\widetilde{Q}}(F_1, \ldots, F_l) \;-\; \mathcal{B}_Q(F_1, \ldots, F_l)\,.$$

Multiplying by $|Q|$ and summing over $Q \in \mathcal{T}$ we obtain

$$\Lambda_{\mathcal{T}}(F_1, \ldots, F_l) \leq \sum_{Q \in \mathcal{L}(\mathcal{T})} |Q|\; \mathcal{B}_Q(F_1, \ldots, F_l) - |Q_{\mathcal{T}}|\; \mathcal{B}_{Q_{\mathcal{T}}}(F_1, \ldots, F_l) \tag{2.8}$$

for $\Lambda_{\mathcal{T}}$ as in (2.1). To verify (2.8), one only has to note that each term

$$|Q|\; \mathcal{B}_Q(F_1, \ldots, F_l) \qquad \text{for } Q \in \mathcal{T} \setminus \{Q_{\mathcal{T}}\}$$

appears exactly once with a positive sign and exactly once with a negative sign and thus all terms but those appearing in (2.8) cancel themselves. Here is where we crucially use the tree structure — a general collection of squares would not work.

The expression $\mathcal{B}$ can be called a *Bellman function* for $\Lambda_{\mathcal{T}}$. It is certainly not unique and other properties required of $\mathcal{B}$ in the actual problem will further narrow the choice. Usefulness of (2.8) is in the fact that it reduces controlling[1] a multi-scale quantity $\Lambda_{\mathcal{T}}$ to controlling two single-scale expressions: one on the level of the "finest scales" $\mathcal{L}(\mathcal{T})$ and another one on the level of the "roughest scale" $Q_{\mathcal{T}}$.

Let us illustrate this with an example that is an immediate consequence of Table 2.1. More important examples are given in Section 2.3. Define

$$\Xi(F,G) := \sum_{I\times J \in \mathcal{C}^{++}} |I \times J|\; \big\langle [F(x,y)]_{x\in I}^2 \big\rangle_{y\in J} \, \big\langle [G(x,y)]_{x\in I} \big\rangle_{y\in J}\;,$$

[1] If we want to control $|\Lambda_{\mathcal{T}}|$, then we also need to find $\mathcal{B}$ with $-\mathcal{A} \leq \Box\mathcal{B}$, or better immediately choose $\mathcal{B}$ satisfying $|\mathcal{A}| \leq \Box\mathcal{B}$.

where $\mathcal{C}^{++}$ is a collection of all dyadic squares in the first quadrant of $\mathbb{R}^2$. This is a particular, quadratic, and dyadic case of one of the forms appearing in [14] and can be handled using the theory of maximal truncated singular integrals. Instead, we present a simple Bellman function proof of [2]

$$\Xi(F,G) \;\lesssim\; \int_0^\infty\!\!\int_0^\infty \big(F(x,y)^3 + G(x,y)^3\big)\,dx\,dy\,,$$

for nonnegative functions F and G.

To keep arguments finite, we also introduce

$$\mathcal{C}_N^{++} := \big\{ I\times J \in \mathcal{C} \,:\, I,J \subseteq [0,2^N),\ |I|=|J|\geq 2^{-N} \big\}\,,$$

for any positive integer N. Define $\mathcal{B} = \mathcal{B}_{I\times J}(F,G)$ by the formula:

$$\mathcal{B} := \frac{1}{3}\,\big[[F]_x^3\big]_y + \frac{1}{3}[F]_{x,y}^3 + \big[[F]_x^2\big]_y\,[G]_{x,y} + [F]_{x,y}^2[G]_{x,y} + 2[F]_{x,y}[G]_{x,y}^2 + \frac{4}{3}[G]_{x,y}^3\,.$$

Using the entries from Table 2.1, rearranging and grouping the terms, and transforming some terms with the aid of (2.6), we arrive at the expression:

$$\begin{aligned}
\Box\mathcal{B} = {} & \big\langle [F(x,y)]_{x\in I}^2 \big\rangle_{y\in J}\, \langle [G(x,y)]_{x\in I}\rangle_{y\in J} \\
& + [F(x,y)]_{x\in I,\,y\in J}\Big(\big\langle \big[F(x,y)+G(x,y)\big]_{x\in I}\big\rangle_{y\in J}^2 + \langle [G(x,y)]_{x\in I}\rangle_{y\in J}^2 \Big) \\
& + [G(x,y)]_{x\in I,\,y\in J}\, \big\langle \big[F(x,y)+2G(x,y)\big]_{x\in I}\big\rangle_{y\in J}^2 \\
& + \frac{1}{2}\Big[[F(x,y)]_{x\in I} \big\langle F(x,y) + [G(x,v)]_{v\in J_{\mathrm{left}}}\big\rangle_{x\in I}^2 \Big]_{y\in J_{\mathrm{left}}} \\
& + \frac{1}{2}\Big[[F(x,y)]_{x\in I} \big\langle F(x,y) + [G(x,v)]_{v\in J_{\mathrm{right}}}\big\rangle_{x\in I}^2 \Big]_{y\in J_{\mathrm{right}}} \\
& + \frac{1}{2}\big[G(x,y)\big]_{x\in I,\,y\in J_{\mathrm{left}}} \Big[\big\langle F(x,y) + 2[G(x,v)]_{v\in J_{\mathrm{left}}}\big\rangle_{x\in I}^2 \Big]_{y\in J_{\mathrm{left}}} \\
& + \frac{1}{2}\big[G(x,y)\big]_{x\in I,\,y\in J_{\mathrm{right}}} \Big[\big\langle F(x,y) + 2[G(x,v)]_{v\in J_{\mathrm{right}}}\big\rangle_{x\in I}^2 \Big]_{y\in J_{\mathrm{right}}}
\end{aligned}$$

[2] Note that Ξ is not always nonnegative and that for illustration we only bound it from above.

$$
\begin{aligned}
&+ \frac{1}{2}\big[F(x,y)\big]_{x\in I,\, y\in J_{\mathrm{left}}} \big[\big\langle F(x,y)+G(x,y)\big\rangle_{x\in I}\big]^2_{y\in J_{\mathrm{left}}} \\
&+ \frac{1}{2}\big[F(x,y)\big]_{x\in I,\, y\in J_{\mathrm{right}}} \big[\big\langle F(x,y)+G(x,y)\big\rangle_{x\in I}\big]^2_{y\in J_{\mathrm{right}}} \\
&+ \frac{1}{2}\big[\langle F(x,y)\rangle_{x\in I}\big]^2_{y\in J_{\mathrm{left}}} \big[G(x,y)\big]_{x\in I,\, y\in J_{\mathrm{left}}} \\
&+ \frac{1}{2}\big[\langle F(x,y)\rangle_{x\in I}\big]^2_{y\in J_{\mathrm{right}}} \big[G(x,y)\big]_{x\in I,\, y\in J_{\mathrm{right}}} \,.
\end{aligned}
$$

Since we have assumed $F, G \geq 0$, each row but the first one is nonnegative, so $\Box\mathcal{B} \geq \mathcal{A}$, where $\mathcal{A} = \mathcal{A}_{I\times J}(F,G)$ is defined by (2.2). Inequality (2.8) with $\mathcal{T} = \mathcal{C}_N^{++}$ then gives

$$
\sum_{I\times J\in\mathcal{C}_N^{++}} |I\times J| \ \big\langle [F(x,y)]_x^2\big\rangle_y \, \langle [G(x,y)]_x\rangle_y \ \leq \sum_{\substack{I,J\subseteq[0,2^N)\\ |I|=|J|=2^{-N-1}}} |I\times J| \ \mathcal{B}_{I\times J}(F,G)\,.
$$

Observe that for nonnegative f by Jensen's inequality for powers we have

$$
[f(x)]_{x\in I} \leq [f(x)^2]_{x\in I}^{1/2} \leq [f(x)^3]_{x\in I}^{1/3}\,,
$$

which easily implies

$$
\mathcal{B}_{I\times J}(F,G) \leq 6\Big([F(x,y)^3]_{x\in I,y\in J} + [G(x,y)^3]_{x\in I,y\in J}\Big)\,.
$$

We have obtained

$$
\sum_{I\times J\in\mathcal{C}_N^{++}} |I\times J| \ \big\langle [F(x,y)]_x^2\big\rangle_y \, \langle [G(x,y)]_x\rangle_y \leq 6\int_0^{2^N}\!\!\int_0^{2^N} \big(F(x,y)^3 + G(x,y)^3\big)\,dx\,dy\,,
$$

and it remains to let $N\to\infty$.

2.2 First order difference formula for paraproduct-type terms

In the previous example we see that the main effort of finding an appropriate Bellman function $\mathcal{B}$ relevant for bounding $\Lambda_{\mathcal{T}} = \sum_{Q\in\mathcal{T}} |Q|\,\mathcal{A}_Q$ consist of the following steps:

- Deciding which terms $\mathcal{B}_1, \ldots, \mathcal{B}_M$ will appear in the Bellman function.
- Computing $\Box\mathcal{B}_i$ for each of these terms $\mathcal{B}_i$.
- Assembling $\mathcal{B}$ as a linear combination of these terms, $\mathcal{B}=\alpha_1\mathcal{B}_1+\ldots+\alpha_M\mathcal{B}_M$.

In the mentioned example, we did not give any motivation for the first step and we did not give any hints on how to easily perform the second one. Fortunately, there is a special type of terms $\mathcal{B}_i$ for which a simple rule for determining $\Box\mathcal{B}_i$ applies. Moreover, the inverse procedure of finding the needed terms $\mathcal{B}_i$ from their first order differences is even more straightforward.

Definition 2.1. *A* paraproduct-type term *is a formal finite product consisting of finitely many two-dimensional functions F, G, H, F_1, F_2, $\ldots$ in finitely many variables $u, v, x, y, x_1, x_2, \ldots,$ with finitely many inserted brackets of two types, $[\cdot]$ and $\langle\cdot\rangle$, each with a variable in its subscript, and satisfying the following set of rules:*

- *The brackets are either nested or enclose disjoint factors of the product, i.e. the string is a "meaningful algebraic expression".*
- *Each bracket is subscripted with a single variable taking values in either I or J and every pair of nested brackets has different subscript variables.*
- *Each argument variable of any of the functions also appears in a subscript of some bracket enclosing it.*

An averaging paraproduct-type term *is a paraproduct-type term containing only brackets of type $[\cdot]$. Linear combinations of paraproduct-type terms are called* paraproduct-type expressions*.*

We also regard $\mathcal{B} \equiv 0$ to be a (trivial) averaging paraproduct-type term. All terms in Tables 2.1 and 2.2 are paraproduct-type terms, but only the terms in the left columns are averaging. Note that formally our definition does not allow appearance of any nonlinear operations, but powers can be interpreted simply as abbreviations for products of repeating factors. Thus, for example (2.4) is just a shorter form of

$$\Big[[F(x,y)]_{x\in I}[F(x,y)]_{x\in I}[F(x,y)]_{x\in I}\Big]_{y\in J}. \tag{2.9}$$

Theorem 2.2. *Let $\mathcal{B}$ be an averaging paraproduct-type term. Then $\Box\mathcal{B}$ is equal to the sum of all non-averaging paraproduct-type terms obtained by replacing some brackets of type $[\cdot]$ with brackets of type $\langle\cdot\rangle$ in any possible way such that:*

- *The number of replacements corresponding to variables in I is even.*
- *The number of replacements corresponding to variables in J is even.*
- *At least two replacements are made, i.e. the derived terms are not averaging.*

In particular, if $\mathcal{B}$ contains m brackets corresponding to variables in I and n brackets corresponding to variables in J, then $\Box\mathcal{B}$ will consist of $2^{m+n-2}-1$ (possibly repeating) terms.

For instance, take (2.4) and first rewrite it as (2.9). There are 3 brackets corresponding to variables in I and 1 bracket corresponding to a variable in J, giving us only $2^{3+1-2}-1=3$ possibilities. All 3 of these terms are equal and we arrive at (2.5). As another example, take the last entry in Table 2.1, and rewrite it as

$$\big[[F(x,y)]_{x\in I}\big]_{y\in J}\big[[F(x,y)]_{x\in I}\big]_{y\in J}\big[[G(x,y)]_{x\in I}\big]_{y\in J}.$$

There are 3 brackets corresponding to variables in I and 3 brackets corresponding to variables in J, leaving us with $2^{3+3-2}-1=15$ choices. We see that the result in Table 2.1 also has $3 + 2 \cdot 2 + 2 \cdot 2 + 2 \cdot 2 = 15$ terms. Easy combinatorial reasoning recovers all of them.

Proof of Theorem 2.2. The first step is to realize that every averaging paraproduct-type term can be written in the following standard form:

$$\mathcal{B} = \big[\Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\big]_{x_1,\ldots,x_m \in I,\ y_1,\ldots,y_n \in J}.$$

Here $\Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)$ depends on functions $F_1, \ldots, F_l$. To achieve that reduction, one has to rename all duplicate variables, i.e. those variables x that appear in the original term in the subscript of more than one bracket $[\cdot]_x$. After that is done, one has to gradually "move" all terms inside the brackets, concentrating on one bracket at a time. For example, (2.9) can be rewritten as

$$\big[\underbrace{F(x_1, y)F(x_2, y)F(x_3, y)}_{\Phi(x_1,x_2,x_3,y)}\big]_{x_1,x_2,x_3 \in I,\ y \in J}.$$

After this simplification, the main statement can be formulated as[3]

$$\Box\mathcal{B} = \sum_{\substack{S \subseteq \{1,\ldots,m\},\ T \subseteq \{1,\ldots,n\} \\ |S|,|T| \text{ even},\ (S,T) \neq (\emptyset,\emptyset)}} \Big[\Big\langle \Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\Big\rangle_{\substack{x_i \in I \text{ for } i \in S \\ y_j \in J \text{ for } j \in T}}\Big]_{\substack{x_i \in I \text{ for } i \in S^c \\ y_j \in J \text{ for } j \in T^c}},$$

which can in turn be rewritten (using the definition of $\Box\mathcal{B}$) as

$$\begin{aligned}
&\frac{1}{4}\big[\Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\big]_{x_1,\ldots,x_m \in I_{\text{left}},\ y_1,\ldots,y_n \in J_{\text{left}}} \\
&+ \frac{1}{4}\big[\Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\big]_{x_1,\ldots,x_m \in I_{\text{left}},\ y_1,\ldots,y_n \in J_{\text{right}}} \\
&+ \frac{1}{4}\big[\Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\big]_{x_1,\ldots,x_m \in I_{\text{right}},\ y_1,\ldots,y_n \in J_{\text{left}}} \\
&+ \frac{1}{4}\big[\Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\big]_{x_1,\ldots,x_m \in I_{\text{right}},\ y_1,\ldots,y_n \in J_{\text{right}}} \\
&= \sum_{\substack{S \subseteq \{1,\ldots,m\},\ T \subseteq \{1,\ldots,n\} \\ |S|,|T| \text{ even}}} \Big[\Big\langle \Phi(x_1, \ldots, x_m, y_1, \ldots, y_n)\Big\rangle_{\substack{x_i \in I \text{ for } i \in S \\ y_j \in J \text{ for } j \in T}}\Big]_{\substack{x_i \in I \text{ for } i \in S^c \\ y_j \in J \text{ for } j \in T^c}}.
\end{aligned} \tag{2.10}$$

[3]Here $S^c := \{1, \ldots, m\} \setminus S$ and $T^c := \{1, \ldots, n\} \setminus T$.

For the purpose of the proof, we denote (for a dyadic interval I)

$$\vartheta_I(x) := \begin{cases} 1, & \text{for } x \in I_{\text{left}}, \\ -1, & \text{for } x \in I_{\text{right}}, \\ 0, & \text{for } x \notin I \end{cases}$$

and immediately observe that

$$\left.\begin{aligned} \langle f(x)\rangle_{x\in I} &= [f(x)\vartheta_I(x)]_{x\in I}\,, \\ [f(x)]_{x\in I_{\text{left}}} &= \big[f(x)\big(1+\vartheta_I(x)\big)\big]_{x\in I}, \\ [f(x)]_{x\in I_{\text{right}}} &= \big[f(x)\big(1-\vartheta_I(x)\big)\big]_{x\in I}. \end{aligned}\right\} \tag{2.11}$$

We start from an obvious algebraic identity

$$\begin{aligned} &\Big(\prod_{i=1}^{m}\big(1+\alpha\,\vartheta_I(x_i)\big)\Big)\Big(\prod_{j=1}^{n}\big(1+\beta\,\vartheta_J(y_j)\big)\Big) \\ &= \sum_{\substack{S\subseteq\{1,\ldots,m\}\\ T\subseteq\{1,\ldots,n\}}} \alpha^{|S|}\beta^{|T|}\Big(\prod_{i\in S}\vartheta_I(x_i)\Big)\Big(\prod_{j\in T}\vartheta_J(y_j)\Big) \end{aligned} \tag{2.12}$$

proved simply by multiplying out the product on the left hand side. Choosing four particular values for the parameters α,β:

$$(\alpha,\beta) \in \big\{(1,1),(1,-1),(-1,1),(-1,-1)\big\}\,,$$

we get

$$\begin{aligned} &\frac{1}{4}\Big(\prod_{i=1}^{m}(1+\vartheta_I(x_i))\Big)\Big(\prod_{j=1}^{n}(1+\vartheta_J(y_j))\Big) + \frac{1}{4}\Big(\prod_{i=1}^{m}(1+\vartheta_I(x_i))\Big)\Big(\prod_{j=1}^{n}(1-\vartheta_J(y_j))\Big) \\ &+ \frac{1}{4}\Big(\prod_{i=1}^{m}(1-\vartheta_I(x_i))\Big)\Big(\prod_{j=1}^{n}(1+\vartheta_J(y_j))\Big) + \frac{1}{4}\Big(\prod_{i=1}^{m}(1-\vartheta_I(x_i))\Big)\Big(\prod_{j=1}^{n}(1-\vartheta_J(y_j))\Big) \\ &= \sum_{\substack{S\subseteq\{1,\ldots,m\},\ T\subseteq\{1,\ldots,n\}\\ |S|,|T|\ \text{even}}} \Big(\prod_{i\in S}\vartheta_I(x_i)\Big)\Big(\prod_{j\in T}\vartheta_J(y_j)\Big)\,. \end{aligned}$$

Multiplying this equality by $\Phi(x_1,\ldots,x_m,y_1,\ldots,y_n)$, averaging over $x_1,\ldots,x_m \in I$ and $y_1,\ldots,y_n \in J$, and using (2.11) gives the desired Identity (2.10). $\square$

Let us state a part of the presented proof as a separate lemma. We will need it in Sections 4.2 and 4.4.

Lemma 2.3. *We add up all expressions of the form*

$$\Big(\dots\Big(\Big(\Psi(x_1,x_2,\dots,x_m)\Big)_{x_1\in I}\Big)_{x_2\in I}\cdots\Big)_{x_m\in I}$$

where an even number of parentheses $(\cdot)$ *is replaced with brackets of type* $\langle\cdot\rangle$ *and the remaining parentheses* $(\cdot)$ *are replaced with brackets of type* $[\cdot]$. *(Possible repeating terms are added multiple times.) The resulting sum is equal to*

$$\frac{1}{2}\big[\Psi(x_1,x_2,\dots,x_m)\big]_{x_1,x_2,\dots,x_m\in I_{\text{left}}}+\frac{1}{2}\big[\Psi(x_1,x_2,\dots,x_m)\big]_{x_1,x_2,\dots,x_m\in I_{\text{right}}}.$$

In particular, if $\Psi(x_1,x_2,\dots,x_m)\geq 0$, *then the sum in question will also be nonnegative.*

Proof. We are using the notation from the proof of Theorem 2.2 and Identity 2.12 with $(\alpha,\beta)=(\pm1,0)$. The sum in question is

$$\begin{aligned}
&\sum_{\substack{S\subseteq\{1,\dots,m\}\\|S|\text{ even}}}\Big[\Big\langle\Psi(x_1,x_2,\dots,x_m)\Big\rangle_{x_i\in I\text{ for }i\in S}\Big]_{x_i\in I\text{ for }i\in S^c}\\
&=\sum_{\substack{S\subseteq\{1,\dots,m\}\\|S|\text{ even}}}\Big[\Psi(x_1,x_2,\dots,x_m)\prod_{i\in S}\vartheta_I(x_i)\Big]_{x_1,x_2,\dots,x_m\in I}\\
&=\frac{1}{2}\Big[\Psi(x_1,x_2,\dots,x_m)\prod_{i=1}^{m}\big(1+\vartheta_I(x_i)\big)\Big]_{x_1,x_2,\dots,x_m\in I}\\
&\quad+\frac{1}{2}\Big[\Psi(x_1,x_2,\dots,x_m)\prod_{i=1}^{m}\big(1-\vartheta_I(x_i)\big)\Big]_{x_1,x_2,\dots,x_m\in I}\\
&=\frac{1}{2}\big[\Psi(x_1,x_2,\dots,x_m)\big]_{x_1,x_2,\dots,x_m\in I_{\text{left}}}+\frac{1}{2}\big[\Psi(x_1,x_2,\dots,x_m)\big]_{x_1,x_2,\dots,x_m\in I_{\text{right}}}.
\end{aligned}$$

□

2.3 Two examples

Examples in this section are relevant to the twisted paraproduct mentioned in Chapter 1. We find it convenient to prove the key estimates here and use them as separate results in Section 3.2.

One of the main objects studied in the next chapter can be written in terms of averages as

$$\Lambda_{\mathrm{d}}(F,G,H) = \sum_{I\times J\in\mathcal{C}} |I\times J| \left\langle \big[F(u,y)G(x,v)H(x,y)\big]_{x,u\in I} \right\rangle_{y,v\in J} . \tag{2.13}$$

Proposition 2.4. *Trilinear form Λ_{d} satisfies the estimate*

$$|\Lambda_{\mathrm{d}}(F,G,H)| \ \lesssim\ \|F\|_{\mathrm{L}^4(\mathbb{R}^2)}\|G\|_{\mathrm{L}^2(\mathbb{R}^2)}\|H\|_{\mathrm{L}^4(\mathbb{R}^2)} .$$

Besides this type-$(4,2,4)$ estimate, one could similarly obtain type-$(2,4,4)$ estimate, and then use multilinear interpolation. However, Λ_{d} is actually bounded in a larger range and the latter proof will require additional ideas and will span over most of the next chapter.

Proof of Proposition 2.4. Note that

$$\Lambda_{\mathrm{d}}(F,G,H) = \sum_{I\times J\in\mathcal{C}} |I\times J|\ \mathcal{A}_{I\times J}(F,G,H) ,$$

with $\mathcal{A} = \mathcal{A}_{I\times J}(F,G,H)$ defined in (2.3). For a positive integer N we denote

$$\mathcal{C}_N := \{I\times J\in\mathcal{C} \,:\, I,J\subseteq[-2^N,2^N),\ |I|=|J|\geq 2^{-N}\} .$$

Observe that $\mathcal{C}_N$ consists of 4 finite convex trees, and that $\mathcal{C}_N$ exhaust $\mathcal{C}$ as $N\to\infty$.

This time, let us build the Bellman function $\mathcal{B} = \mathcal{B}_{I\times J}(F,G,H)$ systematically. We are searching for $\mathcal{B}$ such that $\Box\mathcal{B}\geq|\mathcal{A}|$, but also

$$|\mathcal{B}_{I\times J}(F,G,H)| \ \lesssim\ [F(x,y)^4]_{x,y} + [G(x,y)^2]_{x,y} + [H(x,y)^4]_{x,y} . \tag{2.14}$$

We rewrite $\mathcal{A}_{I\times J}(F,G,H)$ as

$$\big[\langle F(u,y)H(x,y)\rangle_y \langle G(x,v)\rangle_v\big]_{x,u}$$

and realize (using $|AB| \le \frac{1}{2}A^2 + \frac{1}{2}B^2$) that it is dominated by

$$\big[\langle F(u,y)H(x,y)\rangle_y^2\big]_{x,u} \qquad \text{and} \qquad \big[\langle G(x,v)\rangle_v^2\big]_x .$$

The second term is simply $\Box\big(\big[[G(x,v)]_v^2\big]_x\big)$, so we "add" $\big[[G(x,v)]_v^2\big]_x$ to our Bellman function. From Theorem 2.2 we know that the first term appears in $\Box\big[[F(u,y)H(x,y)]_y^2\big]_{x,u}$ and we use the same theorem to compute the actual expression, see Table 2.2. Now we see that one has to dominate the remaining two terms in the same table entry:

$$\big\langle [F(u,y)H(x,y)]_y^2 \big\rangle_{x,u} \qquad \text{and} \qquad \big\langle \langle F(u,y)H(x,y)\rangle_y^2 \big\rangle_{x,u} .$$

We can rewrite them respectively as

$$\begin{aligned}
&\big\langle [F(u,y)H(x,y)F(u,v)H(x,v)]_{y,v} \big\rangle_{x,u} \\
&= \big[\langle F(u,y)F(u,v)\rangle_u \langle H(x,y)H(x,v)\rangle_x\big]_{y,v}
\end{aligned}$$

and

$$\begin{aligned}
&\big\langle F(u,y)H(x,y)F(u,v)H(x,v) \big\rangle_{y,v,x,u} \\
&= \big\langle \langle F(u,y)F(u,v)\rangle_u \langle H(x,y)H(x,v)\rangle_x \big\rangle_{y,v}
\end{aligned}$$

Both of these terms are controlled by

$$\big[\langle F(u,y)F(u,v)\rangle_u^2\big]_{y,v} \qquad \text{and} \qquad \big[\langle H(x,y)H(x,v)\rangle_x^2\big]_{y,v} .$$

The first of the above two terms appears in $\Box\big[[F(u,y)F(u,v)]_u^2\big]_{y,v}$, so Theorem 2.2 applies again and provides the answer in Table 2.2. (The term with H is completely analogous.) The procedure stops at this point, because the remaining terms in the same table entry add up to a nonnegative expression by Lemma 2.3.

$$
\begin{aligned}
&\Big\langle [F(u,y)F(u,v)]_u^2 \Big\rangle_{y,v} + \Big\langle \langle F(u,y)F(u,v)\rangle_u^2 \Big\rangle_{y,v} \\
&= \Big\langle \big[F(u,y)F(u,v)F(x,y)F(x,v) \big]_{x,u} \Big\rangle_{y,v} + \Big\langle F(u,y)F(u,v)F(x,y)F(x,v) \Big\rangle_{x,u,y,v} \\
&= \big[\langle F(u,y)F(x,y)\rangle_y^2 \big]_{x,u} + \Big\langle \langle F(u,y)F(x,y)\rangle_y^2 \Big\rangle_{x,u} \\
&= \frac{1}{2}\big[\langle F(u,y)F(x,y)\rangle_{y\in J}^2 \big]_{x,u\in I_{\mathrm{left}}} + \frac{1}{2}\big[\langle F(u,y)F(x,y)\rangle_{y\in J}^2 \big]_{x,u\in I_{\mathrm{right}}} \geq 0
\end{aligned}
$$

The above reasoning can be graphically represented in the form of a tree diagram, see Figure 2.3.

$$
\begin{tikzcd}[column sep=small]
& \Big\langle \big[F(u,y)G(x,v)H(x,y) \big]_{x,u} \Big\rangle_{y,v} \arrow[d, equal] & \\
& \big[\langle F(u,y)H(x,y)\rangle_y \langle G(x,v)\rangle_v \big]_{x,u} \arrow[dl] \arrow[dr] & \\
\big[\langle F(u,y)H(x,y)\rangle_y^2 \big]_{x,u} \arrow[d, dashed] \arrow[drr, dashed] & & \big[\langle G(x,v)\rangle_v^2 \big]_x \\
\Big\langle [F(u,y)H(x,y)]_y^2 \Big\rangle_{x,u} \arrow[d, equal] & & \Big\langle \langle F(u,y)H(x,y)\rangle_y^2 \Big\rangle_{x,u} \arrow[d, equal] \\
\big[\langle F(u,y)F(u,v)\rangle_u \langle H(x,y)H(x,v)\rangle_x \big]_{y,v} \arrow[d] \arrow[drr] & & \Big\langle \langle F(u,y)F(u,v)\rangle_u \langle H(x,y)H(x,v)\rangle_x \Big\rangle_{y,v} \arrow[d] \arrow[dll] \\
\big[\langle F(u,y)F(u,v)\rangle_u^2 \big]_{y,v} & & \big[\langle H(x,y)H(x,v)\rangle_x^2 \big]_{y,v}
\end{tikzcd}
$$

Figure 2.3: Bellman function tree for type-$(4,2,4)$ estimate for Λ_{d}.

If one decides to choose the coefficients carefully, the resulting Bellman function is

$$\begin{aligned}\mathcal{B} := \frac{1}{2}\big[[G(x,v)]_v^2\big]_x + \frac{1}{2}\big[[F(u,y)H(x,y)]_y^2\big]_{x,u} \\ + \frac{1}{2}\big[[F(u,y)F(u,v)]_u^2\big]_{y,v} + \frac{1}{2}\big[[H(x,y)H(x,v)]_x^2\big]_{y,v}\end{aligned} \tag{2.15}$$

and its first order difference is

$$\begin{aligned}\Box\mathcal{B} = {} & \frac{1}{2}\big[\langle G(x,v)\rangle_v^2\big]_x + \frac{1}{2}\big[\langle F(u,y)H(x,y)\rangle_y^2\big]_{x,u} \\ & + \frac{1}{8}\Big[\big\langle F(x,y)F(x,v) + H(x,y)H(x,v)\big\rangle_x^2\Big]_{y,v\in J_{\mathrm{left}}} \\ & + \frac{1}{8}\Big[\big\langle F(x,y)F(x,v) + H(x,y)H(x,v)\big\rangle_x^2\Big]_{y,v\in J_{\mathrm{right}}} \\ & + \frac{1}{4}\Big[\langle F(x,y)F(x,v)\rangle_{x\in I}^2 + \langle H(x,y)H(x,v)\rangle_{x\in I}^2\Big]_{y\in J_{\mathrm{left}},\, v\in J_{\mathrm{right}}} \\ & + \frac{1}{4}\big[\langle F(u,y)F(x,y)\rangle_{y\in J}^2\big]_{x,u\in I_{\mathrm{left}}} + \frac{1}{4}\big[\langle F(u,y)F(x,y)\rangle_{y\in J}^2\big]_{x,u\in I_{\mathrm{right}}} \\ & + \frac{1}{4}\big[\langle H(u,y)H(x,y)\rangle_{y\in J}^2\big]_{x,u\in I_{\mathrm{left}}} + \frac{1}{4}\big[\langle H(u,y)H(x,y)\rangle_{y\in J}^2\big]_{x,u\in I_{\mathrm{right}}} \\ \geq {} & \Big|\big[\langle G(x,v)\rangle_v \langle F(u,y)H(x,y)\rangle_y\big]_{x,u}\Big| \;=\; |\mathcal{A}|\,.\end{aligned}$$

Also note that $\mathcal{B}$ clearly satisfies (2.14), for instance

$$\begin{aligned}\big[[F(u,y)H(x,y)]_y^2\big]_{x,u} & \leq \big[[F(u,y)^2H(x,y)^2]_y\big]_{x,u} \\ & \leq \frac{1}{2}\big[F(u,y)^4\big]_{u,y} + \frac{1}{2}\big[H(x,y)^4\big]_{x,y}\,.\end{aligned}$$

Applying the general Inequality (2.8) to the four trees that make $\mathcal{C}_N$ and then using (2.14) gives

$$\begin{aligned}\sum_{I\times J\in\mathcal{C}_N} |I\times J|\ |\mathcal{A}_{I\times J}(F,G,H)| & \leq \sum_{\substack{I,J\subseteq[0,2^N) \\ |I|=|J|=2^{-N-1}}} |I\times J|\ \ \mathcal{B}_{I\times J}(F,G,H) \\ & \lesssim \int_{-2^N}^{2^N}\!\!\int_{-2^N}^{2^N}\big(F(x,y)^4 + G(x,y)^2 + H(x,y)^4\big)\,dx\,dy\,.\end{aligned}$$

Since N was arbitrary, we have established

$$|\Lambda_{\mathrm{d}}(F,G,H)| \ \lesssim\ \|F\|_{\mathrm{L}^4(\mathbb{R}^2)}^4 + \|G\|_{\mathrm{L}^2(\mathbb{R}^2)}^2 + \|H\|_{\mathrm{L}^4(\mathbb{R}^2)}^4\,.$$

To prove the desired inequality, it remains to use the "homogeneity trick", i.e. replace F, G, H respectively with $\frac{F}{\|F\|_{\mathrm{L}^4}}, \frac{G}{\|G\|_{\mathrm{L}^2}}, \frac{H}{\|H\|_{\mathrm{L}^4}}$. $\square$

Before we start the next example, let us introduce one notion from additive combinatorics, primarily for notational convenience, but also to emphasize the (somewhat surprising) connection. For any dyadic square $Q = I \times J$ we first define the *Gowers box inner-product* of four functions F_1, F_2, F_3, F_4 as

$$\begin{aligned}[F_1, F_2, F_3, F_4]_{\square(Q)} &:= \frac{1}{|Q|^2} \int_I \int_I \int_J \int_J F_1(u,v) F_2(x,v) F_3(u,y) F_4(x,y) du dx dv dy \\ &= \big[F_1(u,v) F_2(x,v) F_3(u,y) F_4(x,y)\big]_{u,x \in I,\ v,y \in J}.\end{aligned}$$

Then for any function F we introduce the two-dimensional *Gowers box norm* as[4]

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

It plays an important role in a paper by Shkredov [44], while its appearance in an expository paper by Tao [47] is the one we find most inspiring in this context. Gowers box norms (even higher-dimensional ones) arise from the work of Gowers [16], [17] and are more systematically studied and applied in [18] and by Green and Tao in [20]. All these norms can be viewed as particular instances of averaging paraproduct-type terms, this time for functions in $\mathbb{R}^N$, containing precisely two variables in each dimension.[5]

It is easy to prove the *box Cauchy-Schwarz inequality*, as stated in [47], [18], or [20]:

$$[F_1, F_2, F_3, F_4]_{\square(Q)} \leq \|F_1\|_{\square(Q)} \|F_2\|_{\square(Q)} \|F_3\|_{\square(Q)} \|F_4\|_{\square(Q)}. \tag{2.16}$$

[4]We borrow a comment from [47]: If $F(x,y)$ restricted to Q is discretized and viewed as a matrix, then $\|F\|_{\square(Q)}$ can be recognized as its (properly normalized) Schatten 4-norm, i.e. ℓ^4 norm of the sequence of its singular values. This comment gives yet one more immediate proof of Inequality (2.17) below.

[5]The author would like to thank Professor Terence Tao for pointing out the analogy between averaging paraproduct-type terms and higher-dimensional Gowers box norms.

To see (2.16), one has to write $[F_1, F_2, F_3, F_4]_{\Box(Q)}$ as

$$\Big[[F_1(u,v)F_2(x,v)]_v[F_3(u,y)F_4(x,y)]_y\Big]_{u,x},$$

and apply the ordinary Cauchy-Schwarz inequality in $(u,x) \in I \times I$. Then one rewrites the result as

$$\Big[[F_1(u,v)F_1(u,y)]_u[F_2(x,v)F_2(x,y)]_x\Big]_{v,y}^{\frac{1}{2}}\Big[[F_3(u,v)F_3(u,y)]_u[F_4(x,v)F_4(x,y)]_x\Big]_{v,y}^{\frac{1}{2}}$$

and applies the Cauchy-Schwarz inequality again, this time in $(v,y) \in J \times J$. From here it is also easily seen that $\|\cdot\|_{\Box(Q)}$ is really a norm on functions supported on Q. On the other hand, a straightforward application of the (ordinary) Cauchy-Schwarz inequality yields

$$\|F\|_{\Box(Q)} \leq \Big(\frac{1}{|Q|}\int_Q F(x,y)^2\, dxdy\Big)^{1/2} = \big[F(x,y)^2\big]_{x\in I,\, y\in J}^{1/2}. \tag{2.17}$$

An alternative way to verify (2.17) is to notice that it is a special case of the strong $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ estimate for the quadrilinear form

$$(F_1, F_2, F_3, F_4) \mapsto |Q|^2[F_1, F_2, F_3, F_4]_{\Box(Q)}.$$

Since $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ is in the convex hull of $(1,0,0,1)$ and $(0,1,1,0)$, we can use complex interpolation, and it is enough to verify strong type estimates for the latter points, which is trivial.

Note that we have already encountered expressions of that type. For example, the right hand side of (2.15) can now be written as

$$\frac{1}{2}[\mathbf{1}, G, \mathbf{1}, G]_{\Box(Q)} + \frac{1}{2}[F, H, F, H]_{\Box(Q)} + \frac{1}{2}[F, F, F, F]_{\Box(Q)} + \frac{1}{2}[H, H, H, H]_{\Box(Q)},$$

where $\mathbf{1}$ is the constant function.

Let us now provide another elegant example of our Bellman function machinery. For any finite convex tree $\mathcal{T}$ we introduce a quadrilinear form

$$\Theta_{\mathcal{T}}(F_1, F_2, F_3, F_4) := \sum_{I\times J\in\mathcal{T}} |I \times J|\ \Big\langle\big[F_1(u,v)F_2(x,v)F_3(u,y)F_4(x,y)\big]_{u,x\in I}\Big\rangle_{v,y\in J}.$$

In the next chapter, we will need the following inequality.

Proposition 2.5. *Quadrilinear form $\Theta_{\mathcal{T}}$ satisfies the estimate*

$$\big|\Theta_{\mathcal{T}}(F_1, F_2, F_3, F_4)\big| \;\lesssim\; |Q_{\mathcal{T}}| \prod_{j=1}^{4} \max_{Q\in\mathcal{L}(\mathcal{T})} \|F_j\|_{\Box(Q)} \,.$$

Proof of Proposition 2.5. Denote

$$\mathcal{A} = \mathcal{A}_{I\times J}(F_1, F_2, F_3, F_4) := \Big\langle \big[F_1(u,v)F_2(x,v)F_3(u,y)F_4(x,y)\big]_{u,x\in I} \Big\rangle_{v,y\in J}$$

and define the Bellman function $\mathcal{B} = \mathcal{B}_Q(F_1, F_2, F_3, F_4)$ by the formula

$$\mathcal{B} :=\; \frac{1}{2}[F_1, F_2, F_1, F_2]_{\Box(Q)} + \frac{1}{2}[F_3, F_4, F_3, F_4]_{\Box(Q)} + \frac{1}{2}\sum_{j=1}^{4}[F_j, F_j, F_j, F_j]_{\Box(Q)} \,.$$

This choice can be justified similarly as in the previous example and we summarize the reasoning in the form of another tree diagram, see Figure 2.4, where we omit writing treatment of the right (analogous) branch.

$$\begin{tikzcd}
& \big[\langle F_1(u,v)F_2(x,v)\rangle_v \langle F_3(u,y)F_4(x,y)\rangle_y\big]_{u,x} \arrow[dl] \arrow[dr] & \\
\big[\langle F_1(u,y)F_2(x,y)\rangle_y^2\big]_{u,x} \arrow[d, dashed] \arrow[drr, dashed] & & \big[\langle F_3(u,y)F_4(x,y)\rangle_y^2\big]_{u,x} \\
\big\langle [F_1(u,y)F_2(x,y)]_y^2 \big\rangle_{u,x} \arrow[d] \arrow[drr] & & \big\langle \langle F_1(u,y)F_2(x,y)\rangle_y^2 \big\rangle_{u,x} \arrow[d] \arrow[dll] \\
\big[\langle F_1(x,v)F_1(x,y)\rangle_x^2\big]_{v,y} & & \big[\langle F_2(x,v)F_2(x,y)\rangle_x^2\big]_{v,y}
\end{tikzcd}$$

Figure 2.4: Bellman function tree for $\Theta_{\mathcal{T}}$.

Using Theorem 2.2 or Table 2.2 and rearranging the terms with the help of (2.6), we get

$$
\begin{aligned}
\square\mathcal{B} = {} & \frac{1}{2}\big[\langle F_1(u,y)F_2(x,y)\rangle_y^2\big]_{x,u} + \frac{1}{2}\big[\langle F_3(u,y)F_4(x,y)\rangle_y^2\big]_{x,u} \\
& + \frac{1}{8}\Big[\Big\langle F_1(x,y)F_1(x,v) + F_2(x,y)F_2(x,v)\Big\rangle_x^2\Big]_{y,v\in J_{\text{left}}} \\
& + \frac{1}{8}\Big[\Big\langle F_1(x,y)F_1(x,v) + F_2(x,y)F_2(x,v)\Big\rangle_x^2\Big]_{y,v\in J_{\text{right}}} \\
& + \frac{1}{8}\Big[\Big\langle F_3(x,y)F_3(x,v) + F_4(x,y)F_4(x,v)\Big\rangle_x^2\Big]_{y,v\in J_{\text{left}}} \\
& + \frac{1}{8}\Big[\Big\langle F_3(x,y)F_3(x,v) + F_4(x,y)F_4(x,v)\Big\rangle_x^2\Big]_{y,v\in J_{\text{right}}} \\
& + \frac{1}{4}\sum_{j=1}^{4}\big[\langle F_j(x,y)F_j(x,v)\rangle_{x\in I}^2\big]_{y\in J_{\text{left}},\, v\in J_{\text{right}}} \\
& + \frac{1}{4}\sum_{j=1}^{4}\Big(\big[\langle F_j(u,y)F_j(x,y)\rangle_{y\in J}^2\big]_{x,u\in I_{\text{left}}} + \big[\langle F(u,y)F(x,y)\rangle_{y\in J}^2\big]_{x,u\in I_{\text{right}}}\Big)\,.
\end{aligned}
$$

All rows are nonnegative, so

$$\square\mathcal{B} \geq \Big|\big[\langle F_1(u,v)F_2(x,v)\rangle_v \langle F_3(u,y)F_4(x,y)\rangle_y\big]_{u,x}\Big| = |\mathcal{A}|$$

and an immediate consequence of (2.16) is

$$\mathcal{B}_Q(F_1,F_2,F_3,F_4) \;\lesssim\; \sum_{j=1}^{4}\|F_j\|_{\square(Q)}^4\,. \tag{2.18}$$

The main Inequality (2.8) combined with (2.18) and the fact that $\mathcal{L}(\mathcal{T})$ partitions $Q_{\mathcal{T}}$ gives

$$\big|\Theta_{\mathcal{T}}(F_1,F_2,F_3,F_4)\big| \;\lesssim\; |Q_{\mathcal{T}}|\sum_{j=1}^{4}\max_{Q\in\mathcal{L}(\mathcal{T})}\|F_j\|_{\square(Q)}^4\,.$$

The same homogeneity trick as before completes the proof. $\square$

CHAPTER 3

Boundedness of the twisted paraproduct

3.1 Formulation of the result

Let us denote dyadic martingale averages and differences by

$$\mathbb{E}_k f := \sum_{|I|=2^{-k}} \big(\tfrac{1}{|I|}\textstyle\int_I f\big)\, \mathbf{1}_I\,, \qquad \Delta_k f := \mathbb{E}_{k+1} f - \mathbb{E}_k f\,,$$

for every $k \in \mathbb{Z}$, where the sum is taken over dyadic intervals $I \subseteq \mathbb{R}$ of length 2^{-k}. When we apply an operator in only one variable of a two-dimensional function, we mark it with that variable in the superscript. For instance,

$$(\mathbb{E}_k^{(1)} F)(x,y) := \big(\mathbb{E}_k F(\cdot, y)\big)(x)\,.$$

The *dyadic twisted paraproduct* is defined as

$$T_{\mathrm{d}}(F,G) := \sum_{k\in\mathbb{Z}} (\mathbb{E}_k^{(1)} F)(\Delta_k^{(2)} G)\,. \tag{3.1}$$

In the continuous case, let P_φ denote the Fourier multiplier with symbol $\hat{\varphi}$, i.e.

$$\mathrm{P}_\varphi f := f * \varphi\,.$$

Take two functions $\varphi, \psi \in \mathrm{C}^1(\mathbb{R})$ satisfying

$$|\varphi(x)|, |\tfrac{d}{dx}\varphi(x)|, |\psi(x)|, |\tfrac{d}{dx}\psi(x)| \ \lesssim\ (1+|x|)^{-3}, \tag{3.2}$$

and

$$\mathrm{supp}(\hat{\psi}) \subseteq \{\xi \in \mathbb{R} \,:\, \tfrac{1}{2} \le |\xi| \le 2\}\,.$$

For every $k \in \mathbb{Z}$ denote $\varphi_k(x) := 2^k \varphi(2^k x)$ and $\psi_k(x) := 2^k \psi(2^k x)$. The associated *continuous twisted paraproduct* is defined as

$$T_{\mathrm{c}}(F,G) := \sum_{k\in\mathbb{Z}} (\mathrm{P}^{(1)}_{\varphi_k} F)(\mathrm{P}^{(2)}_{\psi_k} G)\,. \tag{3.3}$$

We are interested in strong-type estimates

$$\|T(F,G)\|_{\mathrm{L}^{pq/(p+q)}(\mathbb{R}^2)} \lesssim_{p,q} \|F\|_{\mathrm{L}^p(\mathbb{R}^2)} \|G\|_{\mathrm{L}^q(\mathbb{R}^2)}\,, \tag{3.4}$$

and weak-type estimates

$$\alpha\, \Big|\Big\{(x,y)\in\mathbb{R}^2 : \, |T(F,G)(x,y)| > \alpha\Big\}\Big|^{(p+q)/pq} \lesssim_{p,q} \|F\|_{\mathrm{L}^p(\mathbb{R}^2)} \|G\|_{\mathrm{L}^q(\mathbb{R}^2)} \tag{3.5}$$

for (3.1) and (3.3). The exponent $\frac{pq}{p+q}$ is mandated by scaling invariance. When $p=\infty$ or $q=\infty$, we interpret it as q or p respectively.

The main result of the chapter establishes (3.4) and (3.5) in certain ranges of (p,q).

Theorem 3.1. (a) *Operators T_{d} and T_{c} satisfy the strong bound* (3.4) *if*

$$1 < p,q < \infty, \quad \tfrac{1}{p} + \tfrac{1}{q} > \tfrac{1}{2}\,.$$

(b) *Additionally, operators T_{d} and T_{c} satisfy the weak bound* (3.5) *when*

$$p = 1,\ 1 \le q < \infty \quad \textit{or} \quad q = 1,\ 1 \le p < \infty\,.$$

(c) *The weak estimate* (3.5) *fails for $p=\infty$ or $q=\infty$.*

The name *twisted paraproduct* is indicative because there is a "twist" in the variables in which the convolutions (or the martingale projections) are performed, as opposed to the case of the ordinary paraproduct. No bounds on (3.1) or (3.3) were known prior to this work. A conditional result was shown by Bernicot in

[1], assuming boundedness in some range and expanding the range towards lower exponents using a fiber-wise Calderón-Zygmund decomposition. We repeat his argument in the dyadic setting in Section 3.4, for the purpose of extending the boundedness region established in Section 3.3.

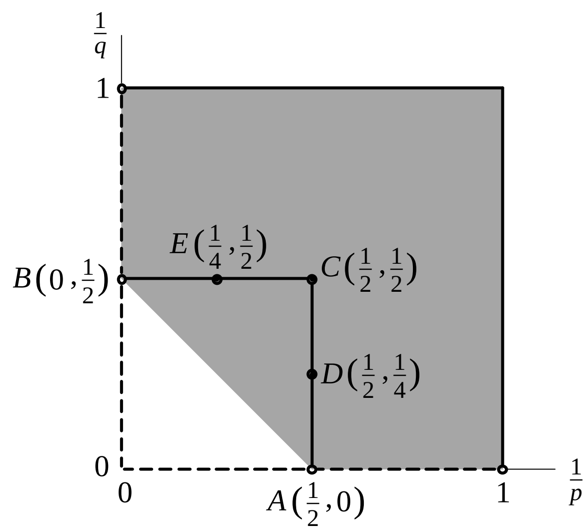

Figure 3.1: The range of exponents for the twisted paraproduct operator.

Figure 3.1 depicts the range of exponents in Theorem 3.1. The shaded region satisfies the strong estimate, while for two solid sides of the unit square we only establish the weak estimates. The two dashed sides of the square represent exponents for which we show that even the weak estimate fails. The white triangle in the lower left corner is the region we do not discuss.

The proof of Theorem 3.1 is organized as follows. Section 3.3 proves estimates for T_{d} in the interior of triangle ABC. In Section 3.4 the rest of bounds for T_{d} are obtained. Section 3.5 establishes bounds for T_{c} by relating T_{c} to T_{d}. Finally, in Section 3.6 we discuss the counterexamples. Recall that in Section 2.3 we gave a simpler proof for points D and E only.

Before going into the proofs, we make several simple observations about $T_{\rm d}$. Note that Theorem 3.1 also gives estimates for a family of shifted operators

$$(F,G) \mapsto \sum_{k\in\mathbb{Z}} (\mathbb{E}^{(1)}_{k+k_0}F)(\Delta^{(2)}_{k}G)$$

uniformly in $k_0 \in \mathbb{Z}$, because the last sum can be rewritten as

$$\mathrm{D}_{(2^{-k_0},1)}\, T_{\rm d}\big(\mathrm{D}_{(2^{k_0},1)}F,\, \mathrm{D}_{(2^{k_0},1)}G\big)\,.$$

Here $\mathrm{D}_{(a,1)}$ denotes the non-isotropic dilation $(\mathrm{D}_{(a,1)}F)(x,y) := F(a^{-1}x,y)$.

If F and G are compactly supported, then one can write

$$T_{\rm d}(F,G) = FG - \sum_{k\in\mathbb{Z}} (\Delta^{(1)}_{k}F)(\mathbb{E}^{(2)}_{k+1}G)\,. \tag{3.6}$$

Combining this with the previous remark and the fact that the pointwise product FG satisfies Hölder's inequality, we see that the set of estimates for $T_{\rm d}(F,G)$ is indeed symmetric under interchanging p and q, F and G. We use this fact to shorten some of the exposition below.

Furthermore, Theorem 3.1 implies bounds on more general dyadic operators of the following type:

$$\Big\| \sum_{k\in\mathbb{Z}} c_k (\mathbb{E}^{(1)}_{k}F)(\Delta^{(2)}_{k}G) \Big\|_{\mathrm{L}^{pq/(p+q)}} \lesssim_{p,q} \|F\|_{\mathrm{L}^p} \|G\|_{\mathrm{L}^q}\,, \tag{3.7}$$

for any numbers c_k such that $|c_k| \leq 1$. Here we restrict ourselves to the interior range $1<p,q<\infty$, $\frac{1}{p}+\frac{1}{q}>\frac{1}{2}$. One simply uses the known bound for $T_{\rm d}(F,\widetilde{G})$ with $\widetilde{G} := \sum_{k\in\mathbb{Z}} c_k\, \Delta^{(2)}_{k}G$, and the dyadic Littlewood-Paley inequality in the second variable. Note that the flexibility of having coefficients c_k is implicit in the definition of $T_{\rm c}$, and indeed we will repeat a continuous variant of this argument in Section 3.5.

3.2 A single tree estimate

This short section serves only to make a connection with the theory developed in Chapter 2, and to retrieve the key result from Section 2.3.

For any dyadic interval I, denote the Haar scaling function $\varphi_I^{\mathrm{d}} := |I|^{-1/2}\mathbf{1}_I$ and the Haar wavelet $\psi_I^{\mathrm{d}} := |I|^{-1/2}(\mathbf{1}_{I_{\mathrm{left}}} - \mathbf{1}_{I_{\mathrm{right}}})$. Martingale averages and differences can be alternatively written in the Haar basis:

$$
\begin{aligned}
(\mathbb{E}_k f)(x) &= \sum_{|I|=2^{-k}} \big(\textstyle\int_{\mathbb{R}} f(u)\varphi_I^{\mathrm{d}}(u)du\big)\, \varphi_I^{\mathrm{d}}(x)\,, \\
(\Delta_k f)(x) &= \sum_{|I|=2^{-k}} \big(\textstyle\int_{\mathbb{R}} f(u)\psi_I^{\mathrm{d}}(u)du\big)\, \psi_I^{\mathrm{d}}(x)\,.
\end{aligned}
$$

Consequently, T_{d} can be rewritten as the sum over the collection of all dyadic squares:

$$
T_{\mathrm{d}}(F,G)(x,y) = \sum_{I\times J\in\mathcal{C}} \int_{\mathbb{R}^2} F(u,y)G(x,v)\ \varphi_I^{\mathrm{d}}(u)\varphi_I^{\mathrm{d}}(x)\psi_J^{\mathrm{d}}(v)\psi_J^{\mathrm{d}}(y)\ du\,dv\,.
$$

It will be more convenient to work with the dualized trilinear form

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

and notice that it can be written in terms of averages introduced in the previous chapter:

$$
\begin{aligned}
\Lambda_{\mathrm{d}}(F,G,H) &= \sum_{I\times J\in\mathcal{C}} \int_{\mathbb{R}^4} F(u,y)G(x,v)H(x,y)\, \varphi_I^{\mathrm{d}}(u)\varphi_I^{\mathrm{d}}(x)\psi_J^{\mathrm{d}}(v)\psi_J^{\mathrm{d}}(y)\, dudxdvdy \\
&= \sum_{I\times J\in\mathcal{C}} |I\times J| \, \big\langle \big[F(u,y)G(x,v)H(x,y)\big]_{u,x\in I} \big\rangle_{v,y\in J}\,.
\end{aligned}
$$

We have justified (2.13) and thus Proposition 2.4 establishes the estimate for T_{d} corresponding to point E in Figure 3.1. Point D is then handled by symmetry following from (3.6).

For the purpose of proving estimates in a larger range we will have to work locally, rather than immediately trying to bound expressions over all dyadic squares. For any finite convex tree $\mathcal{T}$ we introduce a local version of the trilinear form:

$$\Lambda_{\mathcal{T}}(F,G,H) := \sum_{I\times J\in\mathcal{T}} \int_{\mathbb{R}^4} F(u,y)G(x,v)H(x,y)\,\varphi_I^{\mathrm{d}}(u)\varphi_I^{\mathrm{d}}(x)\psi_J^{\mathrm{d}}(v)\psi_J^{\mathrm{d}}(y)\,dudxdvdy$$

and a more symmetric quadrilinear form:

$$\begin{aligned}\Theta_{\mathcal{T}}(F_1,F_2,F_3,F_4) := \sum_{I\times J\in\mathcal{T}} \int_{\mathbb{R}^4} & F_1(u,v)F_2(x,v)F_3(u,y)F_4(x,y) \\ & \varphi_I^{\mathrm{d}}(u)\varphi_I^{\mathrm{d}}(x)\psi_J^{\mathrm{d}}(v)\psi_J^{\mathrm{d}}(y)\,dudxdvdy \\ = \sum_{I\times J\in\mathcal{T}} |I\times J| \, & \big\langle \big[F_1(u,v)F_2(x,v)F_3(u,y)F_4(x,y)\big]_{u,x\in I}\big\rangle_{v,y\in J}\,,\end{aligned}$$

which was already mentioned in Section 2.3. Observe that $\Lambda_{\mathcal{T}}(F,G,H)$ can be recognized as $\Theta_{\mathcal{T}}(\mathbf{1},G,F,H)$, where $\mathbf{1}$ is the unit constant function on $\mathbb{R}^2$.

Here is a key local estimate, which will be "integrated" to a global one in the next section.

Proposition 3.2 (Single tree estimate)**.** *For any finite convex tree $\mathcal{T}$ we have*

$$\big|\Lambda_{\mathcal{T}}(F,G,H)\big| \lesssim |Q_{\mathcal{T}}| \big(\max_{Q\in\mathcal{L}(\mathcal{T})}\|F\|_{\square(Q)}\big)\big(\max_{Q\in\mathcal{L}(\mathcal{T})}\|G\|_{\square(Q)}\big)\big(\max_{Q\in\mathcal{L}(\mathcal{T})}\|H\|_{\square(Q)}\big)\,.$$

It is an immediate consequence of Proposition 2.5 with $F_1=\mathbf{1}$, $F_2=G$, $F_3=F$, $F_4=H$.

3.3 Proving the estimate in the local L^2 case

In this section we show the bound

$$|\Lambda_{\mathrm{d}}(F,G,H)| \lesssim_{p,q,r} \|F\|_{\mathrm{L}^p}\|G\|_{\mathrm{L}^q}\|H\|_{\mathrm{L}^r} \tag{3.8}$$

for $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=1$, $2<p,q,r<\infty$. By duality we get (3.4) for T_{d} in the range $2<p,q<\infty$, $\frac{1}{p}+\frac{1}{q}>\frac{1}{2}$. The following material became somewhat standard over the time, and indeed we are closely following the ideas from [49], actually in a much simpler setting.

Let us fix bounded, compactly supported, measurable functions $F, G, H \geq 0$, none of them being identically 0. To make all arguments finite, in this section we restrict ourselves to considering only dyadic squares Q satisfying $|Q| \geq 2^{-2N}$ for some (large) fixed positive integer N. Since our bounds will be independent of N, letting $N \to \infty$ handles the whole collection $\mathcal{C}$.

We organize the family of dyadic squares in the following way. For any $k \in \mathbb{Z}$ we define the collection

$$\mathcal{P}_k^F := \Big\{ Q \ : \ 2^k \leq \sup_{Q' \supseteq Q} \|F\|_{\square(Q')} < 2^{k+1} \Big\} ,$$

and let $\mathcal{M}_k^F$ denote the family of maximal squares in $\mathcal{P}_k^F$ with respect to the set inclusion. Collections $\mathcal{P}_k^G$, $\mathcal{M}_k^G$, $\mathcal{P}_k^H$, $\mathcal{M}_k^H$ are defined analogously. Furthermore, for any triple of integers k_1, k_2, k_3 we set

$$\mathcal{P}_{k_1,k_2,k_3} := \mathcal{P}_{k_1}^F \cap \mathcal{P}_{k_2}^G \cap \mathcal{P}_{k_3}^H$$

and let $\mathcal{M}_{k_1,k_2,k_3}$ denote the family of maximal squares in $\mathcal{P}_{k_1,k_2,k_3}$.

For each $Q \in \mathcal{M}_{k_1,k_2,k_3}$ note that

$$\mathcal{T}_Q := \big\{ \widetilde{Q} \in \mathcal{P}_{k_1,k_2,k_3} \ : \ \widetilde{Q} \subseteq Q \big\}$$

is a convex tree with root Q and that for different Q the corresponding trees $\mathcal{T}_Q$ occupy disjoint regions in $\mathbb{R}^2$. These trees decompose the collection $\mathcal{P}_{k_1,k_2,k_3}$, for each individual choice of k_1, k_2, k_3.

We apply Proposition 3.2 to each of the trees $\mathcal{T}_Q$. Consider any leaf $\widetilde{Q} \in \mathcal{L}(\mathcal{T}_Q)$, and denote its parent by Q'. From $Q' \in \mathcal{T}_Q \subseteq \mathcal{P}_{k_1,k_2,k_3}$ we get

$$\tfrac{1}{2}\|F\|_{\square(\widetilde{Q})} \leq \|F\|_{\square(Q')} < 2^{k_1+1} ,$$

thus $\|F\|_{\Box(\widetilde{Q})} \lesssim 2^{k_1}$, and similarly $\|G\|_{\Box(\widetilde{Q})} \lesssim 2^{k_2}$, $\|H\|_{\Box(\widetilde{Q})} \lesssim 2^{k_3}$, so "single tree estimate" implies

$$\big|\Lambda_{\mathcal{T}_Q}(F,G,H)\big| \;\lesssim\; 2^{k_1+k_2+k_3}\,|Q|\,.$$

We split Λ_{d} into a sum of $\Lambda_{\mathcal{T}_Q}$ over all $k_1,k_2,k_3\in\mathbb{Z}$ and all $Q\in\mathcal{M}_{k_1,k_2,k_3}$. In order to finish the proof of (3.8), it remains to show

$$\sum_{k_1,k_2,k_3\in\mathbb{Z}} 2^{k_1+k_2+k_3} \sum_{Q\in\mathcal{M}_{k_1,k_2,k_3}} |Q| \;\lesssim_{p,q,r}\; \|F\|_{\mathrm{L}^p}\|G\|_{\mathrm{L}^q}\|H\|_{\mathrm{L}^r}\,. \tag{3.9}$$

The trick from [49] is to observe that for any fixed triple $k_1,k_2,k_3\in\mathbb{Z}$, squares in $\mathcal{M}^F_{k_1}$ cover squares in $\mathcal{M}_{k_1,k_2,k_3}$, and the latter are disjoint. The same is true for $\mathcal{M}^G_{k_2}$ and $\mathcal{M}^H_{k_3}$. Thus, it suffices to prove

$$\sum_{k_1,k_2,k_3\in\mathbb{Z}} 2^{k_1+k_2+k_3}\min\Big(\sum_{Q\in\mathcal{M}^F_{k_1}}|Q|,\sum_{Q\in\mathcal{M}^G_{k_2}}|Q|,\sum_{Q\in\mathcal{M}^H_{k_3}}|Q|\Big) \\ \lesssim_{p,q,r}\; \|F\|_{\mathrm{L}^p}\|G\|_{\mathrm{L}^q}\|H\|_{\mathrm{L}^r}\,. \tag{3.10}$$

Consider the following version of the dyadic maximal function

$$\mathrm{M}_2 F \;:=\; \sup_{Q\in\mathcal{C}}\Big(\frac{1}{|Q|}\int_Q |F|^2\Big)^{1/2}\mathbf{1}_Q\,.$$

For each $Q\in\mathcal{M}^F_k$, from (2.17) and $\|F\|_{\Box(Q)}\geq 2^k$ we have $Q\subseteq\{\mathrm{M}_2F\geq 2^k\}$, so by disjointness

$$\sum_{Q\in\mathcal{M}^F_k}|Q| \;\leq\; |\{\mathrm{M}_2F\geq 2^k\}|\,.$$

Also note that

$$\sum_{k\in\mathbb{Z}} 2^{pk}|\{\mathrm{M}_2F\geq 2^k\}| \;\sim_p\; \|\mathrm{M}_2F\|^p_{\mathrm{L}^p} \;\lesssim_p\; \|F\|^p_{\mathrm{L}^p}\,,$$

because M_2 is bounded on $\mathrm{L}^p(\mathbb{R}^2)$ for $2<p<\infty$. Therefore

$$\sum_{k\in\mathbb{Z}} 2^{pk}\sum_{Q\in\mathcal{M}^F_k}|Q| \;\lesssim_p\; \|F\|^p_{\mathrm{L}^p}\,, \tag{3.11}$$

and completely analogously we get

$$\sum_{k\in\mathbb{Z}} 2^{qk} \sum_{Q\in\mathcal{M}_k^G} |Q| \ \lesssim_q \|G\|_{\mathrm{L}^q}^q\,, \quad \sum_{k\in\mathbb{Z}} 2^{rk} \sum_{Q\in\mathcal{M}_k^H} |Q| \ \lesssim_r \|H\|_{\mathrm{L}^r}^r\,.$$

A purely algebraic "integration lemma" stated and proved in [49] deduces (3.10) from these three estimates. The idea is to split the sum in (3.10) into three parts, depending on which of the numbers

$$\frac{2^{pk_1}}{\|F\|_{\mathrm{L}^p}^p},\ \frac{2^{qk_2}}{\|G\|_{\mathrm{L}^q}^q},\ \frac{2^{rk_3}}{\|H\|_{\mathrm{L}^r}^r}$$

is the largest. For instance, the part of the sum over

$$S_1 := \{(k_1,k_2,k_3)\in\mathbb{Z}^3 :\ \tfrac{2^{pk_1}}{\|F\|_{\mathrm{L}^p}^p} \geq \tfrac{2^{qk_2}}{\|G\|_{\mathrm{L}^q}^q},\ \tfrac{2^{pk_1}}{\|F\|_{\mathrm{L}^p}^p} \geq \tfrac{2^{rk_3}}{\|H\|_{\mathrm{L}^r}^r}\}$$

is controlled as

$$\sum_{k_1\in\mathbb{Z}} \frac{2^{pk_1}}{\|F\|_{\mathrm{L}^p}^p}\Big(\sum_{Q\in\mathcal{M}_{k_1}^F}|Q|\Big) \sum_{\substack{k_2,k_3\in\mathbb{Z}\\ (k_1,k_2,k_3)\in S_1}} \left(\frac{2^{qk_2}/\|G\|_{\mathrm{L}^q}^q}{2^{pk_1}/\|F\|_{\mathrm{L}^p}^p}\right)^{\frac{1}{q}} \left(\frac{2^{rk_3}/\|H\|_{\mathrm{L}^r}^r}{2^{pk_1}/\|F\|_{\mathrm{L}^p}^p}\right)^{\frac{1}{r}} \lesssim_{p,q,r} 1\,,$$

which follows from (3.11) and by summing two convergent geometric series with their largest terms at most 1 and ratios equal to $\frac{1}{2}$.

3.4 Extending the range of exponents

Extension of the main estimate to the range $p\leq 2$ or $q\leq 2$ follows from the conditional result of Bernicot in [1]. Here we repeat his argument in the dyadic case, where it is a bit simpler. His idea is to use one-dimensional Calderón-Zygmund decomposition in each fiber $F(\cdot,y)$ or $G(x,\cdot)$.

We start with an estimate obtained in the previous section[1]:

$$\|T_{\mathrm{d}}(F,G)\|_{\mathrm{L}^{pq/(p+q),\infty}} \leq \|T_{\mathrm{d}}(F,G)\|_{\mathrm{L}^{pq/(p+q)}} \lesssim_{p,q} \|F\|_{\mathrm{L}^p}\|G\|_{\mathrm{L}^q}\,, \tag{3.12}$$

[1]Here $\|F\|_{\mathrm{L}^{p,\infty}} := \sup_{\alpha>0}\alpha\,|\{|F|>\alpha\}|^{1/p}$ denotes the weak L^p norm.

for some $2 < p, q < \infty$, $\frac{1}{p} + \frac{1}{q} > \frac{1}{2}$. If we prove the weak estimate

$$\|T_{\mathrm{d}}(F, G)\|_{\mathrm{L}^{p/(p+1),\infty}} \lesssim_{p,q} \|F\|_{\mathrm{L}^p} \|G\|_{\mathrm{L}^1}\,, \tag{3.13}$$

then T_{d} will be bounded in the whole range of Theorem 3.1, by real interpolation of multilinear operators, as stated for instance in [19] or [50]. We first cover the part $p > 2$, $q \le 2$, then use (3.6) for $p \le 2$, $q > 2$, and finally repeat the argument to tackle the case $p, q \le 2$.

By homogeneity we may assume $\|F\|_{\mathrm{L}^p} = \|G\|_{\mathrm{L}^1} = 1$. For each $x \in \mathbb{R}$ denote by $\mathcal{J}_x$ the collection of all maximal dyadic intervals J with the property

$$\frac{1}{|J|} \int_J |G(x, y)|\, dy > 1\,.$$

Furthermore, set

$$E := \bigcup_{x \in \mathbb{R}} \bigcup_{J \in \mathcal{J}_x} (\{x\} \times J)\,.$$

If G is a dyadic step function, i.e. a finite linear combination of characteristic functions of dyadic squares, then the set E is simply a finite union of dyadic rectangles. Using disjointness of $J \in \mathcal{J}_x$

$$|E| = \int_{\mathbb{R}} \sum_{J \in \mathcal{J}_x} |J|\, dx \le \int_{\mathbb{R}} \Big(\sum_{J \in \mathcal{J}_x} \int_J |G(x, y)|\, dy \Big)\, dx \le 1\,. \tag{3.14}$$

Next, we define "the good part" of G by

$$\widetilde{G}(x, y) := \begin{cases} \frac{1}{|J|} \int_J G(x, v) dv, & \text{for } y \in J \in \mathcal{J}_x, \\ G(x, y), & \text{for } (x, y) \notin E. \end{cases}$$

By the construction of $\mathcal{J}_x$ we have $\|\widetilde{G}\|_{\mathrm{L}^\infty} \le 2$, and from $\|\widetilde{G}\|_{\mathrm{L}^1} \le 1$ we also get $\|\widetilde{G}\|_{\mathrm{L}^q} \le 2$, so using the known Estimate (3.12) we obtain

$$\big|\big\{(x, y) : |T_{\mathrm{d}}(F, \widetilde{G})(x, y)| > 1\big\}\big| \lesssim_{p,q} 1\,. \tag{3.15}$$

As the last ingredient, we show that

$$\Big(\int_{\mathbb{R}} \big(G(x,v) - \widetilde{G}(x,v)\big)\, \psi^{\mathrm{d}}_{J'}(v) dv \Big)\, \psi^{\mathrm{d}}_{J'}(y) = 0 \tag{3.16}$$

for every $J' \in \mathcal{D}$, whenever $(x,y) \notin E$. Since $G(x,\cdot) - \widetilde{G}(x,\cdot)$ is supported on $\bigcup_{J\in\mathcal{J}_x} J$, this in turn will follow from

$$\Big(\int_{\mathbb{R}} \big(G(x,v) - \widetilde{G}(x,v)\big)\, \psi^{\mathrm{d}}_{J'}(v)\, \mathbf{1}_J(v)\, dv \Big)\, \psi^{\mathrm{d}}_{J'}(y) = 0 \tag{3.17}$$

for every $J \in \mathcal{J}_x$. In order to verify (3.17) it is enough to consider $J \cap J' \neq \emptyset$ and $y \in J'$, and since $y \notin J$, we conclude that J is strictly contained in J'. In this case $\psi^{\mathrm{d}}_{J'}(v)\, \mathbf{1}_J(v) = \pm |J'|^{-1/2} \mathbf{1}_J(v)$, so we only have to observe $\int_J \big(G(x,v) - \widetilde{G}(x,v)\big) dv = 0$, by the definition of $\widetilde{G}$.

Equation (3.16) immediately gives $T_{\mathrm{d}}(F, G - \widetilde{G})(x,y) = 0$ for $(x,y) \notin E$, so

$$\big\{(x,y) : |T_{\mathrm{d}}(F,G)(x,y)| > 1\big\} \subseteq E \cup \big\{(x,y) : |T_{\mathrm{d}}(F,\widetilde{G})(x,y)| > 1\big\},$$

and then from (3.14) and (3.15)

$$\big|\big\{(x,y) : |T_{\mathrm{d}}(F,G)(x,y)| > 1\big\}\big| \ \lesssim_{p,q}\ 1\,.$$

This establishes (3.13) by dyadic scaling.

3.5 Transition to the continuous case

Now we turn to the task of proving strong estimates for T_{c} in the range from part (a) of Theorem 3.1:

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

for $1 < p, q < \infty$, $\frac{1}{p} + \frac{1}{q} > \frac{1}{2}$. In order to get the boundary weak estimates, one can later proceed as in [1].

Let φ and ψ be as in the introduction. If $\int_{\mathbb{R}} \varphi = 0$, then $T_{\mathrm{c}}(F,G)$ is dominated by

$$\Big(\sum_{k\in\mathbb{Z}} |\mathrm{P}^{(1)}_{\varphi_k} F|^2\Big)^{1/2} \Big(\sum_{k\in\mathbb{Z}} |\mathrm{P}^{(2)}_{\psi_k} G|^2\Big)^{1/2},$$

and it is enough to use bounds for the two square functions. Otherwise, we have $0 < |\int_{\mathbb{R}} \varphi| \lesssim 1$, so let us normalize $\int_{\mathbb{R}} \varphi = 1$.

A tool that comes in very handy here is the square function introduced by Calderón and generalized by Jones, Seeger, and Wright in [22]. It effectively compares convolutions to martingale averages, allowing us to do the transition easily.

Proposition 3.3 (from [22])**.** *Let φ be a function satisfying* (3.2) *and* $\int_{\mathbb{R}} \varphi = 1$. *The square function*

$$\mathcal{S}_{\mathrm{JSW},\varphi} f := \Big(\sum_{k\in\mathbb{Z}} \big|\mathrm{P}_{\varphi_k} f - \mathbb{E}_k f\big|^2\Big)^{1/2}$$

is bounded from $\mathrm{L}^p(\mathbb{R})$ to $\mathrm{L}^p(\mathbb{R})$ for $1 < p < \infty$, with the constant depending only on p.

Let ϕ be a nonnegative C^∞ function such that $\hat{\phi}(\xi) = 1$ for $|\xi| \leq 2^{-0.6}$, and $\hat{\phi}(\xi) = 0$ for $|\xi| \geq 2^{-0.4}$. We regard it as fixed, so we do not keep track of dependence of constants on ϕ. For any $a \in \mathbb{R}$ define ϕ_a, ϑ_a, ρ_a by

$$\begin{aligned}
\hat{\phi}_a(\xi) &:= \hat{\phi}(2^{-a}\xi)\,,\\
\hat{\vartheta}_a(\xi) &:= \hat{\phi}(2^{-a-1}\xi) - \hat{\phi}(2^{-a}\xi) \;=\; \hat{\phi}_{a+1}(\xi) - \hat{\phi}_a(\xi)\,,\\
\hat{\rho}_a(\xi) &:= \hat{\phi}(2^{-a-0.6}\xi) - \hat{\phi}(2^{-a-0.5}\xi)\,,
\end{aligned}$$

so that in particular

$$\hat{\vartheta}_a = 1 \quad \text{on } \mathrm{supp}(\hat{\rho}_a)\,, \tag{3.18}$$

$$\textstyle\sum_{i=-20}^{20} \hat{\rho}_{k+0.1i} = 1 \quad \text{on } \mathrm{supp}(\hat{\psi}_k)\,, \tag{3.19}$$

$$\textstyle\sum_{i=-20}^{20} \hat{\rho}_{k+0.1i} = 0 \quad \text{on } \mathrm{supp}(\hat{\psi}_{k'}) \;\text{ if } |k'-k| \geq 10\,. \tag{3.20}$$

We first use Proposition 3.3 to obtain bounds for a special case of our continuous twisted paraproduct:

$$T_{\varphi,\vartheta,b}(F,G) := \sum_{k\in\mathbb{Z}} (\mathrm{P}^{(1)}_{\varphi_k} F)(\mathrm{P}^{(2)}_{\vartheta_{k+b}} G)\,, \tag{3.21}$$

where $b \in \mathbb{R}$ is a fixed parameter. The constants can depend on b, as later b will take only finitely many concrete values. Since we have already established estimates for (3.1), it is enough to bound their difference:

$$\big\| T_{\varphi,\vartheta,b}(F,G) - T_{\mathrm{d}}(F,G) \big\|_{\mathrm{L}^{pq/(p+q)}} \lesssim_{p,q,b} \|F\|_{\mathrm{L}^p} \|G\|_{\mathrm{L}^q}\,. \tag{3.22}$$

We introduce a mixed-type operator

$$T_{\mathrm{aux},b}(F,G) := \sum_{k\in\mathbb{Z}} (\mathbb{E}^{(1)}_k F)(\mathrm{P}^{(2)}_{\vartheta_{k+b}} G)\,.$$

Using the Cauchy-Schwarz inequality in $k \in \mathbb{Z}$, one gets

$$\big| T_{\varphi,\vartheta,b}(F,G) - T_{\mathrm{aux},b}(F,G) \big| \;\leq\; \Big(\sum_{k\in\mathbb{Z}} \big| \mathrm{P}^{(1)}_{\varphi_k} F - \mathbb{E}^{(1)}_k F \big|^2 \Big)^{1/2} \Big(\sum_{k\in\mathbb{Z}} \big| \mathrm{P}^{(2)}_{\vartheta_{k+b}} G \big|^2 \Big)^{1/2}.$$

The first term on the right hand side is $\mathcal{S}^{(1)}_{\mathrm{JSW},\varphi} F$, while the second one is the ordinary square function in the second variable, as $\int_{\mathbb{R}} \vartheta_b = 0$. Next, one can rewrite $T_{\mathrm{aux},b}$ and T_{d} as

$$T_{\mathrm{aux},b}(F,G) = FG - \sum_{k\in\mathbb{Z}} (\Delta^{(1)}_k F)(\mathrm{P}^{(2)}_{\phi_{k+1+b}} G)\,,$$

$$T_{\mathrm{d}}(F,G) = FG - \sum_{k\in\mathbb{Z}} (\Delta^{(1)}_k F)(\mathbb{E}^{(2)}_{k+1} G)\,.$$

Subtracting and using the Cauchy-Schwarz inequality in $k \in \mathbb{Z}$, this time we obtain

$$\big| T_{\mathrm{aux},b}(F,G) - T_{\mathrm{d}}(F,G) \big| \;\leq\; \Big(\sum_{k\in\mathbb{Z}} \big| \Delta^{(1)}_k F \big|^2 \Big)^{1/2} \Big(\sum_{k\in\mathbb{Z}} \big| \mathrm{P}^{(2)}_{\phi_{k+b}} G - \mathbb{E}^{(2)}_k G \big|^2 \Big)^{1/2}.$$

The first term on the right hand side is just the dyadic square function in the first variable, while the second term is $\mathcal{S}^{(2)}_{\mathrm{JSW},\phi_b} G$. The Estimate (3.22) now follows from Proposition 3.3 and bounds on the two common square functions.

Actually, we need a "sparser" paraproduct than the one in (3.21):

$$T^{10\mathbb{Z}}_{\varphi,\rho,b,l}(F,G) := \sum_{j\in\mathbb{Z}} (\mathrm{P}^{(1)}_{\varphi_{10j+l}} F)(\mathrm{P}^{(2)}_{\rho_{10j+l+b}} G)\,, \tag{3.23}$$

for $l = 0, 1, \ldots, 9$. To see that (3.23) is bounded too, we define

$$\widetilde{G}_{b,l} := \sum_{j\in\mathbb{Z}} \mathrm{P}^{(2)}_{\rho_{10j+l+b}} G\,.$$

Notice that because of (3.18) we have

$$\mathrm{P}^{(2)}_{\vartheta_{k+b}} \widetilde{G}_{b,l} = \begin{cases} \mathrm{P}^{(2)}_{\rho_{10j+l+b}} G, & \text{for } k = 10j + l \in 10\mathbb{Z} + l, \\ 0, & \text{for } k \in \mathbb{Z},\ k \notin 10\mathbb{Z} + l \end{cases}$$

and the Littlewood-Paley inequality gives

$$\|\widetilde{G}_{b,l}\|_{\mathrm{L}^q} \lesssim_{q,b,l} \|G\|_{\mathrm{L}^q}\,.$$

It remains to write

$$T^{10\mathbb{Z}}_{\varphi,\rho,b,l}(F,G) = T_{\varphi,\vartheta,b}(F,\widetilde{G}_{b,l})$$

and use boundedness of (3.21).

Finally, we tackle the original operator (3.3). The following computation is possible because of (3.19) and (3.20).

$$\begin{aligned} \sum_{k\in\mathbb{Z}} \hat{\varphi}_k(\xi)\hat{\psi}_k(\eta) &= \sum_{l=0}^{9} \sum_{j\in\mathbb{Z}} \hat{\varphi}_{10j+l}(\xi)\hat{\psi}_{10j+l}(\eta) \\ &= \sum_{l=0}^{9} \sum_{i=-20}^{20} \sum_{j\in\mathbb{Z}} \hat{\varphi}_{10j+l}(\xi)\hat{\rho}_{10j+l+0.1i}(\eta)\hat{\psi}_{10j+l}(\eta) \\ &= \sum_{l=0}^{9} \sum_{i=-20}^{20} \sum_{j\in\mathbb{Z}} \hat{\varphi}_{10j+l}(\xi)\hat{\rho}_{10j+l+0.1i}(\eta)\hat{\Psi}_l(\eta) \end{aligned}$$

Above we have set $\Psi_l := \sum_{m\in\mathbb{Z}} \psi_{10m+l}$. This "symbol identity" leads us to

$$T_{\mathrm{c}}(F,G) = \sum_{l=0}^{9} \sum_{i=-20}^{20} T^{10\mathbb{Z}}_{\varphi,\,\rho,\,0.1i,\,l}(F, \mathrm{P}^{(2)}_{\Psi_l} G)\,. \tag{3.24}$$

Since $\hat{\psi}$ has a compact support and $|\hat{\psi}(\eta)|$, $|\frac{d}{d\eta}\hat{\psi}(\eta)| \lesssim 1$ by (3.2), scaling gives $|\hat{\Psi}_l(\eta)| \lesssim 1$, $\left|\frac{d}{d\eta}\hat{\Psi}_l(\eta)\right| \lesssim |\eta|^{-1}$, and thus the Hörmander-Mikhlin multiplier theorem (in one variable) implies

$$\big\|\mathrm{P}^{(2)}_{\Psi_l} G\big\|_{\mathrm{L}^q} \lesssim_{q,l} \|G\|_{\mathrm{L}^q}\,.$$

It remains to use (3.24) and boundedness of (3.23).

3.6 Endpoint counterexamples

We give the arguments in the dyadic setting, the continuous case being similar. First we show that T_{d} does not map boundedly

$$\mathrm{L}^\infty(\mathbb{R}^2) \times \mathrm{L}^q(\mathbb{R}^2) \to \mathrm{L}^{q,\infty}(\mathbb{R}^2)$$

for $1 \le q < \infty$. Take G to be

$$G(x,y) := \mathbf{1}_{[0,2^{-n})}(x) \sum_{k=1}^{n} R_k(y)\,,$$

for some positive integer n, where R_k denotes the k-th Rademacher function[2] on $[0,1)$, i.e.

$$R_k := \sum_{J\subseteq[0,1),\,|J|=2^{-k+1}} (\mathbf{1}_{J_{\mathrm{left}}} - \mathbf{1}_{J_{\mathrm{right}}})\,.$$

Recall Khintchine's inequality, which can be formulated as:

$$\Big\| \sum_{k=1}^{n} c_k R_k \Big\|_{\mathrm{L}^q} \sim_q \Big(\sum_{k=1}^{n} |c_k|^2 \Big)^{1/2}, \qquad \text{for } 0 < q < \infty\,,$$

[2]Linear combinations of Rademacher functions $\sum_k c_k R_k(t)$ are dyadic analogues of lacunary trigonometric series $\sum_k c_k e^{i2^k t}$.

giving us $\|G\|_{\mathrm{L}^q} \sim_q 2^{-n/q} n^{1/2}$. Observe that

$$(\Delta_k^{(2)} G)(x,y) = \mathbf{1}_{[0,2^{-n})}(x) R_{k+1}(y), \quad \text{for } k = 0,1,\ldots,n-1\,.$$

We choose F supported in the unit square $[0,1)^2$ and defined by

$$F(x,y) := \begin{cases} 2R_j(y) - R_{j+1}(y), & \text{for } x \in [2^{-j}, 2^{-j+1}),\ j = 1,\ldots,n-1, \\ R_n(y), & \text{for } x \in [0, 2^{-n+1}). \end{cases}$$

Note that $\|F\|_{\mathrm{L}^\infty} \leq 3$ and $(\mathbb{E}_k^{(1)} F)(x,y) = R_{k+1}(y)$ for $x \in [0,2^{-n})$, $k = 0,1,\ldots,n-1$. Since the output function is now simply $T_{\mathrm{d}}(F,G) = n\,\mathbf{1}_{[0,2^{-n})\times[0,1)}$, we have

$$\frac{\|T_{\mathrm{d}}(F,G)\|_{\mathrm{L}^{q,\infty}}}{\|F\|_{\mathrm{L}^\infty}\|G\|_{\mathrm{L}^q}} \gtrsim_q \frac{2^{-n/q} n}{2^{-n/q} n^{1/2}} = n^{1/2}\,,$$

which shows unboundedness.

The remaining estimate $\|T_{\mathrm{d}}(F,G)\|_{\mathrm{L}^\infty} \lesssim \|F\|_{\mathrm{L}^\infty}\|G\|_{\mathrm{L}^\infty}$ is even easier to disprove. For a positive integer n, take

$$F(x,y) := \begin{cases} 1, & \text{for } x \in \bigcup_{j=0}^{n-1} [2^{-2j-1}, 2^{-2j}),\ y \in [0,1), \\ 0, & \text{otherwise} \end{cases}$$

and $G(x,y) := F(y,x)$. It is easy to see that $|T_{\mathrm{d}}(F,G)(x,y)| \sim n$ on the square $(x,y) \in [0,2^{-2n})^2$.

CHAPTER 4

Two-dimensional paraproduct-type multilinear forms

4.1 Generalized paraproducts

This chapter is devoted to a study of a somewhat general class of multilinear forms that generalizes ordinary paraproducts as well as the twisted paraproduct from Chapter 3. This time we confine ourselves to bounding only dyadic model sums using the machinery from Chapter 2, because the simple transference trick used in Section 3.5 (from the dyadic to the continuous model) is no longer available. Some continuous results can still be obtained by averaging over translated and dilated dyadic grids, but we do not discuss those ideas here. Furthermore, the "fiber-wise" Calderón-Zygmund decomposition from [1] does not apply either (as it did in Section 3.4), so there does not seem to be a pre-existing result allowing any extension of the exponent range. Thus, Theorem 4.1 below does not include Theorem 3.1. On the other hand, the generalization presented in this chapter requires more involved combinatorial reasoning.

Let m, n be positive integers and choose

$$E \subseteq \{1, \ldots, m\} \times \{1, \ldots, n\}, \quad S \subseteq \{1, \ldots, m\}, \quad T \subseteq \{1, \ldots, n\}\,.$$

It will be convenient to represent E as the set of edges of a simple bipartite undirected graph with vertices $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$, where x_i and y_j

are connected by an edge if and only if $(i,j) \in E$. Also, we regard elements of S and T as "selected" vertices from these two vertex-sets respectively. We additionally require

$$|S| \geq 2 \quad \text{or} \quad |T| \geq 2\,.$$

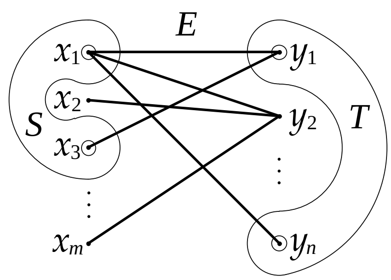

Figure 4.1: Graph interpretation of a triple (E,S,T). Selected vertices are circled.

To each such triple (E,S,T) we associate a multilinear form $\Lambda = \Lambda_{E,S,T}$ acting on $|E|$ functions by

$$\Lambda\big((F_{i,j})_{(i,j)\in E}\big) \;:=\; \sum_{I\times J\in\mathcal{C}} |I|^{2-\frac{m+n}{2}} \int_{\mathbb{R}^{m+n}} \Big(\prod_{(i,j)\in E} F_{i,j}(x_i,y_j)\Big)$$
$$\Big(\prod_{i\in S}\psi_I^{\mathrm{d}}(x_i)\Big)\Big(\prod_{i\in S^c}\varphi_I^{\mathrm{d}}(x_i)\Big)\Big(\prod_{j\in T}\psi_J^{\mathrm{d}}(y_j)\Big)\Big(\prod_{j\in T^c}\varphi_J^{\mathrm{d}}(y_j)\Big)\, dx_1\ldots dx_m\, dy_1\ldots dy_n\,.$$

As before, $\mathcal{C}$ denotes the collection of all dyadic squares in $\mathbb{R}^2$ and

$$\varphi_I^{\mathrm{d}} := |I|^{-1/2}\mathbf{1}_I, \qquad \psi_I^{\mathrm{d}} := |I|^{-1/2}(\mathbf{1}_{I_{\mathrm{left}}} - \mathbf{1}_{I_{\mathrm{right}}})\,.$$

To make sure that the above summand is well-defined for each $I\times J\in\mathcal{C}$ and that all of the succeeding arguments are finitary, we suppose that $F_{i,j}$ are measurable, bounded, and compactly supported functions.[1] Since this is only a qualitative assumption, any quantitative bounds can be extended by density arguments. The

[1] Absolute convergence of the series $\sum_{I\times J\in\mathcal{C}}$ will be a part of Theorem 4.1.

normalization $|I|^{2-(m+n)/2}$ is chosen so that Λ is invariant under simultaneous dyadic dilations of all functions:

$$\Lambda\big((\mathrm{D}_{2^l} F_{i,j})_{(i,j)\in E}\big) \;=\; 2^{2l}\,\Lambda\big((F_{i,j})_{(i,j)\in E}\big)\,,$$

where $l \in \mathbb{Z}$ and $(\mathrm{D}_{2^l} F)(x,y) := F(2^{-l}x, 2^{-l}y)$.

We can describe the structure of Λ in words:

- Every edge (x_i, y_j) contributes with a function $F_{i,j}$.
- Each vertex, x_i or y_j, carries a "dyadic bump function" (either φ^{d} or ψ^{d}).
- Selected vertices carry "dyadic bump functions" of mean zero (i.e. ψ^{d}).
- At least one bipartition class, $\{x_1, \ldots, x_m\}$ or $\{y_1, \ldots, y_n\}$, contains at least two selected vertices.

The last condition is an analogue of the standard cancellation condition for classical paraproducts, see Section 1.1.

The form Λ can be rewritten using the notation from Chapter 2 as

$$\Lambda\big((F_{i,j})_{(i,j)\in E}\big) \;= \sum_{Q\in\mathcal{C}} |Q|\,\mathcal{A}_Q\big((F_{i,j})_{(i,j)\in E}\big)\,,$$

with

$$\mathcal{A}_{I\times J}\big((F_{i,j})_{(i,j)\in E}\big) = \Big[\Big\langle \prod_{(i,j)\in E} F_{i,j}(x_i, y_j)\Big\rangle_{\substack{x_i\in I \text{ for } i\in S\\ y_j\in J \text{ for } j\in T}}\Big]_{\substack{x_i\in I \text{ for } i\in S^c\\ y_j\in J \text{ for } j\in T^c}}\,.$$

The latter formulation also justifies the factor $|I|^{2-(m+n)/2}$ in the definition of Λ.

Let $d_{i,j}$ denote larger size of the two bipartition classes of the connected component containing an edge (x_i, y_j). In more details, the graph described above splits into connected components, i.e. maximal connected subgraphs. Suppose that there are k components that contain at least one edge and list their vertex-sets as

$$X_1 \cup Y_1,\; X_2 \cup Y_2,\; \ldots,\; X_k \cup Y_k\,,$$

where $X_l \subseteq \{x_1, \ldots, x_m\}$ and $Y_l \subseteq \{y_1, \ldots, y_n\}$ for $l = 1, \ldots, k$. We define $d_{i,j}$ to be

$$d_{i,j} = d^{(l)} := \max\{|X_l|, |Y_l|\}$$

if $x_i \in X_l,\; y_j \in Y_l$ for some $l \in \{1, 2, \ldots, k\}$. Since $d_{i,j}$ depends on the component only, we sometimes prefer to write it as $d^{(l)}$. Vertices in $\{x_1, \ldots, x_m\} \setminus (X_1 \cup \ldots \cup X_k)$ and $\{y_1, \ldots, y_n\} \setminus (Y_1 \cup \ldots \cup Y_k)$ are *isolated*, i.e. no edge is incident to any of them. For notational convenience we also denote

$$\mathcal{X}_l := \big\{i \in \{1, \ldots, m\} \,:\, x_i \in X_l\big\},$$
$$\mathcal{Y}_l := \big\{j \in \{1, \ldots, n\} \,:\, y_j \in Y_l\big\},$$

for $l = 1, \ldots, k$.

Connected components are useful because the associated form factorizes according to the splitting.

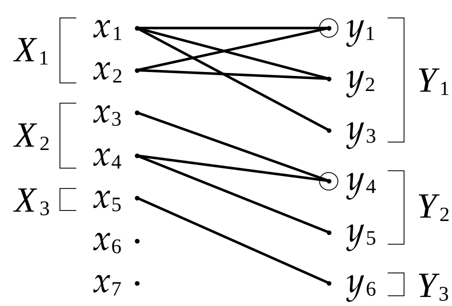

Figure 4.2: An example of a bipartite graph and its splitting.

An illustrative example is given in Figure 4.2 and it has

$m = 7,\;\; n = 6,\;\; E = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(3,4),(4,4),(4,5),(5,6)\big\},$

$S = \emptyset,\;\; T = \{1,4\},\;\; d_{1,1} = d_{1,2} = d_{1,3} = d_{2,1} = d_{2,2} = 3,\;\; d_{3,4} = d_{4,4} = d_{4,5} = 2,\;\; d_{5,6} = 1.$

The splitting into connected components is

$$\mathcal{X}_1 = \{1,2\},\ \ \mathcal{Y}_1 = \{1,2,3\},\ \ \mathcal{X}_2 = \{3,4\},\ \ \mathcal{Y}_2 = \{4,5\},\ \ \mathcal{X}_3 = \{5\},\ \ \mathcal{Y}_3 = \{6\}.$$

The form we associate to this graph is

$$\Lambda = \sum_{Q\in\mathcal{C}} |Q|\,\mathcal{A}_Q^{(1)}\,\mathcal{A}_Q^{(2)}\,\mathcal{A}_Q^{(3)}\,,$$

where

$$\begin{aligned}
\mathcal{A}^{(1)} &= \big[\big\langle F_{1,1}(x_1,y_1)F_{1,2}(x_1,y_2)F_{1,3}(x_1,y_3)F_{2,1}(x_2,y_1)F_{2,2}(x_2,y_2)\big\rangle_{y_1}\big]_{x_1,x_2,y_2,y_3},\\
\mathcal{A}^{(2)} &= \big[\big\langle F_{3,4}(x_3,y_4)F_{4,4}(x_4,y_4)F_{4,5}(x_4,y_5)\big\rangle_{y_4}\big]_{x_3,x_4,y_5},\\
\mathcal{A}^{(3)} &= \big[F_{5,6}(x_5,y_6)\big]_{x_5,y_6}.
\end{aligned}$$

Now we state the main result.

Theorem 4.1. *Let (E,S,T) and $(d_{i,j})_{(i,j)\in E}$ be as above. (Recall that we assume $|S|\geq 2$ or $|T|\geq 2$.) The associated form Λ satisfies the estimate*

$$\big|\Lambda\big((F_{i,j})_{(i,j)\in E}\big)\big| \ \lesssim_{m,n,(p_{i,j})} \prod_{(i,j)\in E} \|F_{i,j}\|_{\mathrm{L}^{p_{i,j}}(\mathbb{R}^2)}$$

whenever the exponents $(p_{i,j})_{(i,j)\in E}$ are such that $\sum_{(i,j)\in E}\frac{1}{p_{i,j}} = 1$ and $d_{i,j} < p_{i,j} < \infty$ for each $(i,j)\in E$. Moreover, the series defining Λ converges absolutely.

Before the proof, let us comment on a couple of already familiar particular instances.

Classical paraproducts.

$m=n,\ \ E = \big\{(i,i) : i\in\{1,\ldots,n\}\big\},\ \ |S|\geq 2,\ \ T=\emptyset,\ \ d_{i,i}=1.$

This special case leads to ordinary two-dimensional dyadic paraproducts[2]

$$\Lambda(F_1,\ldots,F_n) = \sum_{I\times J\in\mathcal{C}} |I|^{2-n}\Big(\prod_{i\in S}\big\langle F_i, \psi_I^{\mathrm{d}}\otimes\varphi_J^{\mathrm{d}}\big\rangle_{\mathrm{L}^2(\mathbb{R}^2)}\Big)\Big(\prod_{i\in S^c}\big\langle F_i, \varphi_I^{\mathrm{d}}\otimes\varphi_J^{\mathrm{d}}\big\rangle_{\mathrm{L}^2(\mathbb{R}^2)}\Big)$$

[2]Here $\langle\,\cdot\,,\cdot\,\rangle_{\mathrm{L}^2(\mathbb{R}^2)}$ denotes the standard inner product on $\mathrm{L}^2(\mathbb{R}^2)$.

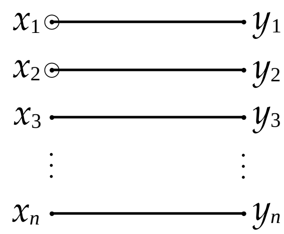

Figure 4.3: Bipartite graph corresponding to a classical paraproduct.

and Theorem 4.1 yields the inequality

$$|\Lambda(F_1, \ldots, F_n)| \lesssim_{n,(p_i)} \prod_{i=1}^{n} \|F_i\|_{\mathrm{L}^{p_i}(\mathbb{R}^2)}$$

for $\sum_{i=1}^{n} \frac{1}{p_i} = 1,\ 1 < p_i < \infty$.

Twisted paraproduct.

$m = n = 2,\ \ E = \big\{(1,2),(2,1),(2,2)\big\},\ \ S = \emptyset,\ \ T = \{1,2\},\ \ d_{i,j} = 2.$

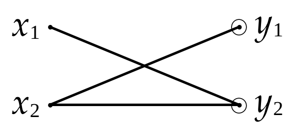

Figure 4.4: Bipartite graph corresponding to the twisted paraproduct.

This case is exactly the dyadic variant of the twisted paraproduct (2.13) and Theorem 4.1 claims Estimate (3.8) in the range $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1,\ 2 < p, q, r < \infty$ only.

The next two sections are devoted to the proof of Theorem 4.1. Note that isolated vertices contribute to the form Λ in a trivial way. If x_i is a non-selected isolated vertex, then no functions $F_{i,j}$ will have x_i as a variable, so the contribution of x_i is only in the factor

$$|I|^{-\frac{1}{2}} \int_{\mathbb{R}} \varphi_I^{\mathrm{d}}(x_i)\, dx_i = 1\,.$$

On the other hand, if x_i is a selected isolated vertex, then that factor will be

$$|I|^{-\frac{1}{2}} \int_{\mathbb{R}} \psi_I^{\mathrm{d}}(x_i)\, dx_i = 0\,,$$

so consequently $\Lambda \equiv 0$. The same reasoning holds for isolated vertices y_j. Therefore, from now on we assume that the associated bipartite graph has no isolated vertices.

In addition, we may suppose that all the functions $F_{i,j}$ are nonnegative, since otherwise they can be split into positive and negative (real and imaginary) parts and one uses multilinearity of the form.

4.2 A single tree estimate

The main step towards the proof of Theorem 4.1 is an estimate for a single tree, similar to Proposition 3.2. We retain the notation and all assumptions from the previous section.

As in Chapter 3, we need to introduce a local version of Λ. For a finite convex tree of dyadic squares $\mathcal{T}$ we define

$$\Lambda_{\mathcal{T}}\big((F_{i,j})_{(i,j)\in E}\big) := \sum_{Q\in\mathcal{T}} |Q|\,\big|\mathcal{A}_Q\big((F_{i,j})_{(i,j)\in E}\big)\big|\,.$$

Notice an absolute value in the definition, which makes $\Lambda_{\mathcal{T}}$ only multi-sublinear.

Proposition 4.2 (Single tree estimate)**.** *For any finite convex tree $\mathcal{T}$ with root $Q_{\mathcal{T}}$ and leaves $\mathcal{L}(\mathcal{T})$ we have*[3]

$$\Lambda_{\mathcal{T}}\big((F_{i,j})_{(i,j)\in E}\big) \;\lesssim_{m,n} |Q_{\mathcal{T}}| \prod_{(i,j)\in E} \max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})} \big[F_{i,j}^{d_{i,j}}\big]_Q^{1/d_{i,j}}. \tag{4.1}$$

[3]Indeed, it is easy to see that the maximum in (4.1) must be attained at some leaf, so it can be replaced by $\max_{Q\in\mathcal{L}(\mathcal{T})}$. We do not use this fact as it does not simplify any arguments in Section 4.3.

Note that the implicit constant is independent of the tree $\mathcal{T}$ and the functions $F_{i,j}$. We can allow it to depend on the graph, because for each pair (m,n) there are only finitely many choices for (E,S,T). Moreover, $d_{i,j}$ are determined by the graph. By homogeneity of Estimate (4.1) and its invariance under dyadic dilations we can normalize the tree and the functions by $|Q_{\mathcal{T}}| = 1$ and

$$\max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})} \big[F_{i,j}^{d_{i,j}}\big]_Q^{1/d_{i,j}} = 1 \qquad \text{for every } (i,j)\in E\,. \tag{4.2}$$

Thus, our task is to prove

$$\Lambda_{\mathcal{T}}\big((F_{i,j})_{(i,j)\in E}\big) \lesssim_{m,n} 1\,. \tag{4.3}$$

4.2.1 Proof of Proposition 4.2 for complete bipartite graphs

Most of our effort will be spent in this special case, when $E = \{1,\ldots,m\}\times\{1,\ldots,n\}$. In particular, there is only one connected component and $d_{i,j} = \max\{m,n\}$. Later we will reduce the general case to this one.

We begin with a simple estimate for functions on a single square.

Lemma 4.3. *For nonnegative functions $G_{i,j}$ on a dyadic square $I\times J$ the following inequality holds,*

$$\Big[\prod_{\substack{1\le i\le m\\ 1\le j\le n}} G_{i,j}(x_i,y_j)\Big]_{\substack{x_1,\ldots,x_m\in I\\ y_1,\ldots,y_n\in J}} \le \prod_{\substack{1\le i\le m\\ 1\le j\le n}} \Big[G_{i,j}^{\max\{m,n\}}\Big]_{I\times J}^{1/\max\{m,n\}}.$$

Proof of Lemma 4.3. Because of the obvious symmetry, we can assume $m\ge n$. With two applications of (generalized) Hölder's inequality for n and m functions respectively, we estimate the left hand side as

$$\begin{aligned}\Big[\prod_{i=1}^{m}\Big[\prod_{j=1}^{n} G_{i,j}(x_i,y_j)\Big]_{x_i}\Big]_{y_1,\ldots,y_n} &\le \Big[\prod_{i=1}^{m}\prod_{j=1}^{n}\big[G_{i,j}(x_i,y_j)^n\big]_{x_i}^{1/n}\Big]_{y_1,\ldots,y_n}\\ = \prod_{j=1}^{n}\Big[\prod_{i=1}^{m}\big[G_{i,j}(x,y)^n\big]_x^{1/n}\Big]_y &\le \prod_{j=1}^{n}\prod_{i=1}^{m}\big[\big[G_{i,j}(x,y)^n\big]_x^{m/n}\big]_y^{1/m}.\end{aligned}$$

By Jensen's inequality for the power function with exponent $\frac{m}{n} \geq 1$ we have

$$\big[G_{i,j}(x,y)^n\big]_x^{m/n} \leq \big[G_{i,j}(x,y)^m\big]_x,$$

so

$$\Big[\big[G_{i,j}(x,y)^n\big]_x^{m/n}\Big]_y^{1/m} \leq \big[G_{i,j}(x,y)^m\big]_{x,y}^{1/m},$$

which completes the proof. $\square$

In the following discussion, all constructions and (implicit) constants are understood to depend on m and n, but not on the tree $\mathcal{T}$ or the functions $F_{i,j}$. We introduce the notion of a *selective* (m,n)*-partition* as a $(2m+2n)$-tuple of integers

$$\mathbf{p} = (a_1,\ldots,a_m;\, b_1,\ldots,b_n;\, \alpha_1,\ldots,\alpha_m;\, \beta_1,\ldots,\beta_n) \tag{4.4}$$

satisfying:

- $0 \leq \alpha_i \leq a_i$ for $i=1,\ldots,m$ and $0 \leq \beta_j \leq b_j$ for $j=1,\ldots,n$,
- $a_1+\ldots+a_m = m$ and $b_1+\ldots+b_n = n$,
- $\alpha_1+\ldots+\alpha_m$ and $\beta_1+\ldots+\beta_n$ are even,
- $\alpha_1+\ldots+\alpha_m \neq 0$ or $\beta_1+\ldots+\beta_n \neq 0$.

The set of all selective (m,n)-partitions will be denoted $\Omega_{m,n}$. To every $\mathbf{p} \in \Omega_{m,n}$ we associate a paraproduct-type term $\mathcal{A}^{(\mathbf{p})} = \mathcal{A}^{(\mathbf{p})}_{I\times J}\big((F_{i,j})_{(i,j)\in E}\big)$ by

$$\mathcal{A}^{(\mathbf{p})}_{I\times J} := \Bigg[\Bigg\langle \prod_{\substack{1\leq i\leq m\\ 1\leq j\leq n}} \prod_{\substack{1\leq\mu\leq a_i\\ 1\leq\nu\leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)}) \Bigg\rangle_{\substack{x_i^{(\mu)}\in I \text{ for all } (i,\mu)\\ \text{such that } 1\leq\mu\leq\alpha_i\\ y_j^{(\nu)}\in J \text{ for all } (j,\nu)\\ \text{such that } 1\leq\nu\leq\beta_j}} \Bigg]_{\substack{x_i^{(\mu)}\in I \text{ for all } (i,\mu)\\ \text{such that } \alpha_i+1\leq\mu\leq a_i\\ y_j^{(\nu)}\in J \text{ for all } (j,\nu)\\ \text{such that } \beta_j+1\leq\nu\leq b_j}}.$$

Pairs (i,j) with $a_i=0$ or $b_j=0$ do not exist in the above product as we interpret (sub)products over empty ranges to be identically 1. In words, we average the product of mn terms of the pattern $F_{i,j}(x_i, y_j)$ that contains precisely a_i copies of x_i and b_j copies of y_j. Averages of type $\langle\cdot\rangle_{x_i}$ are taken over α_i of the x_i's (i.e. these x_i's are "selected"), while averages of type $[\cdot]_{x_i}$ are taken over $a_i - \alpha_i$ remaining ones. Similarly for y_j's. For instance, to

$$\mathbf{p} = (2,0;\ 2,0,1;\ 0,0;\ 1,0,1) \in \Omega_{2,3}$$

we associate

$$\begin{aligned}\mathcal{A}^{(\mathbf{p})}_{I\times J} = &\Big[\Big\langle F_{1,1}(x_1,y_1)F_{1,1}(x_1,y_1')F_{1,3}(x_1,y_3)\\ &\quad F_{1,1}(x_1',y_1)F_{1,1}(x_1',y_1')F_{1,3}(x_1',y_3)\Big\rangle_{y_1,y_3\in J}\Big]_{x_1,x_1'\in I,\ y_1'\in J}.\end{aligned}$$

For $\mathbf{p} \in \Omega_{m,n}$ given by (4.4) we define the *composition type* of $\mathbf{p}$ to be the vector of first $m+n$ components,

$$\mathrm{comp}(\mathbf{p}) := (a_1,\ldots,a_m;\ b_1,\ldots,b_n)\,,$$

and the *partition type* of $\mathbf{p}$ (and $\mathcal{A}^{(\mathbf{p})}$) to be an $(m+n)$-tuple

$$\mathrm{part}(\mathbf{p}) := (a_1^*,\ldots,a_m^*;\ b_1^*,\ldots,b_n^*)\,,$$

where $a_1^*,\ldots,a_m^*$ is the decreasing rearrangement[4] of $a_1,\ldots,a_m$ and $b_1^*,\ldots,b_n^*$ is the decreasing rearrangement of $b_1,\ldots,b_n$. The set of all these partition types will be denoted $\Omega^*_{m,n}$. Note that $\Omega^*_{m,n}$ has $\mathfrak{p}_\#(m)\mathfrak{p}_\#(n)$ elements, where $\mathfrak{p}_\#(n)$ denotes the number of distinct order-independent positive integer partitions of n, i.e. the number of Young diagrams with n boxes. Actually, we only use that $|\Omega_{m,n}|$ and $|\Omega^*_{m,n}|$ are finite numbers depending solely on m,n.

Finally, we define a strict total order relation $\prec$ on $\Omega^*_{m,n}$ simply as the restriction of the inverse of the lexicographical order on $(m+n)$-tuples of inte-

[4]This means: $a_1^* \geq \ldots \geq a_m^*$ and $a_1^*,\ldots,a_m^*$ is a permutation of the multiset $a_1,\ldots,a_m$.

gers.[5] Since every finite totally ordered set is isomorphic to an initial segment of positive integers, we have a natural rank (i.e. order) function, $\operatorname{ord}\colon \Omega^*_{m,n} \to \{1,2,\ldots,\mathfrak{p}_\#(m)\mathfrak{p}_\#(n)\}$. We simply write $\operatorname{ord}(\mathbf{p})$ for $\operatorname{ord}\big(\operatorname{part}(\mathbf{p})\big)$. For example, the total order on $\Omega^*_{2,3}$ and its rank function are

$$\underset{\operatorname{ord}=1}{(2,0;3,0,0)}\prec\underset{\operatorname{ord}=2}{(2,0;2,1,0)}\prec\underset{\operatorname{ord}=3}{(2,0;1,1,1)}\prec\underset{\operatorname{ord}=4}{(1,1;3,0,0)}\prec\underset{\operatorname{ord}=5}{(1,1;2,1,0)}\prec\underset{\operatorname{ord}=6}{(1,1;1,1,1)}\,.$$

Our goal is to dominate all terms $\mathcal{A}^{(\mathbf{p})}$ by $\Box\mathcal{B}$ for some averaging paraproduct-type expression $\mathcal{B} = \mathcal{B}_Q\big((F_{i,j})_{(i,j)\in E}\big)$ that is controlled in the sense

$$\max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})} \big|\mathcal{B}_Q\big((F_{i,j})_{(i,j)\in E}\big)\big| \lesssim_{m,n} 1\,. \tag{4.5}$$

This expression $\mathcal{B}$ will be the desired Bellman function. The goal will be achieved by mathematical induction on $\operatorname{ord}(\mathbf{p})$ and for this we will need the following crucial reduction lemma.

Lemma 4.4. *For any $\mathbf{p}\in\Omega_{m,n}$ there exists an averaging paraproduct-type term $\mathcal{B}^{(\mathbf{p})} = \mathcal{B}^{(\mathbf{p})}_Q\big((F_{i,j})_{(i,j)\in E}\big)$ satisfying (4.5) and such that for any $0<\delta<1$ we have the estimate*

$$|\mathcal{A}^{(\mathbf{p})}| \;\leq\; \Box\mathcal{B}^{(\mathbf{p})} \;+\; C_{m,n}\,\delta^{-1}\sum_{\substack{\widetilde{\mathbf{p}}\in\Omega_{m,n}\\ \operatorname{ord}(\widetilde{\mathbf{p}})<\operatorname{ord}(\mathbf{p})}} |\mathcal{A}^{(\widetilde{\mathbf{p}})}| \;+\; C_{m,n}\,\delta \sum_{\substack{\widetilde{\mathbf{p}}\in\Omega_{m,n}\\ \operatorname{ord}(\widetilde{\mathbf{p}})\geq\operatorname{ord}(\mathbf{p})}} |\mathcal{A}^{(\widetilde{\mathbf{p}})}|$$

with some constant $C_{m,n}>0$.

Proof of Lemma 4.4. We distinguish two cases depending on positions of selected vertices, i.e. on the last $m+n$ coordinates in (4.4).

[5]Lexicographical order on partitions of a single positive integer extends the common *dominance order*, which is only a partial order. Even though the latter one is already strong enough for intended purpose, we prefer to have linear order to avoid invoking well-founded induction in the proof. For the same reason we decide to order pairs of partitions totally (for two numbers m and n), although we will only need to compare partitions of a single number.

Case 1. $\alpha_i \neq 0$ for at least two indices $i \in \{1, \ldots, m\}$ or $\beta_j \neq 0$ for at least two indices $j \in \{1, \ldots, n\}$.

By symmetry we may assume that $\alpha_1, \alpha_2 \geq 1$ and $a_1 \geq a_2$. In this case we simply take $\mathcal{B}^{(\mathbf{p})} \equiv 0$. Using $|\langle f(y) \rangle_y| \leq [|f(y)|]_y$ and $|AB| \leq \frac{1}{2\delta} A^2 + \frac{\delta}{2} B^2$ we estimate:

$$
\begin{aligned}
|\mathcal{A}^{(\mathbf{p})}| \leq \Bigg[\bigg| \Big\langle \prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{1,j}(x_1, y_j^{(\nu)}) \Big\rangle_{x_1} \Big\langle \prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{2,j}(x_2, y_j^{(\nu)}) \Big\rangle_{x_2} \bigg| \\
\prod_{(i,\mu) \neq (1,1),(2,1)} \Big[\prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)}) \Big]_{x_i^{(\mu)}} \Bigg]_{\text{all } y_j^{(\nu)}} \\
\leq \frac{1}{2\delta} \Bigg[\Big\langle \prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{1,j}(x_1, y_j^{(\nu)}) \Big\rangle_{x_1}^2 \prod_{(i,\mu) \neq (1,1),(2,1)} \Big[\prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)}) \Big]_{x_i^{(\mu)}} \Bigg]_{\text{all } y_j^{(\nu)}} \\
+ \frac{\delta}{2} \Bigg[\Big\langle \prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{2,j}(x_2, y_j^{(\nu)}) \Big\rangle_{x_2}^2 \prod_{(i,\mu) \neq (1,1),(2,1)} \Big[\prod_{\substack{1 \leq j \leq n \\ 1 \leq \nu \leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)}) \Big]_{x_i^{(\mu)}} \Bigg]_{\text{all } y_j^{(\nu)}} \\
= \frac{1}{2} \Big(\delta^{-1} \mathcal{A}^{(\widetilde{\mathbf{p}})} + \delta \, \mathcal{A}^{(\overline{\mathbf{p}})} \Big) .
\end{aligned}
$$

Here $\widetilde{\mathbf{p}}$, $\overline{\mathbf{p}} \in \Omega_{m,n}$ are defined as follows. If $\mathbf{p}$ is given by a $(2m+2n)$-tuple (4.4), then $\widetilde{\mathbf{p}}$ and $\overline{\mathbf{p}}$ will have coordinates:

$$
\widetilde{a}_i = \begin{cases} a_1 + 1, & \text{for } i = 1, \\ a_2 - 1, & \text{for } i = 2, \\ a_i, & \text{for } i \neq 1, 2, \end{cases} \qquad \overline{a}_i = \begin{cases} a_1 - 1, & \text{for } i = 1, \\ a_2 + 1, & \text{for } i = 2, \\ a_i, & \text{for } i \neq 1, 2, \end{cases}
$$

$$
\widetilde{\alpha}_i = \begin{cases} 2, & \text{for } i = 1, \\ 0, & \text{for } i \neq 1, \end{cases} \qquad \overline{\alpha}_i = \begin{cases} 2, & \text{for } i = 2, \\ 0, & \text{for } i \neq 2, \end{cases}
$$

$$
\widetilde{b}_j = \overline{b}_j = b_j, \ \ \widetilde{\beta}_j = \overline{\beta}_j = 0 \quad \text{for every } j \, .
$$

Observe that a_1 appears to the left from a_2 in the list $a_1^* \geq \ldots \geq a_m^*$. Therefore simultaneously increasing a_1 by 1 and decreasing a_2 by 1 we produce a lexicographically larger partition of m, so we conclude $\operatorname{ord}(\widetilde{\mathbf{p}}) < \operatorname{ord}(\mathbf{p})$. On the other hand, both $\operatorname{ord}(\overline{\mathbf{p}}) < \operatorname{ord}(\mathbf{p})$ and $\operatorname{ord}(\overline{\mathbf{p}}) \geq \operatorname{ord}(\mathbf{p})$ are possible, where in the former case we use $\delta < \delta^{-1}$.

Case 2. $\alpha_i \neq 0$ for at most one index $i \in \{1, \ldots, m\}$ and $\beta_j \neq 0$ for at most one index $j \in \{1, \ldots, n\}$.

Without loss of generality we assume $\alpha_i = 0$ for $i \neq 1$ and $\beta_j = 0$ for $j \neq 1$. Note that α_1 and β_1 are even and at least one of them is nonzero, say $\alpha_1 \geq 2$. Since

$$|\mathcal{A}^{(\mathbf{p})}| \leq \Big[\Big\langle \prod_{\substack{1\leq j\leq n \\ 1\leq \nu\leq b_j}} F_{1,j}(x_1, y_j^{(\nu)}) \Big\rangle_{x_1}^2 \prod_{(i,\mu)\neq(1,1),(1,2)} \Big[\prod_{\substack{1\leq j\leq n \\ 1\leq \nu\leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)}) \Big]_{x_i^{(\mu)}} \Big]_{\text{all } y_j^{(\nu)}},$$

we can also assume $\alpha_1 = 2$, $\beta_1 = 0$. Consider

$$\mathcal{B}_{I\times J}^{(\mathbf{p})} := \Big[\prod_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} \prod_{\substack{1\leq \mu\leq a_i \\ 1\leq \nu\leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)}) \Big]_{\substack{x_i^{(\mu)}\in I \text{ for all } (i,\mu) \\ y_j^{(\nu)}\in J \text{ for all } (j,\nu)}}.$$

Indeed, $\mathcal{B}^{(\mathbf{p})}$ depends only on $\mathbf{q} = \operatorname{comp}(\mathbf{p})$. Observe that Lemma 4.3 and Normalization (4.2) guarantee Condition (4.5). Theorem 2.2 gives the equality

$$\Box\mathcal{B}^{(\mathbf{p})} = \sum_{\substack{\mathbf{p}'\in\Omega_{m,n} \\ \operatorname{comp}(\mathbf{p}')=\mathbf{q}}} \binom{\mathbf{q}}{\mathbf{p}'} \mathcal{A}^{(\mathbf{p}')}, \tag{4.6}$$

where

$$\binom{\mathbf{q}}{\mathbf{p}'} := \prod_{i=1}^{m} \binom{a_i}{\alpha_i'} \prod_{j=1}^{n} \binom{b_j}{\beta_j'}.$$

Note that $1 \leq \binom{\mathbf{q}}{\mathbf{p}'} \lesssim_{m,n} 1$. Let us split the summation set

$$\Omega_{m,n,\mathbf{q}} := \big\{ \mathbf{p}' \in \Omega_{m,n} : \operatorname{comp}(\mathbf{p}') = \mathbf{q} \big\}$$

into three parts,

$$\Omega_{m,n,\mathbf{q}}^{(1)} := \big\{ \mathbf{p}' \in \Omega_{m,n,\mathbf{q}} : \alpha_i' \neq 0 \text{ for exactly one } i \text{ and } \beta_j' = 0 \text{ for every } j \big\},$$
$$\Omega_{m,n,\mathbf{q}}^{(2)} := \big\{ \mathbf{p}' \in \Omega_{m,n,\mathbf{q}} : \beta_j' \neq 0 \text{ for exactly one } j \big\},$$

$$\Omega^{(3)}_{m,n,\mathbf{q}} := \big\{\mathbf{p}' \in \Omega_{m,n,\mathbf{q}} : \alpha'_i \neq 0 \text{ for at least two } i \text{ and } \beta'_j = 0 \text{ for every } j\big\}$$
$$\cup \big\{\mathbf{p}' \in \Omega_{m,n,\mathbf{q}} : \beta'_j \neq 0 \text{ for at least two } j\big\}.$$

First, observe that $\mathbf{p} \in \Omega^{(1)}_{m,n,\mathbf{q}}$ and that

$$\mathcal{A}^{(\mathbf{p}')} = \Bigg[\Big\langle \prod_{\substack{1\leq j\leq n\\ 1\leq\nu\leq b_j}} F_{i_0,j}(x_{i_0}, y_j^{(\nu)})\Big\rangle^{2\sigma}_{x_{i_0}} \prod_{\substack{(i,\mu) \text{ such that}\\ i\neq i_0 \text{ or } \mu>2\sigma}} \Big[\prod_{\substack{1\leq j\leq n\\ 1\leq\nu\leq b_j}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)})\Big]_{x_i^{(\mu)}}\Bigg]_{\text{all } y_j^{(\nu)}} \geq 0$$

for every $\mathbf{p}' \in \Omega^{(1)}_{m,n,\mathbf{q}}$, where i_0 is chosen such that $\alpha'_{i_0} = 2\sigma \neq 0$. In particular,

$$0 \leq \mathcal{A}^{(\mathbf{p})} \leq \sum_{\mathbf{p}'\in\Omega^{(1)}_{m,n,\mathbf{q}}} \binom{\mathbf{q}}{\mathbf{p}'} \mathcal{A}^{(\mathbf{p}')}. \tag{4.7}$$

Next, Lemma 2.3 applied to

$$\begin{aligned}\Psi &= \Psi\big((x_i^{(\mu)})_{1\leq i\leq m,\ 1\leq\mu\leq a_i}\big)\\ &= \sum_{j_0=1}^{n}\sum_{\tau=1}^{\lfloor b_{j_0}/2\rfloor} \binom{b_{j_0}}{2\tau} \Big\langle \prod_{\substack{1\leq i\leq m\\ 1\leq\mu\leq a_i}} F_{i,j_0}(x_i^{(\mu)}, y_{j_0})\Big\rangle^{2\tau}_{y_{j_0}} \prod_{\substack{(j,\nu) \text{ such that}\\ j\neq j_0 \text{ or } \nu>2\tau}} \Big[\prod_{\substack{1\leq i\leq m\\ 1\leq\mu\leq a_i}} F_{i,j}(x_i^{(\mu)}, y_j^{(\nu)})\Big]_{y_j^{(\nu)}} \geq 0\end{aligned}$$

(where $\beta'_{j_0} = 2\tau \neq 0$) yields

$$\sum_{\mathbf{p}'\in\Omega^{(2)}_{m,n,\mathbf{q}}} \binom{\mathbf{q}}{\mathbf{p}'} \mathcal{A}^{(\mathbf{p}')} \geq 0. \tag{4.8}$$

Finally, each $\mathcal{A}^{(\mathbf{p}')}$ for $\mathbf{p}' \in \Omega^{(3)}_{m,n,\mathbf{q}}$ can be controlled as in Case 1 to obtain

$$|\mathcal{A}^{(\mathbf{p}')}| \ \leq\ \delta^{-1} \sum_{\substack{\widetilde{\mathbf{p}}\in\Omega_{m,n}\\ \mathrm{ord}(\widetilde{\mathbf{p}})<\mathrm{ord}(\mathbf{p})}} |\mathcal{A}^{(\widetilde{\mathbf{p}})}| \ +\ \delta \sum_{\substack{\widetilde{\mathbf{p}}\in\Omega_{m,n}\\ \mathrm{ord}(\widetilde{\mathbf{p}})\geq\mathrm{ord}(\mathbf{p})}} |\mathcal{A}^{(\widetilde{\mathbf{p}})}|. \tag{4.9}$$

Combining (4.6)–(4.9) proves the stated inequality. □

Figure 4.5 depicts partition types in $\Omega^*_{2,4}$ and ways of controlling $\mathcal{A}^{(\mathbf{p})}$ as in Case 1 of the previous proof. Different kinds of arrows represent different possibilities for various $\mathbf{p} \in \Omega_{2,4}$ with the same $\mathrm{part}(\mathbf{p})$. Labels δ and δ^{-1} represent

$$
\begin{tikzcd}
(1,1;\ 1,1,1,1) \arrow[r, dashed, "{\delta,\delta^{-1}}"] \arrow[d, dotted, "{\delta,\delta^{-1}}"] & (2,0;\ 1,1,1,1) \arrow[d, dotted, "{\delta,\delta^{-1}}"] \\
(1,1;\ 2,1,1,0) \arrow[loop left, "\delta"] \arrow[r, dashed, "{\delta,\delta^{-1}}"] \arrow[d, dotted, "{\delta,\delta^{-1}}"] \arrow[dd, bend right=60, "\delta^{-1}"'] & (2,0;\ 2,1,1,0) \arrow[loop right, "\delta"] \arrow[d, dotted, "{\delta,\delta^{-1}}"] \arrow[dd, bend left=60, "\delta^{-1}"] \\
(1,1;\ 2,2,0,0) \arrow[r, dashed, "{\delta,\delta^{-1}}"] \arrow[d, dotted, "{\delta,\delta^{-1}}"'] & (2,0;\ 2,2,0,0) \arrow[d, dotted, "{\delta,\delta^{-1}}"] \\
(1,1;\ 3,1,0,0) \arrow[u, bend right, "\delta"'] \arrow[r, dashed, "{\delta,\delta^{-1}}"] \arrow[d, "\delta^{-1}"] & (2,0;\ 3,1,0,0) \arrow[u, bend left, "\delta"] \arrow[d, "\delta^{-1}"] \\
(1,1;\ 4,0,0,0) \arrow[r, dashed, "{\delta,\delta^{-1}}"] & (2,0;\ 4,0,0,0)
\end{tikzcd}
$$

Figure 4.5: Recursive control of partition types in $\Omega^*_{2,4}$.

coefficients in the "reduction inequality" for $\mathcal{A}^{(\mathbf{p})}$. It is important that always at least one arrow (the one marked by δ^{-1}) points to a partition type with smaller rank, i.e. points downwards or to the right in the picture.

Lemma 4.5. *There exists a "universal" averaging paraproduct-type expression* $\mathcal{B}^{(m,n)} = \mathcal{B}^{(m,n)}_{I\times J}\big((F_{i,j})_{(i,j)\in E}\big)$ *satisfying* (4.5) *and*

$$\sum_{\mathbf{p}\in\Omega_{m,n}} |\mathcal{A}^{(\mathbf{p})}| \ \le\ \Box\mathcal{B}^{(m,n)}\,.$$

Proof of Lemma 4.5. We prove the following claim by mathematical induction on $\kappa\in\{0,1,2,\ldots,\mathfrak{p}_\#(m)\mathfrak{p}_\#(n)\}$: *For every* $0<\varepsilon<1$ *there exists an averaging paraproduct-type expression* $\mathcal{B}^{(\kappa,\varepsilon)}$ *satisfying*

$$\max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})} \big|\mathcal{B}^{(\kappa,\varepsilon)}_Q\big((F_{i,j})_{(i,j)\in E}\big)\big| \lesssim_{m,n,\varepsilon} 1$$

and

$$\sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\le\kappa}} |\mathcal{A}^{(\mathbf{p})}| \ \le\ \Box\mathcal{B}^{(\kappa,\varepsilon)} \ +\ \varepsilon \sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})>\kappa}} |\mathcal{A}^{(\mathbf{p})}|\,.$$

The bound we need to prove is a special instance for $\kappa = \mathfrak{p}_\#(m)\mathfrak{p}_\#(n)$ and any ε, since then the sum on the right disappears. On the other hand, the sum on the left is 0 for $\kappa = 0$, so the inequality (trivially) holds with $\mathcal{B}^{(\kappa,\varepsilon)} \equiv 0$. We take $\kappa = 0$ as the induction basis.

In the induction step we suppose that the claim holds for some $0 \leq \kappa \leq \mathfrak{p}_\#(m)\mathfrak{p}_\#(n) - 1$. To prove the claim for $\kappa + 1$ we take arbitrary $0 < \varepsilon < 1$. Let $C_{m,n}$ be the constant from Lemma 4.4. Induction hypothesis applied with $\varepsilon' = \big(\frac{\varepsilon}{4C_{m,n}|\Omega_{m,n}|}\big)^2$ gives $\mathcal{B}^{(\kappa,\varepsilon')}$ such that

$$\sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa}} |\mathcal{A}^{(\mathbf{p})}| \;\leq\; \Box\mathcal{B}^{(\kappa,\varepsilon')} \;+\; \big(\tfrac{\varepsilon}{4C_{m,n}|\Omega_{m,n}|}\big)^2 \sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})>\kappa}} |\mathcal{A}^{(\mathbf{p})}|\,. \tag{4.10}$$

Lemma 4.4 applied to each $\mathbf{p} \in \Omega_{m,n}$, $\mathrm{ord}(\mathbf{p}) \leq \kappa + 1$, and $\delta = \frac{\varepsilon}{4C_{m,n}|\Omega_{m,n}|}$ yields

$$\sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa+1}} |\mathcal{A}^{(\mathbf{p})}| \;\leq\; \sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa+1}} \Box\mathcal{B}^{(\mathbf{p})} \;+\; \tfrac{4C_{m,n}^2|\Omega_{m,n}|^2}{\varepsilon} \sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa}} |\mathcal{A}^{(\mathbf{p})}| \;+\; \tfrac{\varepsilon}{4} \sum_{\mathbf{p}\in\Omega_{m,n}} |\mathcal{A}^{(\mathbf{p})}|\,,$$

which combined with (4.10) leads to

$$\sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa+1}} |\mathcal{A}^{(\mathbf{p})}| \;\leq\; \sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa+1}} \Box\mathcal{B}^{(\mathbf{p})} \;+\; \tfrac{4C_{m,n}^2|\Omega_{m,n}|^2}{\varepsilon}\,\Box\mathcal{B}^{(\kappa,\varepsilon')} \;+\; \tfrac{\varepsilon}{2} \sum_{\mathbf{p}\in\Omega_{m,n}} |\mathcal{A}^{(\mathbf{p})}|\,.$$

It remains to split $\sum_{\mathbf{p}\in\Omega_{m,n}}$ according to $\mathrm{ord}(\mathbf{p}) \leq \kappa + 1$ or $\mathrm{ord}(\mathbf{p}) > \kappa + 1$, move the former summands to the left hand side, and multiply the inequality by 2. We set

$$\mathcal{B}^{(\kappa+1,\,\varepsilon)} \;=\; 2 \sum_{\substack{\mathbf{p}\in\Omega_{m,n}\\ \mathrm{ord}(\mathbf{p})\leq\kappa+1}} \mathcal{B}^{(\mathbf{p})} \;+\; \tfrac{8C_{m,n}^2|\Omega_{m,n}|^2}{\varepsilon}\,\mathcal{B}^{(\kappa,\varepsilon')}\,.$$

$\Box$

Let us finally complete the proof of Proposition 4.2 in this special case. Even though the general term $\mathcal{A}$ of Λ is not necessarily among $\mathcal{A}^{(\mathbf{p})}$, $\mathbf{p} \in \Omega_{m,n}$, it can still be easily dominated by two of these terms. Without loss of generality suppose $|S| \geq 2$ and take $i_1, i_2 \in S,\ i_1 \neq i_2$.

$$
\begin{aligned}
|\mathcal{A}| \ \leq \ & \Big[\Big| \Big\langle \prod_{1\leq j\leq n} F_{i_1,j}(x_{i_1}, y_j) \Big\rangle_{x_{i_1}} \Big\langle \prod_{1\leq j\leq n} F_{i_2,j}(x_{i_2}, y_j) \Big\rangle_{x_{i_2}} \Big| \\
& \qquad\qquad \prod_{i\neq i_1,i_2} \Big[\prod_{1\leq j\leq n} F_{i,j}(x_i, y_j) \Big]_{x_i} \Big]_{y_1,\ldots,y_n} \\
\leq \ & \frac{1}{2} \Big[\Big\langle \prod_{1\leq j\leq n} F_{i_1,j}(x_{i_1}, y_j) \Big\rangle_{x_{i_1}}^2 \prod_{i\neq i_1,i_2} \Big[\prod_{1\leq j\leq n} F_{i,j}(x_i, y_j) \Big]_{x_i} \Big]_{y_1,\ldots,y_n} \\
& + \frac{1}{2} \Big[\Big\langle \prod_{1\leq j\leq n} F_{i_2,j}(x_{i_2}, y_j) \Big\rangle_{x_{i_2}}^2 \prod_{i\neq i_1,i_2} \Big[\prod_{1\leq j\leq n} F_{i,j}(x_i, y_j) \Big]_{x_i} \Big]_{y_1,\ldots,y_n} \\
= \ & \frac{1}{2} \mathcal{A}^{(\widetilde{\mathbf{p}})} + \frac{1}{2} \mathcal{A}^{(\overline{\mathbf{p}})}
\end{aligned}
$$

Above $\widetilde{\mathbf{p}}$, $\overline{\mathbf{p}} \in \Omega_{m,n}$ are given by their components:

$$
\widetilde{a}_i = \begin{cases} 2, & \text{for } i = i_1, \\ 0, & \text{for } i = i_2, \\ 1, & \text{for } i \neq i_1, i_2, \end{cases} \qquad \overline{a}_i = \begin{cases} 0, & \text{for } i = i_1, \\ 2, & \text{for } i = i_2, \\ 1, & \text{for } i \neq i_1, i_2, \end{cases}
$$

$$
\widetilde{\alpha}_i = \begin{cases} 2, & \text{for } i = i_1, \\ 0, & \text{for } i \neq i_1, \end{cases} \qquad \overline{\alpha}_i = \begin{cases} 2, & \text{for } i = i_2, \\ 0, & \text{for } i \neq i_2, \end{cases}
$$

$$
\widetilde{b}_j = \overline{b}_j = 1,\ \ \widetilde{\beta}_j = \overline{\beta}_j = 0 \quad \text{for every } j\,.
$$

Lemma 4.5, Bound (4.5), and the general Bellman function Estimate (2.8) establish (4.3).

4.2.2 Proof of Proposition 4.2 in full generality

Let us now consider a general triple (E, S, T). By taking $F_{i,j} \equiv \mathbf{1}$ whenever $(i,j) \notin E$ and "completing" each component of the graph, we can assume

$$
E = (\mathcal{X}_1 \times \mathcal{Y}_1) \cup \ldots \cup (\mathcal{X}_k \times \mathcal{Y}_k)\,.
$$

Note that the completed graph still has the same characteristic quantities $d_{i,j}$ and that functions $F_{i,j}$ identically equal to 1 do not contribute to either of the sides in (4.1). The associated form splits as

$$\Lambda_{\mathcal{T}}\big((F_{i,j})_{(i,j)\in E}\big) := \sum_{Q\in\mathcal{T}} |Q| \prod_{l=1}^{k} \big|\mathcal{A}_Q^{(l)}\big((F_{i,j})_{(i,j)\in\mathcal{X}_l\times\mathcal{Y}_l}\big)\big| ,$$

with

$$\mathcal{A}_{I\times J}^{(l)} = \Big[\Big\langle \prod_{\substack{i\in\mathcal{X}_l\\ j\in\mathcal{Y}_l}} F_{i,j}(x_i,y_j)\Big\rangle_{\substack{x_i\in I \text{ for } i\in S\cap\mathcal{X}_l\\ y_j\in J \text{ for } j\in T\cap\mathcal{Y}_l}}\Big]_{\substack{x_i\in I \text{ for } i\in S^c\cap\mathcal{X}_l\\ y_j\in J \text{ for } j\in T^c\cap\mathcal{Y}_l}} .$$

We distinguish two cases with respect to the distribution of selected vertices in the graph. Recall once again that $|S|\geq 2$ or $|T|\geq 2$, which guarantees that the following two cases cover all possibilities, although they are not disjoint.

Case 1. $|S\cap\mathcal{X}_{l_0}|\geq 2$ or $|T\cap\mathcal{Y}_{l_0}|\geq 2$ for some index $l_0\in\{1,\ldots,k\}$.

In words, some connected component contains two selected vertices in one of its bipartition classes. From the previously proven case of Proposition 4.2 applied to the complete bipartite graph with vertex-sets X_{l_0}, Y_{l_0} we have the estimate

$$\sum_{Q\in\mathcal{T}} |Q| \, \big|\mathcal{A}_Q^{(l_0)}\big((F_{i,j})_{(i,j)\in\mathcal{X}_{l_0}\times\mathcal{Y}_{l_0}}\big)\big| \lesssim_{m,n} |Q_{\mathcal{T}}| \prod_{\substack{i\in\mathcal{X}_{l_0}\\ j\in\mathcal{Y}_{l_0}}} \max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})} \big[F_{i,j}^{d^{(l_0)}}\big]_Q^{1/d^{(l_0)}} = 1 .$$

Applying Lemma 4.3 to the functions $(F_{i,j})_{i\in\mathcal{X}_l,\, j\in\mathcal{Y}_l}$ we get

$$\big|\mathcal{A}_Q^{(l)}\big((F_{i,j})_{(i,j)\in\mathcal{X}_l\times\mathcal{Y}_l}\big)\big| \leq \Big[\prod_{\substack{i\in\mathcal{X}_l\\ j\in\mathcal{Y}_l}} F_{i,j}(x_i,y_j)\Big]_{\substack{x_i\in I \text{ for } i\in\mathcal{X}_l\\ y_j\in J \text{ for } j\in\mathcal{Y}_l}} \leq \prod_{\substack{i\in\mathcal{X}_l\\ j\in\mathcal{Y}_l}} \big[F_{i,j}^{d^{(l)}}\big]_Q^{1/d^{(l)}} \leq 1$$

for each $Q = I\times J\in\mathcal{T}$ and each $l\neq l_0$. This establishes (4.3).

Case 2. $S\cap\mathcal{X}_{l_1}\neq\emptyset\neq S\cap\mathcal{X}_{l_2}$ or $T\cap\mathcal{Y}_{l_1}\neq\emptyset\neq T\cap\mathcal{Y}_{l_2}$ for two different indices $l_1, l_2\in\{1,\ldots,k\}$.

In words, there exist two connected components each containing at least one selected x-vertex or each containing at least one selected y-vertex. Without loss of generality suppose $S\cap\mathcal{X}_1\neq 0$ and $S\cap\mathcal{X}_2\neq 0$. Using Lemma 4.3 we can estimate the general term as

$$|\mathcal{A}| \;=\; |\mathcal{A}^{(1)}||\mathcal{A}^{(2)}||\mathcal{A}^{(3)}|\ldots|\mathcal{A}^{(k)}| \;\leq\; |\mathcal{A}^{(1)}||\mathcal{A}^{(2)}| \;\leq\; \frac{1}{2}\big(\mathcal{A}^{(1)}\big)^2+\frac{1}{2}\big(\mathcal{A}^{(2)}\big)^2.$$

By symmetry it is enough to handle $(\mathcal{A}^{(1)})^2$ and by renaming vertices we may assume $\mathcal{X}_1=\{1,\ldots,m_1\}$, $\mathcal{Y}_1=\{1,\ldots,n_1\}$, and $1\in S$.

If $m_1=n_1=1$, then by a simple computation using Theorem 2.2,

$$\begin{aligned}
(\mathcal{A}^{(1)})^2 \;&=\; [\langle F_{1,1}(x_1,y_1)\rangle_{x_1}]_{y_1}^2 \text{ or } \langle F_{1,1}(x_1,y_1)\rangle_{x_1,y_1}^2 \\
&\leq\; [\langle F_{1,1}(x_1,y_1)\rangle_{x_1}]_{y_1}^2+\langle F_{1,1}(x_1,y_1)\rangle_{x_1,y_1}^2+\langle[F_{1,1}(x_1,y_1)]_{x_1}\rangle_{y_1}^2 \\
&=\; \Box\big(\underbrace{[F_{1,1}(x_1,y_1)]_{x_1,y_1}^2}_{\mathcal{B}}\big)\,.
\end{aligned}$$

We separate this special case because now $d^{(1)}=1$ (as for classical paraproducts), so the normalization (4.2) does not control averages of higher powers of $F_{1,1}$. However, the first power is enough here:

$$\max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})}\mathcal{B}_Q(F_{1,1}) \;=\; \Big(\max_{Q\in\mathcal{T}\cup\mathcal{L}(\mathcal{T})}[F_{1,1}]_Q\Big)^2=1\,.$$

On the other hand, the condition $d^{(1)}\geq 2$ is ensured if $m_1=1,\ n_1\geq 2,$ so then we bound $(\mathcal{A}^{(1)})^2$ as

$$\begin{aligned}
(\mathcal{A}^{(1)})^2 \;&\leq\; \Bigg[\bigg|\Big\langle\prod_{1\leq j\leq n_1}F_{1,j}(x_1,y_j)\Big\rangle_{x_1}\bigg|\Bigg]_{y_1,\ldots,y_{n_1}}^2 \\
&\leq\; \Bigg[\Big\langle\prod_{1\leq j\leq n_1}F_{1,j}(x_1,y_j)\Big\rangle_{x_1}^2\Bigg]_{y_1,\ldots,y_{n_1}}.
\end{aligned}$$

We can recognize the last row as $\mathcal{A}^{(\widetilde{\mathbf{p}})}$ for $\widetilde{\mathbf{p}}\in\Omega_{2,n_1}$ given by

$$\begin{aligned}
&\widetilde{a}_1=\widetilde{\alpha}_1=2,\ \ \widetilde{a}_2=\widetilde{\alpha}_2=0,\\
&\widetilde{b}_j=1,\ \widetilde{\beta}_j=0\quad\text{for } 1\leq j\leq n_1\,.
\end{aligned}$$

Thus, $(\mathcal{A}^{(1)})^2\leq\mathcal{A}^{(\widetilde{\mathbf{p}})}\leq\Box\mathcal{B}^{(2,n_1)},$ where $\mathcal{B}^{(2,n_1)}$ is from Lemma 4.5.

If $m_1 \geq 2$, then one can write with the help of the Cauchy-Schwarz inequality and Lemma 4.3 once again:

$$
\begin{aligned}
(\mathcal{A}^{(1)})^2 \ \leq \ & \Big[\Big| \Big\langle \prod_{1 \leq j \leq n_1} F_{1,j}(x_1, y_j) \Big\rangle_{x_1} \Big| \Big[\prod_{1 \leq j \leq n_1} F_{2,j}(x_2, y_j) \Big]_{x_2} \\
& \qquad\qquad \prod_{3 \leq i \leq m_1} \Big[\prod_{1 \leq j \leq n_1} F_{i,j}(x_i, y_j) \Big]_{x_i} \Big]^2_{y_1, \ldots, y_{n_1}} \\
\leq \ & \underbrace{\Big[\Big\langle \prod_{1 \leq j \leq n_1} F_{1,j}(x_1, y_j) \Big\rangle^2_{x_1} \prod_{3 \leq i \leq m_1} \Big[\prod_{1 \leq j \leq n_1} F_{i,j}(x_i, y_j) \Big]_{x_i} \Big]_{y_1, \ldots, y_{n_1}}}_{\mathcal{A}^{(\overline{\mathbf{p}})} \leq \square \mathcal{B}^{(m_1, n_1)}} \\
& \times \underbrace{\Big[\Big[\prod_{1 \leq j \leq n_1} F_{2,j}(x_2, y_j) \Big]^2_{x_2} \prod_{3 \leq i \leq m_1} \Big[\prod_{1 \leq j \leq n_1} F_{i,j}(x_i, y_j) \Big]_{x_i} \Big]_{y_1, \ldots, y_{n_1}}}_{\leq 1},
\end{aligned}
$$

where $\overline{\mathbf{p}} \in \Omega_{m_1, n_1}$ has coordinates

$$
\overline{a}_i = \begin{cases} 2, & \text{for } i = 1, \\ 0, & \text{for } i = 2, \\ 1, & \text{for } 3 \leq i \leq m_1, \end{cases} \qquad \overline{\alpha}_i = \begin{cases} 2, & \text{for } i = 1, \\ 0, & \text{for } 2 \leq i \leq m_1, \end{cases}
$$

$$
\overline{b}_j = 1, \ \overline{\beta}_j = 0 \quad \text{for } 1 \leq j \leq n_1 \, .
$$

and $\mathcal{B}^{(m_1, n_1)}$ is from Lemma 4.5.

In all three possibilities above the proof is finished by invoking the general Estimate (2.8).

4.3 Completing the proof of Theorem 4.1

To establish the global estimate we adapt the approach by Thiele from [49].

Proof of Theorem 4.1. Fix a positive integer N and consider only squares with sidelength at least 2^{-N},

$$
\mathcal{C}_N := \Big\{ Q = I \times J \in \mathcal{C} \, : \, |I| = |J| \geq 2^{-N} \Big\} .
$$

We prove the bound

$$\sum_{Q\in\mathcal{C}_N} |Q|\,\big|\mathcal{A}_Q\big((F_{i,j})_{(i,j)\in E}\big)\big| \;\lesssim_{m,n,(p_{i,j})} \prod_{(i,j)\in E} \|F_{i,j}\|_{\mathrm{L}^{p_{i,j}}(\mathbb{R}^2)}\,, \tag{4.11}$$

with the implicit constant independent of N, so that it implies the result for the whole collection $\mathcal{C}$. Using homogeneity, this time we normalize

$$\|F_{i,j}\|_{\mathrm{L}^{p_{i,j}}(\mathbb{R}^2)} = 1 \quad \text{for every } (i,j)\in E\,.$$

For each $|E|$-tuple of integers $\mathbf{k}=(k_{i,j})_{(i,j)\in E}\in\mathbb{Z}^{|E|}$ we denote

$$\mathcal{P}_{\mathbf{k}} := \Big\{ Q\in\mathcal{C}_N \,:\, 2^{k_{i,j}} \leq \sup_{\substack{Q'\in\mathcal{C}_N\\ Q'\supseteq Q}} \big[F_{i,j}^{d_{i,j}}\big]_{Q'}^{1/d_{i,j}} < \, 2^{k_{i,j}+1} \text{ for every } (i,j)\in E \Big\}.$$

Note that squares in $\mathcal{P}_{\mathbf{k}}$ satisfy $|Q|\leq 2^{-p_{i,j}(k_{i,j}-1)}$ for any $(i,j)\in E$, which limits their size from above. To verify this, we take $Q\in\mathcal{P}_{\mathbf{k}}$ and choose $Q'\supseteq Q$ such that $\big[F_{i,j}^{d_{i,j}}\big]_{Q'}^{1/d_{i,j}} > 2^{k_{i,j}-1}$. By the monotonicity of normalized L^p norms

$$2^{k_{i,j}-1} < \big[F_{i,j}^{d_{i,j}}\big]_{Q'}^{1/d_{i,j}} \leq \big[F_{i,j}^{p_{i,j}}\big]_{Q'}^{1/p_{i,j}} = |Q'|^{-1/p_{i,j}}\|F_{i,j}\|_{\mathrm{L}^{p_{i,j}}(Q')} \leq |Q'|^{-1/p_{i,j}}$$

and thus $|Q|\leq|Q'|\leq 2^{-p_{i,j}(k_{i,j}-1)}$.

Define $\mathcal{M}_{\mathbf{k}}$ to be the collection of maximal squares in $\mathcal{P}_{\mathbf{k}}$ with respect to the set inclusion. For each $Q\in\mathcal{M}_{\mathbf{k}}$ the family

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

is a finite convex[6] tree with root Q and for different squares $Q\in\mathcal{M}_{\mathbf{k}}$ the corresponding trees cover disjoint regions in the plane. For each $\widetilde{Q}\in\mathcal{T}_Q$ by the construction of $\mathcal{P}_{\mathbf{k}}$ we have $\big[F_{i,j}^{d_{i,j}}\big]_{\widetilde{Q}}^{1/d_{i,j}} < 2^{k_{i,j}+1}$. Also, if $\widetilde{Q}\in\mathcal{L}(\mathcal{T}_Q)$ and $\overline{Q}$ is the parent of $\widetilde{Q}$, then

$$\big[F_{i,j}^{d_{i,j}}\big]_{\widetilde{Q}}^{1/d_{i,j}} \leq \; 4\,\big[F_{i,j}^{d_{i,j}}\big]_{\overline{Q}}^{1/d_{i,j}} < 2^{k_{i,j}+3}\,.$$

[6]Convexity of $\mathcal{T}_Q$ follows from monotonicity of $\widetilde{Q}\mapsto\sup_{Q'\in\mathcal{C}_N,\ Q'\supseteq\widetilde{Q}}$.

Therefore, Proposition 4.2 gives

$$\Lambda_{\mathcal{T}_Q}\big((F_{i,j})_{(i,j)\in E}\big) \ \lesssim_{m,n} |Q| \ 2^{\sum_{(i,j)\in E} k_{i,j}} \,.$$

We decompose[7] and estimate,

$$\begin{aligned}\sum_{Q\in\mathcal{C}_N} |Q| \, \big|\mathcal{A}_Q\big((F_{i,j})_{(i,j)\in E}\big)\big| \ &= \sum_{\mathbf{k}\in\mathbb{Z}^{|E|}} \sum_{Q\in\mathcal{M}_{\mathbf{k}}} \Lambda_{\mathcal{T}_Q}\big((F_{i,j})_{(i,j)\in E}\big) \\ &\lesssim_{m,n} \sum_{\mathbf{k}\in\mathbb{Z}^{|E|}} 2^{\sum_{(i,j)\in E} k_{i,j}} \Big(\sum_{Q\in\mathcal{M}_{\mathbf{k}}} |Q| \Big) \,. \end{aligned} \tag{4.12}$$

For any $d \geq 1$ it is a classical result[8] that "power d variant" of the *dyadic maximal function*

$$(\mathrm{M}_d F)(x,y) := \sup_{\substack{Q\in\mathcal{C} \\ \text{s.t. } (x,y)\in Q}} \big[|F|^d\big]_Q^{1/d}$$

is bounded on $\mathrm{L}^p(\mathbb{R}^2)$ whenever $d < p \leq \infty$. We have $\mathbb{Z}^{|E|} = \bigcup_{(i_0,j_0)\in E} \mathcal{K}_{(i_0,j_0)}$, where the subsets $\mathcal{K}_{(i_0,j_0)}$ are defined by

$$\mathcal{K}_{(i_0,j_0)} := \big\{ \mathbf{k} = (k_{i,j})_{(i,j)\in E} \ : \ p_{i_0,j_0} k_{i_0,j_0} \geq p_{i,j} k_{i,j} \text{ for every } (i,j)\in E \big\} \,.$$

Observe that for $(x,y)\in Q\in\mathcal{P}_{\mathbf{k}}$ we have by the definition of $\mathcal{P}_{\mathbf{k}}$,

$$(\mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0})(x,y) \geq \sup_{\substack{Q'\supseteq Q \\ Q'\in\mathcal{C}_N}} \big[F_{i_0,j_0}^{d_{i_0,j_0}}\big]_{Q'}^{1/d_{i_0,j_0}} \geq 2^{k_{i_0,j_0}} \,,$$

so

$$\bigcup_{Q\in\mathcal{P}_{\mathbf{k}}} Q \ \subseteq \ \big\{ (x,y)\in\mathbb{R}^2 : (\mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0})(x,y) \geq 2^{k_{i_0,j_0}} \big\} \,.$$

This together with disjointness of $\mathcal{M}_{\mathbf{k}}$ gives

$$\sum_{Q\in\mathcal{M}_{\mathbf{k}}} |Q| = \Big| \bigcup_{Q\in\mathcal{M}_{\mathbf{k}}} Q \Big| = \Big| \bigcup_{Q\in\mathcal{P}_{\mathbf{k}}} Q \Big| \leq \big| \big\{ \mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0} \geq 2^{k_{i_0,j_0}} \big\} \big| \tag{4.13}$$

[7]Here we use that for each $\widetilde{Q} \in \mathcal{C}_N \setminus \bigcup_{\mathbf{k}\in\mathbb{Z}^{|E|}} \mathcal{P}_{\mathbf{k}}$ at least one of the functions constantly vanishes on $\widetilde{Q}$, so the corresponding summand is equal to 0.

[8]One simply writes $\mathrm{M}_d F = \big(\mathrm{M}_1 |F|^d\big)^{1/d}$, where M_1 is the standard dyadic maximal function.

for any choice of $(i_0, j_0) \in E$.

Combining (4.12) and (4.13) allows the final computation:

$$\begin{aligned}
&\sum_{Q\in\mathcal{C}_N} |Q| \, \big|\mathcal{A}_Q\big((F_{i,j})_{(i,j)\in E}\big)\big| \\
&\lesssim_{m,n} \sum_{(i_0,j_0)\in E} \sum_{\mathbf{k}\in\mathcal{K}_{(i_0,j_0)}} 2^{\sum_{(i,j)\in E} k_{i,j}} \, \big|\big\{\mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0} \geq 2^{k_{i_0,j_0}}\big\}\big| \\
&= \sum_{(i_0,j_0)\in E} \sum_{k_{i_0,j_0}\in\mathbb{Z}} 2^{p_{i_0,j_0} k_{i_0,j_0}} \, \big|\big\{\mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0} \geq 2^{k_{i_0,j_0}}\big\}\big| \\
&\qquad\qquad \prod_{\substack{(i,j)\in E \\ (i,j)\neq(i_0,j_0)}} \sum_{\substack{k_{i,j}\in\mathbb{Z} \\ k_{i,j} \leq p_{i_0,j_0} k_{i_0,j_0}/p_{i,j}}} 2^{\,k_{i,j} - p_{i_0,j_0} k_{i_0,j_0}/p_{i,j}} \\
&\lesssim_{|E|} \sum_{(i_0,j_0)\in E} \sum_{k_{i_0,j_0}\in\mathbb{Z}} 2^{p_{i_0,j_0} k_{i_0,j_0}} \, \big|\big\{\mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0} \geq 2^{k_{i_0,j_0}}\big\}\big| \\
&\lesssim_{(p_{i,j})} \sum_{(i_0,j_0)\in E} \|\mathrm{M}_{d_{i_0,j_0}} F_{i_0,j_0}\|_{\mathrm{L}^{p_{i_0,j_0}}(\mathbb{R}^2)}^{p_{i_0,j_0}} \\
&\lesssim_{(d_{i,j}),\,(p_{i,j})} \sum_{(i_0,j_0)\in E} \|F_{i_0,j_0}\|_{\mathrm{L}^{p_{i_0,j_0}}(\mathbb{R}^2)}^{p_{i_0,j_0}} \;\lesssim_{|E|}\; 1\,,
\end{aligned}$$

which is exactly (4.11). We used $\sum_{(i,j)\in E} \frac{1}{p_{i,j}} = 1$ and added up $|E|-1$ geometric series with initial terms in $(\frac{1}{2}, 1]$ and common ratios equal to $\frac{1}{2}$. $\square$

4.4 An example that illustrates the proof

In this short section we a take the example from Figure 4.2 and show how steps from Section 4.2 should be performed in a concrete case. These steps gradually reduce the general term of Λ to the "simpler" ones and the procedure is presented as recursive rather than inductive, i.e. we only care about paraproduct-type terms relevant for dominating Λ. We keep assuming $F_{i,j} \geq 0$ and Normalization (4.2).

In order to deal with more symmetric situation, we "complete" the graph by introducing $F_{2,3} \equiv \mathbf{1}$ and $F_{3,5} \equiv \mathbf{1}$. Since $|T| = 2$, we decide to rewrite

$$\begin{aligned}
\mathcal{A}^{(1)} &= \Big[\big\langle F_{1,1}(x_1, y_1)F_{2,1}(x_2, y_1)\big\rangle_{y_1} \big[F_{1,2}(x_1, y_2)F_{2,2}(x_2, y_2)\big]_{y_2} \\
&\qquad\qquad\qquad\qquad \big[F_{1,3}(x_1, y_3)F_{2,3}(x_2, y_3)\big]_{y_3}\Big]_{x_1,x_2}, \\
\mathcal{A}^{(2)} &= \Big[\big\langle F_{3,4}(x_3, y_4)F_{4,4}(x_4, y_4)\big\rangle_{y_4} \big[F_{3,5}(x_3, y_5)F_{4,5}(x_4, y_5)\big]_{y_5}\Big]_{x_3,x_4}
\end{aligned}$$

and start by estimating the general term as in Subsection 4.2.2, Case 2,

$$\begin{aligned}
|\mathcal{A}| = |\mathcal{A}^{(1)}|\,|\mathcal{A}^{(2)}|\,|\mathcal{A}^{(3)}| &\le \frac{1}{2}\big(\mathcal{A}^{(1)}\big)^2 + \frac{1}{2}\big(\mathcal{A}^{(2)}\big)^2 \\
&\le \frac{1}{2} \underbrace{\Big[\big\langle F_{1,1}(x_1, y_1)F_{2,1}(x_2, y_1)\big\rangle_{y_1}^2 \big[F_{1,3}(x_1, y_3)F_{2,3}(x_2, y_3)\big]_{y_3}\Big]_{x_1,x_2}}_{\mathcal{A}^{(4)}} \\
&\qquad \times \underbrace{\Big[\big[F_{1,2}(x_1, y_2)F_{2,2}(x_2, y_2)\big]_{y_2}^2 \big[F_{1,3}(x_1, y_3)F_{2,3}(x_2, y_3)\big]_{y_3}\Big]_{x_1,x_2}}_{\le 1} \\
&\quad + \frac{1}{2} \underbrace{\Big[\big\langle F_{3,4}(x_3, y_4)F_{4,4}(x_4, y_4)\big\rangle_{y_4}^2\Big]_{x_3,x_4}}_{\mathcal{A}^{(5)}} \\
&\qquad \times \underbrace{\Big[\big[F_{3,5}(x_3, y_5)F_{4,5}(x_4, y_5)\big]_{y_5}^2\Big]_{x_3,x_4}}_{\le 1}.
\end{aligned}$$

Let us proceed with $\mathcal{A}^{(4)}$ since $\mathcal{A}^{(5)}$ was already handled in Section 2.3. Following the approach from the proof of Lemma 4.4, Case 2, we define

$$\begin{aligned}
\mathcal{B}^{(4)} &= \Big[\big[F_{1,1}(x_1, y_1)F_{2,1}(x_2, y_1)\big]_{y_1}^2 \big[F_{1,3}(x_1, y_3)F_{2,3}(x_2, y_3)\big]_{y_3}\Big]_{x_1,x_2} \\
&= \Big[\big[F_{1,1}(x_1, y_1)F_{1,1}(x_1, y_1')F_{1,3}(x_1, y_3)\big]_{x_1} \\
&\qquad \big[F_{2,1}(x_2, y_1)F_{2,1}(x_2, y_1')F_{2,3}(x_2, y_3)\big]_{x_2}\Big]_{y_1,y_1',y_3}.
\end{aligned}$$

By Theorem 2.2 every term in $\Box\mathcal{B}^{(4)}$ is either $\mathcal{A}^{(4)}$, or

$$\begin{aligned}
\mathcal{A}^{(6)} &= \Big[\big\langle F_{1,1}(x_1, y_1)F_{2,1}(x_2, y_1)\big\rangle_{y_1} \big[F_{1,1}(x_1, y_1)F_{2,1}(x_2, y_1)\big]_{y_1} \\
&\qquad\qquad\qquad\qquad \big\langle F_{1,3}(x_1, y_3)F_{2,3}(x_2, y_3)\big\rangle_{y_3}\Big]_{x_1,x_2},
\end{aligned}$$

or is of the form

$$\mathcal{A}^{(7)} = \Big(\big\langle F_{1,1}(x_1, y_1) F_{1,1}(x_1, y_1') F_{1,3}(x_1, y_3) \big\rangle_{x_1} \big\langle F_{2,1}(x_2, y_1) F_{2,1}(x_2, y_1') F_{2,3}(x_2, y_3) \big\rangle_{x_2} \Big)_{y_1, y_1', y_3},$$

where parentheses $(\cdot)$ are understood as either $[\cdot]$ or $\langle \cdot \rangle$ for each of the variables y_1, y_1', y_3 independently. We dominate $\mathcal{A}^{(6)}$, $\mathcal{A}^{(7)}$ as in the proof of Lemma 4.4, Case 1.

$$|\mathcal{A}^{(6)}| \leq \frac{1}{2\delta} \underbrace{\Big[\big\langle F_{1,1}(x_1, y_1) F_{2,1}(x_2, y_1) \big\rangle_{y_1}^2 \big[F_{1,1}(x_1, y_1) F_{2,1}(x_2, y_1) \big]_{y_1} \Big]_{x_1, x_2}}_{\mathcal{A}^{(8)}}$$

$$+ \frac{\delta}{2} \underbrace{\Big[\big\langle F_{1,3}(x_1, y_3) F_{2,3}(x_2, y_3) \big\rangle_{y_3}^2 \big[F_{1,1}(x_1, y_1) F_{2,1}(x_2, y_1) \big]_{y_1} \Big]_{x_1, x_2}}_{\mathcal{A}^{(9)}}$$

$$|\mathcal{A}^{(7)}| \leq \frac{1}{2} \underbrace{\Big[\big\langle F_{1,1}(x_1, y_1) F_{1,1}(x_1, y_1') F_{1,3}(x_1, y_3) \big\rangle_{x_1}^2 \Big]_{y_1, y_1', y_3}}_{\mathcal{A}^{(10)}}$$

$$+ \frac{1}{2} \underbrace{\Big[\big\langle F_{2,1}(x_2, y_1) F_{2,1}(x_2, y_1') F_{2,3}(x_2, y_3) \big\rangle_{x_2}^2 \Big]_{y_1, y_1', y_3}}_{\mathcal{A}^{(11)}}$$

Since $\mathcal{A}^{(9)}$ is analogous to $\mathcal{A}^{(4)}$, the algorithm loops (with a small "weight" δ) and we continue with $\mathcal{A}^{(8)}$ only. Also, by symmetry, we can consider $\mathcal{A}^{(10)}$ and disregard $\mathcal{A}^{(11)}$.

To control $\mathcal{A}^{(8)}$ we define

$$\begin{aligned} \mathcal{B}^{(8)} &= \Big[\big[F_{1,1}(x_1, y_1) F_{2,1}(x_2, y_1) \big]_{y_1}^3 \Big]_{x_1, x_2} \\ &= \Big[\big[F_{1,1}(x_1, y_1) F_{1,1}(x_1, y_1') F_{1,1}(x_1, y_1'') \big]_{x_1} \\ &\qquad \big[F_{2,1}(x_2, y_1) F_{2,1}(x_2, y_1') F_{2,1}(x_2, y_1'') \big]_{x_2} \Big]_{y_1, y_1', y_1''}. \end{aligned}$$

Every term in $\Box\mathcal{B}^{(8)}$ different from $\mathcal{A}^{(8)}$ is of the shape

$$\mathcal{A}^{(12)} = \Big(\big\langle F_{1,1}(x_1, y_1) F_{1,1}(x_1, y_1') F_{1,1}(x_1, y_1'') \big\rangle_{x_1} \big\langle F_{2,1}(x_2, y_1) F_{2,1}(x_2, y_1') F_{2,1}(x_2, y_1'') \big\rangle_{x_2} \Big)_{y_1, y_1', y_1''}$$

and is estimated as

$$
\begin{aligned}
|\mathcal{A}^{(12)}| \leq\ & \underbrace{\frac{1}{2}\Big[\big\langle F_{1,1}(x_1,y_1)F_{1,1}(x_1,y_1')F_{1,1}(x_1,y_1'')\big\rangle_{x_1}^2\Big]_{y_1,y_1',y_1''}}_{\mathcal{A}^{(13)}} \\
&+ \frac{1}{2}\Big[\big\langle F_{2,1}(x_2,y_1)F_{2,1}(x_2,y_1')F_{2,1}(x_2,y_1'')\big\rangle_{x_2}^2\Big]_{y_1,y_1',y_1''}.
\end{aligned}
$$

Finally, $\mathcal{A}^{(13)}$ corresponds to a partition with the smallest rank. We introduce

$$
\begin{aligned}
\mathcal{B}^{(13)} &= \Big[\big[F_{1,1}(x_1,y_1)F_{1,1}(x_1,y_1')F_{1,1}(x_1,y_1'')\big]_{x_1}^2\Big]_{y_1,y_1',y_1''} \\
&= \Big[\big[F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big]_{y_1}^3\Big]_{x_1,x_1'}
\end{aligned}
$$

and observe that by Theorem 2.2 each term in $\Box\mathcal{B}^{(13)}$ different from $\mathcal{A}^{(13)}$ is of the form

$$
\Big(\big\langle F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big\rangle_{y_1}^2\big[F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big]_{y_1}\Big)_{x_1,x_1'}.
$$

Sum of these terms is nonnegative by Lemma 2.3 and can be discarded as it only increases $\Box\mathcal{B}^{(13)}$.

On the other hand, to deal with $\mathcal{A}^{(10)}$ we define

$$
\begin{aligned}
\mathcal{B}^{(10)} &= \Big[\big[F_{1,1}(x_1,y_1)F_{1,1}(x_1,y_1')F_{1,3}(x_1,y_3)\big]_{x_1}^2\Big]_{y_1,y_1',y_3} \\
&= \Big[\big[F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big]_{y_1}^2\big[F_{1,3}(x_1,y_3)F_{1,3}(x_1',y_3)\big]_{y_3}\Big]_{x_1,x_1'}.
\end{aligned}
$$

Every term in $\Box\mathcal{B}^{(10)}$ other than $\mathcal{A}^{(10)}$ takes either the shape

$$
\Big(\big\langle F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big\rangle_{y_1}^2\big[F_{1,3}(x_1,y_3)F_{1,3}(x_1',y_3)\big]_{y_3}\Big)_{x_1,x_1'},
$$

or the shape

$$
\begin{aligned}
\Big(\big\langle F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big\rangle_{y_1}&\big[F_{1,1}(x_1,y_1)F_{1,1}(x_1',y_1)\big]_{y_1} \\
&\big\langle F_{1,3}(x_1,y_3)F_{1,3}(x_1',y_3)\big\rangle_{y_3}\Big)_{x_1,x_1'}.
\end{aligned}
$$

Sum of the former terms is nonnegative by Lemma 2.3, while the latter ones are treated similarly as $\mathcal{A}^{(6)}$.

A concrete Bellman function $\mathcal{B}$ satisfying $|\mathcal{A}| \leq \Box\mathcal{B}$ and (4.5) under Normalization (4.2) can be assembled from all averaging paraproduct-type terms that appear in the proof, including the omitted ones. Estimate (2.8) gives (4.3) once again.

We close this chapter by a comment on the setup from Section 4.1. Even though it is possible to associate multilinear forms to general undirected graphs, the machinery from Chapter 2 cannot be applied to forms arising from graphs that are not bipartite. For instance, a multilinear form associated to a triangle, i.e. a cycle of length 3, could be

$$\widetilde{\Lambda}(F,G,H) := \sum_{\substack{I,J,K\in\mathcal{D} \\ |I|=|J|=|K| \\ c(I,J,K)=0}} |I|^{1/2} \int_{\mathbb{R}^3} F(x,y)G(y,z)H(z,x)\, \psi_I^{\mathrm{d}}(x)\psi_J^{\mathrm{d}}(y)\psi_K^{\mathrm{d}}(z)\, dx\, dy\, dz\,,$$

where $c(I,J,K)=0$ is some constraint, making the sum effectively indexed by only two of the intervals. Such forms seem to share many characteristics with "singular bilinear averages" (1.2), for which no L^p bounds are known at the time of writing.

CHAPTER 5

Uniform constants in Hausdorff-Young inequalities for the Cantor group model of the scattering transform

5.1 Statement of the main result

Fix an integer $d \geq 2$ and denote $\mathbb{Z}_d := \mathbb{Z}/d\mathbb{Z}$. For any $x, \xi \in [0, \infty)$ that can be written uniquely[1] in base d number system as $x = \sum_{n\in\mathbb{Z}} x_n d^n$ and $\xi = \sum_{n\in\mathbb{Z}} \xi_n d^n$, we define

$$E_d(x, \xi) := e^{(2\pi i/d) \sum_{n\in\mathbb{Z}} x_n \xi_{-1-n}} .$$

Then the L^∞ function $E_d \colon [0, \infty) \times [0, \infty) \to \mathrm{S}^1$ is called the *Cantor group character function*. To justify the name, we identify $[0, \infty)$ with a subgroup $\mathbb{A}_d$ of the infinite group product $\mathbb{Z}_d^{\mathbb{Z}}$ given by

$$\begin{aligned} \mathbb{A}_d := \big\{ (x_n)_{n\in\mathbb{Z}} \ : \ & x_n \in \mathbb{Z}_d \text{ for every } n \in \mathbb{Z}, \text{ and there exists} \\ & n_0 \in \mathbb{Z} \text{ such that } x_n = \mathbf{0} \text{ for every } n \geq n_0 \big\} , \end{aligned}$$

via the identification $\mathbb{A}_d \to [0, \infty)$, $(x_n)_{n\in\mathbb{Z}} \mapsto \sum_{n\in\mathbb{Z}} x_n d^n$. Then $E_d(\cdot, \cdot)$ realizes duality between $\mathbb{A}_d$ and its dual group $\hat{\mathbb{A}}_d \cong \mathbb{A}_d$.

[1]Because of ambiguous base d representation of some reals, the function E_d is not well-defined on a set of measure zero. The same comment applies to the later identification of $\mathbb{A}_d$ with $[0, \infty)$.

For a compactly supported integrable function $f\colon [0,\infty) \to \mathbb{C}$ and $\xi \in [0,\infty)$ consider the initial value problem on $[0,\infty)$:

$$\frac{\partial}{\partial x} G(x,\xi) = G(x,\xi) W(x,\xi), \qquad G(0,\xi) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where

$$G(x,\xi) = \begin{bmatrix} a(x,\xi) & \overline{b(x,\xi)} \\ b(x,\xi) & \overline{a(x,\xi)} \end{bmatrix}, \qquad W(x,\xi) = \begin{bmatrix} 0 & \overline{f(x)\, E_d(x,\xi)} \\ f(x)\, E_d(x,\xi) & 0 \end{bmatrix}.$$

The limit

$$G(\xi) = \begin{bmatrix} a(\xi) & \overline{b(\xi)} \\ b(\xi) & \overline{a(\xi)} \end{bmatrix} := \lim_{x\to\infty} \begin{bmatrix} a(x,\xi) & \overline{b(x,\xi)} \\ b(x,\xi) & \overline{a(x,\xi)} \end{bmatrix}$$

defines a function $\xi \mapsto G(\xi)$ from $[0,\infty)$ to $\mathrm{SU}(1,1)$, which we call the *Cantor group model Dirac scattering transform* of f. Dependence on d is not notationally emphasized but is understood. If for some interval $I \subseteq [0,\infty)$ we replace f by $f\mathbf{1}_I$, then we will denote the corresponding G, a, b respectively by G_I, a_I, b_I.

Our main result on this topic is the following theorem.

Theorem 5.1. *For every integer $d \geq 2$ there exists a constant $C_d > 0$ such that for any pair of conjugated exponents $1 \leq p \leq 2$ and $2 \leq q \leq \infty$ and every compactly supported integrable function f one has*

$$\|(\ln|a(\xi)|)^{1/2}\|_{\mathrm{L}^q_\xi(\mathbb{R})} \leq C_d \|f\|_{\mathrm{L}^p(\mathbb{R})}\,.$$

In the following exposition we need a couple of simple facts proven in [33].

Lemma 5.2 (from [33])**.** *If I and ω are two d-adic intervals*[2] *with $|I||\omega| = 1$, then $\xi \mapsto |a_I(\xi)|$ and $\xi \mapsto |b_I(\xi)|$ are constant functions on ω.*

[2]These are intervals of the form $\big[d^k l, d^k(l+1)\big)$ for some $k, l \in \mathbb{Z}$, $l \geq 0$.

We will be working in the phase space $\mathbb{A}_d \times \hat{\mathbb{A}}_d$, which is identified with $[0, \infty) \times [0, \infty)$. *Tiles* and *multitiles* are rectangles of the form $I \times \omega$ for two d-adic intervals I, ω satisfying $|I||\omega| = 1$ and $|I||\omega| = d$ respectively. Every multitile $I \times \omega$ can be partitioned into d tiles by subdividing either I or ω into d congruent d-adic intervals. Lemma 5.2 motivates us to define G_P, a_P, b_P for any tile $P = I \times \omega$ simply as $G_I(\xi_\omega)$, $a_I(\xi_\omega)$, $b_I(\xi_\omega)$, where ξ_ω is the left endpoint of ω.

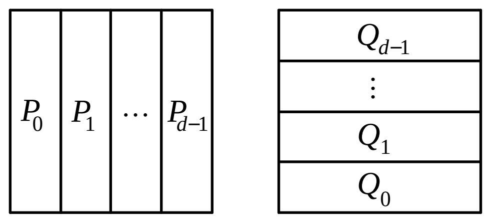

Figure 5.1: A multitile partitioned in two ways.

Lemma 5.3 (from [33]). *Suppose that a multitile is divided horizontally into tiles $P_0, \ldots, P_{d-1}$ and vertically into tiles $Q_0, \ldots, Q_{d-1}$, as in Figure 5.1. Then*

$$\begin{bmatrix} a_{Q_k} & \overline{b_{Q_k}} \\ b_{Q_k} & \overline{a_{Q_k}} \end{bmatrix} = \prod_{j=0}^{d-1} \begin{bmatrix} a_{P_j} & \overline{b_{P_j}}\, e^{-2\pi ijk/d} \\ b_{P_j}\, e^{2\pi ijk/d} & \overline{a_{P_j}} \end{bmatrix}$$

for $k = 0, 1, \ldots, d-1$. (The matrix product has to be taken in ascending order.)

5.2 Proof of Theorem 5.1

This section proves the main theorem assuming that the following proposition holds. This way, the main technical construction is postponed until the next section.

Proposition 5.4. *There exist a constant $C_d > 0$ and a function $\beta_d\colon [0,\infty) \to [0,\infty)$ such that for every* $\left[\begin{smallmatrix} a & \bar{b} \\ b & \bar{a} \end{smallmatrix}\right] \in \mathrm{SU}(1,1)$

$$C_d^{-1}(\ln|a|)^{1/2} \,\leq\, \beta_d(|b|) \,\leq\, C_d(\ln|a|)^{1/2}\,, \tag{5.1}$$

and whenever matrices $\left[\begin{smallmatrix} a_j & \overline{b_j} \\ b_j & \overline{a_j} \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} A_k & \overline{B_k} \\ B_k & \overline{A_k} \end{smallmatrix}\right] \in \mathrm{SU}(1,1)$, $j,k = 0,1,\ldots,d-1$ *satisfy*

$$\begin{bmatrix} A_k & \overline{B_k} \\ B_k & \overline{A_k} \end{bmatrix} = \prod_{j=0}^{d-1} \begin{bmatrix} a_j & \overline{b_j}\, e^{-2\pi ijk/d} \\ b_j\, e^{2\pi ijk/d} & \overline{a_j} \end{bmatrix}, \tag{5.2}$$

then for any pair of conjugated exponents $1 < p \leq 2$ *and* $2 \leq q < \infty$ *one has*

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1} \beta_d(|B_k|)^q\Big)^{\frac{1}{q}} \leq \Big(\sum_{j=0}^{d-1} \beta_d(|b_j|)^p\Big)^{\frac{1}{p}}. \tag{5.3}$$

This proposition is proven in the next section, by giving an explicit construction of β_d. The construction might seem a bit tedious, but we have to satisfy (5.3) with the exact constant at most 1, since we will be repeatedly applying that inequality in the proof of Theorem 5.1. Iterating an inequality with a constant $C > 1$ would not yield an estimate independent of the number of scales.

A consequence of Lemma 5.3 and (5.3) is that for $P_0,\ldots,P_{d-1},Q_0,\ldots,Q_{d-1}$ as above we get

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1} \beta_d(|b_{Q_k}|)^q\Big)^{\frac{1}{q}} \leq \Big(\sum_{j=0}^{d-1} \beta_d(|b_{P_j}|)^p\Big)^{\frac{1}{p}}. \tag{5.4}$$

Proof of Theorem 5.1 assuming Proposition 5.4. We can consider $1 < p \leq 2$, as for $p = 1$ the estimate is an immediate consequence of Gronwall's inequality. Fix a positive integer N (large enough) so that f is supported in $[0, d^N)$. In all of the following we consider only those tiles $I \times \omega$ that are subsets of $[0, d^N) \times [0, d^N)$. For any $n \in \mathbb{Z}$, $-N \leq n \leq N$ consider the following quantity:

$$\mathcal{B}_n := \Bigg(\sum_{|I|=d^n} \Big(d^{-n}\sum_{|\omega|=d^{-n}} \beta_d(|b_{I\times\omega}|)^q\Big)^{\frac{p}{q}}\Bigg)^{\frac{1}{p}}.$$

In words, we consider all tiles P of type $d^n \times d^{-n}$, then we take normalized ℓ^q-norm of the numbers $\beta_d(|b_P|)$ for all tiles in the same column, and finally we take ℓ^p-norm of those numbers over all columns.

The quantity $\mathcal{B}$ is our "Bellman function", but in this case an appropriate name would be a *monovariant* (a common term in combinatorial game theory), because we only need its monotonicity over scales and do not require that it dominates any extra terms. Let us prove that $\mathcal{B}_n$ is decreasing in n.

$$\begin{aligned}
\mathcal{B}_{n+1}^p &= \sum_{|I|=d^{n+1}} \Big(d^{-n} \sum_{|\omega|=d^{-n}} d^{-1} \sum_{\substack{\omega'\subseteq\omega \\ |\omega'|=d^{-n-1}}} \beta_d(|b_{I\times\omega'}|)^q\Big)^{\frac{p}{q}} \\
&\quad \text{using (5.4) for the multitile } I\times\omega \\
&\leq \sum_{|I|=d^{n+1}} \bigg(d^{-n} \sum_{|\omega|=d^{-n}} \Big(\sum_{\substack{I'\subseteq I \\ |I'|=d^n}} \beta_d(|b_{I'\times\omega}|)^p\Big)^{\frac{q}{p}}\bigg)^{\frac{p}{q}} \\
&\quad \text{using Minkowski's inequality, since } q/p \geq 1 \\
&\leq \sum_{|I|=d^{n+1}} \sum_{\substack{I'\subseteq I \\ |I'|=d^n}} \Big(d^{-n} \sum_{|\omega|=d^{-n}} \beta_d(|b_{I'\times\omega}|)^q\Big)^{\frac{p}{q}} = \mathcal{B}_n^p
\end{aligned}$$

Furthermore, when $n = -N$ we have:

$$\begin{aligned}
\mathcal{B}_{-N} = \Big(\sum_{|I|=d^{-N}} (d^N)^{\frac{p}{q}}\ \beta_d(|b_{I\times[0,d^N)}|)^p\Big)^{\frac{1}{p}} &\leq C_d (d^N)^{\frac{1}{q}} \Big(\sum_{|I|=d^{-N}} (\ln|a_I(0)|)^{\frac{p}{2}}\Big)^{\frac{1}{p}} \\
&\leq C_d (d^N)^{\frac{1}{q}} \Big(\sum_{|I|=d^{-N}} \|f\mathbf{1}_I\|_{\mathrm{L}^1}^p\Big)^{\frac{1}{p}} \leq C_d \|f\|_{\mathrm{L}^p} .
\end{aligned}$$

Here we have applied the trivial L^1–L^∞ estimate, a.k.a. the nonlinear Riemann-Lebesgue estimate, a consequence of Gronwall's lemma, $(\ln|a_I(0)|)^{1/2} \leq \|f\mathbf{1}_I\|_{\mathrm{L}^1}$ and Hölder's inequality $\|f\mathbf{1}_I\|_{\mathrm{L}^1} \leq \|f\mathbf{1}_I\|_{\mathrm{L}^p}\|\mathbf{1}_I\|_{\mathrm{L}^q}$. On the other hand, for $n = N$

we have:

$$\mathcal{B}_N = \Big(d^{-N} \sum_{|\omega|=d^{-N}} \beta_d(|b_{[0,d^N)\times\omega}|)^q\Big)^{\frac{1}{q}} \geq C_d^{-1}\Big(d^{-N} \sum_{|\omega|=d^{-N}} (\ln|a(\xi_\omega)|)^{\frac{q}{2}}\Big)^{\frac{1}{q}}$$
$$= C_d^{-1}\Big(\int_0^{d^N} (\ln|a(\xi)|)^{\frac{q}{2}}\, d\xi\Big)^{\frac{1}{q}}.$$

Above ξ_ω denotes the left endpoint of ω and we have used that $\xi \mapsto |a(\xi)|$ is constant on intervals of length d^{-N}, by Lemma 5.2. From the monotonicity of $(\mathcal{B}_n)$ we conclude:

$$\Big(\int_0^{d^N} (\ln|a(\xi)|)^{\frac{q}{2}}\, d\xi\Big)^{\frac{1}{q}} \leq C_d\mathcal{B}_N \leq C_d\mathcal{B}_{-N} \leq C_d^2\|f\|_{\mathrm{L}^p},$$

and by taking $\lim_{N\to\infty}$ we deduce the theorem. □

5.3 The swapping inequality

This technical section is devoted to the proof of Proposition 5.4. An arbitrary function on $\mathbb{Z}_d$ can be presented as a complex d-tuple $(z_0, z_1, \ldots, z_{d-1})$. Its Fourier transform is the d-tuple $(Z_0, Z_1, \ldots, Z_{d-1})$ given by

$$Z_k := \sum_{j=0}^{d-1} z_j\, e^{2\pi ijk/d}.$$

Lemma 5.5. *For a pair of conjugated exponents* $1 < p \leq 2$ *and* $2 \leq q < \infty$ *and* (z_j), (Z_k) *as above, one has*

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1} |Z_k|^q\Big)^{\frac{1}{q}} \leq \Big(\sum_{j=0}^{d-1} |z_j|^p\Big)^{\frac{1}{p}}.$$

Lemma 5.5 is a particular consequence of the general theory of the Fourier transform on locally compact abelian groups (see [15]). Indeed, one observes that the (non-stated) case $p = 1$ is trivial from the triangle inequality, while for $p = 2$

we indeed have an equality that follows from orthonormality of group characters. Intermediate cases are deduced by interpolating these two endpoint ones using the Riesz-Thorin theorem, since the transformation $(z_j) \mapsto (Z_k)$ is linear.

For any integer $d \geq 2$ let t_d be the unique solution of the equation

$$te^{-t} = (2d)^{-5}\sqrt{1 + \operatorname{arsinh} t}$$

that lies in $[0, 1]$. One can easily see

$$2^{-5}d^{-5} < t_d < 2^{-4}d^{-5}\,, \tag{5.5}$$

and indeed $t_d = (2d)^{-5} + \frac{3}{2}(2d)^{-10} + \mathrm{O}(d^{-15})$ as $d \to \infty$, but we do not need bounds on t_d that are more precise than (5.5).

Now we define $\beta_d \colon [0, \infty) \to [0, \infty)$ by the formula

$$\beta_d(t) := \begin{cases} te^{-t}, & \text{for } t \leq t_d\,, \\ (2d)^{-5}\sqrt{1 + \operatorname{arsinh} t}, & \text{for } t > t_d\,. \end{cases}$$

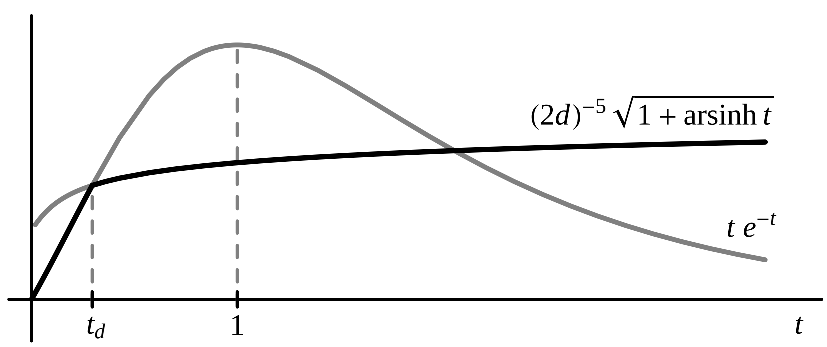

Figure 5.2: Graph of β_d. Not drawn to scale.

Using only basic calculus, one can easily establish the following properties of β_d:

$$2^{-6}d^{-5}\sqrt{\ln(1+t^2)} \;\leq\; \beta_d(t) \;\leq\; 2\sqrt{\ln(1+t^2)}\,, \tag{5.6}$$

$$\beta_d(t) \;\leq\; te^{-t}, \qquad \text{for } 0 \leq t \leq 1\,, \tag{5.7}$$

$$\beta_d(t) \;\leq\; (2d)^{-5}\sqrt{1+\operatorname{arsinh} t}, \qquad \text{for any } t \geq 0\,. \tag{5.8}$$

Since (5.6) is exactly (5.1), it is enough to verify (5.3).

By performing matrix multiplication in (5.2), one can write B_k explicitly as a sum of 2^{d-1} terms of the form

$$\overline{a_0}\ldots\overline{a_{j_1-1}}b_{j_1}a_{j_1+1}\ldots a_{j_2-1}\overline{b_{j_2}}\,\overline{a_{j_2+1}}\ldots\overline{a_{j_3-1}}b_{j_3}a_{j_3+1}\ldots$$
$$\ldots a_{j_{2r}-1}\overline{b_{j_{2r}}}\,\overline{a_{j_{2r}+1}}\ldots\overline{a_{j_{2r+1}-1}}b_{j_{2r+1}}a_{j_{2r+1}+1}\ldots a_{d-1}\cdot e^{(2\pi ik/d)(j_1-j_2+j_3-\ldots-j_{2r}+j_{2r+1})}$$

where the summation is taken over all integers $0 \leq r \leq \lfloor\frac{d-1}{2}\rfloor$ and over all possible choices of indices $0 \leq j_1 < j_2 < \ldots < j_{2r+1} \leq d-1$. In particular, observe that each term contains an odd number of b's. Terms that contain exactly one of the b's could be called *linear terms*, and so the "linear part" of B_k is

$$B'_k := \sum_{j=0}^{d-1}\overline{a_0a_1\ldots a_{j-1}}b_ja_{j+1}\ldots a_{d-1}\,e^{2\pi ijk/d}, \quad \text{for } k=0,1,\ldots,d-1.$$

Other terms in B_k are called *nonlinear terms*. Observe that Lemma 5.5 gives

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B'_k|^q\Big)^{\frac{1}{q}} \leq \Big(\sum_{j=0}^{d-1}|b'_j|^p\Big)^{\frac{1}{p}}\,, \tag{5.9}$$

where $b'_j := \overline{a_0\ldots a_{j-1}}b_ja_{j+1}\ldots a_{d-1}$. In the case when some $|b_m|$ is "large" and all other $|b_j|$, $j \neq m$ are "small" we find the following variant more useful:

$$\begin{aligned} B''_k \;:=\; &\sum_{j=0}^{m-1}\overline{c_0\ldots c_{j-1}}b_jc_{j+1}\ldots c_{m-1}a_mc_{m+1}\ldots c_{d-1}\,e^{2\pi ijk/d} \\ &+\overline{c_0\ldots c_{m-1}}b_mc_{m+1}\ldots c_{d-1}\,e^{2\pi imk/d} \\ &+\sum_{j=m+1}^{d-1}\overline{c_0\ldots c_{m-1}a_mc_{m+1}\ldots c_{j-1}}b_jc_{j+1}\ldots c_{d-1}\,e^{2\pi ijk/d}\,, \end{aligned}$$

where we have denoted $c_j := a_j/|a_j|$. This time Lemma 5.5 implies

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B''_k|^q\Big)^{\frac{1}{q}} \;\le\; \Big(|b_m|^p + |a_m|^p\sum_{j\neq m}|b_j|^p\Big)^{\frac{1}{p}}. \tag{5.10}$$

The proof strategy is to compare B_k to B'_k or B''_k by estimating nonlinear terms, and then use inequalities (5.9) or (5.10). As we will soon see, β_d is carefully chosen so that it compensates for the perturbation caused by nonlinear terms.

Choose indices $m, m^* \in \{0,\dots,d-1\}$ such that $|b_m|$ is the largest among the numbers $|b_j|$, and $|b_{m^*}|$ is the largest among the numbers $|b_j|$; $j \neq m$, i.e. the second largest among $|b_j|$. We distinguish the following three cases.

Case 1. $|b_j| \le t_d$ for every j.

Recall that $|a_j|^2 - |b_j|^2 = 1$, which implies $|a_j| \le 1 + |b_j| \le 1 + t_d$. We begin with a rough estimate obtained using (5.5):

$$|B_k| \;\le\; \sum_{j=0}^{d-1}|b_j|\Big(\prod_{l\neq j}(|a_l|+|b_l|)\Big) \;\le\; dt_d(1+2t_d)^{d-1} \;\le\; 2^{-3}d^{-4},$$

which guarantees $|B_k| \le 1$, and thus $\beta_d(|B_k|) \le |B_k|e^{-|B_k|}$ by (5.7). Therefore it is enough to prove

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^q e^{-q|B_k|}\Big)^{\frac{1}{q}} \le \Big(\sum_{j=0}^{d-1}|b_j|^p e^{-p|b_j|}\Big)^{\frac{1}{p}}.$$

Lemma 5.6.

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^q e^{-q|B_k|}\Big)^{\frac{1}{q}} \le \Big(\frac{1}{d}\sum_{k=0}^{d-1}|B'_k|^q e^{-q|B'_k|}\Big)^{\frac{1}{q}} + 2^{-3}d^{-2}|b_{m^*}|^2 \tag{5.11}$$

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B'_k|^q e^{-q|B'_k|}\Big)^{\frac{1}{q}} \le \|b'\|_{\ell^p}\, e^{-\|b'\|_{\ell^p}} \tag{5.12}$$

$$\|b'\|_{\ell^p}\, e^{-\|b'\|_{\ell^p}} \le \Big(\sum_{j=0}^{d-1}|b_j|^p e^{-p|b_j|}\Big)^{\frac{1}{p}} - 2^{-3}d^{-2}|b_{m^*}|^2 \tag{5.13}$$

Here we have denoted $\|b'\|_{\ell^p} := \big(\sum_{j=0}^{d-1}|b'_j|^p\big)^{1/p}$.

The desired inequality is obtained simply by adding the three estimates above.

Proof of Lemma 5.6. We start by showing (5.11). Since $B_k - B'_k$ contains only nonlinear terms and these have at least 3 b's, we have the following error estimate:

$$\begin{aligned}|B_k - B'_k| &\leq \sum_{j_1<j_2<j_3} |b_{j_1}||b_{j_2}||b_{j_3}|\Big(\prod_{l\neq j_1,j_2,j_3}(|a_l|+|b_l|)\Big)\\ &\leq d^3|b_m||b_{m^*}|^2(1+2t_d)^{d-3} \leq 2^{-3}d^{-2}|b_{m^*}|^2.\end{aligned}$$

(For $d=2$ this difference is 0.) By the mean value theorem for te^{-t}:

$$\Big||B_k|e^{-|B_k|} - |B'_k|e^{-|B'_k|}\Big| \leq 2^{-3}d^{-2}|b_{m^*}|^2,$$

and it remains to use Minkowski's inequality.

In order to prove (5.12) we consider the function $\varphi(t) := te^{-qt^{1/q}}$, which is increasing and concave on $[0,1]$ since:

$$\varphi'(t) = e^{-qt^{1/q}}(1-t^{1/q}) > 0, \quad \varphi''(t) = \frac{1}{q}e^{-qt^{1/q}}t^{1/q-1}(-1-q+qt^{1/q}) < 0,$$

for $0<t<1$. Now (5.12) follows using Jensen's inequality and (5.9):

$$\frac{1}{d}\sum_{k=0}^{d-1}\varphi(|B'_k|^q) \leq \varphi\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B'_k|^q\Big) \leq \varphi\bigg(\Big(\sum_{j=0}^{d-1}|b'_j|^p\Big)^{\frac{q}{p}}\bigg).$$

To show (5.13), we observe that $|b'_j| \geq |b_j|$, and thus it suffices to prove

$$\Big(\sum_{j=0}^{d-1}|b_j|^p e^{-p|b_j|}\Big)^{\frac{1}{p}} - \Big(\sum_{j=0}^{d-1}|b'_j|^p e^{-p\,\|b\|_{\ell^p}}\Big)^{\frac{1}{p}} \geq 2^{-3}d^{-2}|b_{m^*}|^2. \tag{5.14}$$

From the mean value theorem we obtain

$$e^{-p|b_j|} - e^{-p\,\|b\|_{\ell^p}} \geq p\,e^{-p\,\|b\|_{\ell^p}}\left(\|b\|_{\ell^p} - |b_j|\right),$$

and using $e^{-p\,\|b\|_{\ell^p}} \geq e^{-pd|b_m|} \geq \frac{1}{2}$ we come to the inequality

$$|b_j|^p e^{-p|b_j|} - |b'_j|^p e^{-p\,\|b\|_{\ell^p}} \geq \frac{1}{2}|b_j|^p\Big(p\,\|b\|_{\ell^p} - p\,|b_j| - \prod_{l\neq j}|a_l|^p + 1\Big). \tag{5.15}$$

Another application of the mean value theorem, this time for the function $t^{1/p}$, gives

$$p\,\|b\|_{\ell^p} - p|b_j| \;\geq\; d^{-1}|b_m|^{1-p}\Big(\sum_{l\neq j}|b_l|^p\Big)\,.$$

On the other hand, we estimate:

$$\begin{aligned}
&\prod_{l\neq j}|a_l|^p - 1 \leq \prod_{l\neq j}|a_l|^2 - 1 = \prod_{l\neq j}(1+|b_l|^2) - 1 \leq e^{\sum_{l\neq j}|b_l|^2} - 1\\
&\leq\; e^{2^{-8}d^{-9}}\sum_{l\neq j}|b_l|^2 \leq 2\Big(\sum_{l\neq j}|b_l|^p\Big)^{\frac{2}{p}} \leq 2d\,|b_m|^{2-p}\Big(\sum_{l\neq j}|b_l|^p\Big)\,,
\end{aligned}$$

to conclude for every $j\neq m$:

$$p\,\|b\|_{\ell^p} - p\,|b_j| - \prod_{l\neq j}|a_l|^p + 1 \;\geq\; 0,$$

and for $j=m$:

$$p\,\|b\|_{\ell^p} - p\,|b_m| - \prod_{l\neq m}|a_l|^p + 1 \;\geq\; 2^{-1}d^{-1}|b_m|^{1-p}|b_{m^*}|^p\,.$$

Now by summing (5.15) over all $j=0,\ldots,d-1$ we get

$$\sum_{j=0}^{d-1}|b_j|^p e^{-p|b_j|} - \sum_{j=0}^{d-1}|b'_j|^p e^{-p\,\|b\|_{\ell^p}} \geq 2^{-2}d^{-1}|b_m||b_{m^*}|^p,$$

and then finally obtain (using the mean value theorem for $t^{1/p}$):

$$\begin{aligned}
&\Big(\sum_{j=0}^{d-1}|b_j|^p e^{-p|b_j|}\Big)^{\frac{1}{p}} - \Big(\sum_{j=0}^{d-1}|b'_j|^p e^{-p\,\|b\|_{\ell^p}}\Big)^{\frac{1}{p}}\\
&\geq \frac{1}{p}(d|b_m|^p)^{\frac{1}{p}-1}2^{-2}d^{-1}|b_m||b_{m^*}|^p \geq 2^{-3}d^{-2}|b_m|^{2-p}|b_{m^*}|^p \geq 2^{-3}d^{-2}|b_{m^*}|^2\,.
\end{aligned}$$

This is exactly (5.14), which completes the proof of Lemma 5.6. $\square$

Case 2. $|b_m| > t_d$, but $|b_j| \le t_d$ for every $j \neq m$.

By (5.8) it is enough to prove

$$\begin{aligned}(2d)^{-5}\Big(\frac{1}{d}\sum_{k=0}^{d-1}(1+\operatorname{arsinh}|B_k|)^{q/2}\Big)^{\frac{1}{q}} \\ \le \Big((2d)^{-5p}(1+\operatorname{arsinh}|b_m|)^{p/2}+\sum_{j\neq m}|b_j|^p e^{-p|b_j|}\Big)^{\frac{1}{p}},\end{aligned}$$

and because $e^{-|b_{m^*}|} \ge e^{-t_d} \ge \frac{1}{2}$, it suffices to show

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}(1+\operatorname{arsinh}|B_k|)^{q/2}\Big)^{\frac{p}{q}} \le (1+\operatorname{arsinh}|b_m|)^{p/2} + 2^{4p}d^{5p}|b_{m^*}|^p\,. \tag{5.16}$$

Lemma 5.7.

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}(1+\operatorname{arsinh}|B_k|)^{q/2}\Big)^{\frac{p}{q}} \le \Big(1+\operatorname{arsinh}\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^q\Big)^{\frac{1}{q}}\Big)^{\frac{p}{2}} \tag{5.17}$$

$$\begin{aligned}\Big(1+\operatorname{arsinh}\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^q\Big)^{\frac{1}{q}}\Big)^{\frac{p}{2}} \le\ & \big(1+\operatorname{arsinh}|b_m|\big)^{p/2} \\ & + \frac{1}{|a_m|}\Big(\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^q\Big)^{\frac{1}{q}} - |b_m|\Big)\end{aligned} \tag{5.18}$$

$$\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^q\Big)^{\frac{1}{q}} \le \Big(|b_m|^p+|a_m|^p\sum_{j\neq m}|b_j|^p\Big)^{\frac{1}{p}} + 2^{4p-1}d^{5p}|a_m||b_{m^*}|^p \tag{5.19}$$

$$\Big(|b_m|^p+|a_m|^p\sum_{j\neq m}|b_j|^p\Big)^{\frac{1}{p}} \le |b_m| + 2^{4p-1}d^{5p}|a_m||b_{m^*}|^p \tag{5.20}$$

Estimate (5.16) follows by successively substituting left hand side of each Inequality (5.18)–(5.20) into the preceding one. Also, we may assume $(\frac{1}{d}\sum_k |B_k|^q)^{\frac{1}{q}} \ge |b_m|$ in (5.18), since otherwise the desired Estimate (5.16) trivially follows from (5.17).

Proof of Lemma 5.7. In order to prove (5.17), we consider the function

$$\psi(t) := \Big(1+\operatorname{arsinh}(t^{2/q})\Big)^{\frac{q}{2}}.$$

One can calculate:

$$
\begin{aligned}
\psi'(t) &= t^{2/q-1}(1+t^{4/q})^{-\frac{1}{2}}\Big(1+\operatorname{arsinh}(t^{2/q})\Big)^{\frac{q}{2}-1}, \\
\psi''(t) &= -\frac{1}{2q}t^{2/q-2}(1+t^{4/q})^{-\frac{3}{2}}\Big(1+\operatorname{arsinh}(t^{2/q})\Big)^{\frac{q}{2}-2} \\
&\cdot \Big(2\big((q-2)+q\,t^{4/q}\big)\operatorname{arsinh}(t^{2/q}) + (q-2)\big((1+t^{4/q})^{\frac{1}{2}}-t^{2/q}\big)^2 + (q-2) + 4t^{4/q}\Big),
\end{aligned}
$$

and conclude (using $q \geq 2$) that ψ is increasing and concave on $[0,\infty)$. Jensen's inequality and elementary inequalities between power means (see [30]) give (5.17):

$$
\frac{1}{d}\sum_{k=0}^{d-1}\psi(|B_k|^{q/2}) \leq \psi\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^{q/2}\Big) \leq \psi\bigg(\Big(\frac{1}{d}\sum_{k=0}^{d-1}|B_k|^{q}\Big)^{\frac{1}{2}}\bigg).
$$

A couple of applications of the mean value theorem, for $(1+t)^{p/2}$ and for $\operatorname{arsinh} t$, together with $1 \leq p \leq 2$, $\sqrt{1+|b_m|^2} = |a_m|$, yield (5.18).

For (5.19) we first estimate the perturbation due to nonlinear terms:

$$
\begin{aligned}
|B_k - B_k'| &\leq \sum_{j_1<j_2<j_3}|b_{j_1}||b_{j_2}||b_{j_3}|\Big(\prod_{l\neq j_1,j_2,j_3}(|a_l|+|b_l|)\Big) \\
&\leq d^3(1+2t_d)^{d-3}(|a_m|+|b_m|)|b_{m^*}|^2 \leq 4d^3|a_m||b_{m^*}|^2\,,
\end{aligned}
$$

and furthermore compare:

$$
\begin{aligned}
|B_k' - B_k''| &\leq |b_m|\Big(\prod_{l\neq m}|a_l| - 1\Big) + \sum_{j\neq m}|a_m||b_j|\Big(\prod_{l\neq m,j}|a_l| - 1\Big) \\
&\leq d|a_m|\Big(\prod_{l\neq m}|a_l| - 1\Big) \leq d|a_m|\Big(e^{\frac{1}{2}\sum_{l\neq m}|b_l|^2} - 1\Big) \leq d^2|a_m||b_{m^*}|^2\,,
\end{aligned}
$$

where in the last line we used $e^x - 1 \leq e^x x$ for $x \geq 0$. These two estimates can be combined, so that Minkowski's inequality, together with (5.10), $5d^3 \leq 2^{4p-1}d^{5p}$, and $|b_{m^*}|^2 \leq |b_{m^*}|^p$, gives (5.19).

To deduce the last Estimate (5.20), we use the mean value theorem for $t^{1/p}$, and $\frac{|a_m|}{|b_m|} \leq 1 + \frac{1}{|b_m|} \leq \frac{2}{t_d} \leq 2^6 d^5$.

$$\begin{aligned}\Big(|b_m|^p + |a_m|^p \sum_{j\neq m} |b_j|^p\Big)^{\frac{1}{p}} - |b_m| \ &\leq \ (|b_m|^p)^{\frac{1}{p}-1} |a_m|^p \sum_{j\neq m} |b_j|^p \\ \leq \ \frac{d|a_m|^p |b_{m^*}|^p}{|b_m|^{p-1}} \ &\leq \ 2^{6p-6} d^{5p-4} |a_m||b_{m^*}|^p \leq \ 2^{4p-1} d^{5p} |a_m||b_{m^*}|^p \,.\end{aligned}$$

This proves Lemma 5.7. $\square$

Case 3. $|b_m| > t_d$ and $|b_{m^*}| > t_d$.

Observe that it suffices to prove

$$\beta_d(|B_k|)^2 \leq \sum_{j=0}^{d-1} \beta_d(|b_j|)^2,$$

for every $k = 0, \ldots, d-1$, because then by elementary inequalities for ℓ^p norms (see [30]) we have

$$\Big(\frac{1}{d} \sum_{k=0}^{d-1} \beta_d(|B_k|)^q\Big)^{\frac{1}{q}} \leq \max_{0\leq k\leq d-1} \beta_d(|B_k|) \leq \Big(\sum_{j=0}^{d-1} \beta_d(|b_j|)^2\Big)^{\frac{1}{2}} \leq \Big(\sum_{j=0}^{d-1} \beta_d(|b_j|)^p\Big)^{\frac{1}{p}}.$$

The rest of the proof is a simple observation taken for instance from [33] or [48], but we include it for completeness. Split the set of indices $\{0, \ldots, d-1\}$ into

$$J_{\text{big}} := \{j \,:\, |b_j| > t_d\} \qquad \text{and} \qquad J_{\text{small}} := \{j \,:\, |b_j| \leq t_d\},$$

so in this case $|J_{\text{big}}| \geq 2$. Since the spectral norm of any $\left[\begin{smallmatrix} a & \bar{b} \\ b & \bar{a} \end{smallmatrix}\right] \in \mathrm{SU}(1,1)$ is equal to $|a| + |b|$, using submultiplicativity of operator norms and (5.2) we deduce

$$|A_k| + |B_k| \leq \prod_{j=0}^{d-1} (|a_j| + |b_j|),$$

which can, after taking logarithms, be written as

$$\operatorname{arsinh} |B_k| \leq \sum_{j=0}^{d-1} \operatorname{arsinh} |b_j|.$$

By (5.8) one always has

$$\beta_d(|B_k|)^2 \le (2d)^{-10}(1 + \operatorname{arsinh}|B_k|)\,,$$

and thus we estimate:

$$\begin{aligned}\beta_d(|B_k|)^2 &\le (2d)^{-10}\Big(1 + \sum_{j\in J_{\mathrm{big}}} \operatorname{arsinh}|b_j| + |J_{\mathrm{small}}| \operatorname{arsinh} t_d\Big) \\ &\le (2d)^{-10} \sum_{j\in J_{\mathrm{big}}} (1 + \operatorname{arsinh}|b_j|) = \sum_{j\in J_{\mathrm{big}}} \beta_d(|b_j|)^2 \le \sum_{j=0}^{d-1} \beta_d(|b_j|)^2\,.\end{aligned}$$

In the above calculation we used $d \operatorname{arsinh} t_d \le dt_d \le 1$ and $|J_{\mathrm{big}}| \ge 2$.

This concludes the last case and therefore Proposition 5.4 is established.

Let us conclude this topic with a somewhat deterring remark about the corresponding result in the real case. While the estimate of Theorem 5.1 is independent of p, the proof makes it seriously dependent on d. It is not clear if the latter dependence can be avoided, but if so, it would require genuinely new methods. Suppose for a moment that we can construct $\beta = \beta_d$ as in Proposition 5.4 that does not depend on d. If we take $d \to \infty$ in (5.3), we will recover an analogue of Conjecture 1.3 on the group $\mathbb{Z}$ (as stated in [48]), and by an easy transference principle the actual Conjecture 1.3 (on $\mathbb{R}$) would also follow. Therefore, uniformization of the constants in d turns out to be an even harder problem than the original one.

References

[1] F. Bernicot, *Fiber-wise Calderón-Zygmund decomposition and application to a bi-dimensional paraproduct*, to appear in Illinois J. Math. arXiv:1011.3614v1 [math.CA]

[2] F. Bernicot, *Uniform estimates for paraproducts and related multilinear multipliers*, Rev. Mat. Iberoam., **25** (2009), no. 3, 1055–1088.

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

[4] J.-M. Bony, *Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires*, Ann. Sci. École Norm. Sup., **14** (1981), 209–246.

[5] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math., **404** (1990), 140–161.

[6] D. L. Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab., **12** (1984), no. 3, 647–702.

[7] A. P. Calderón, *Ergodic theory and translation invariant operators*, Proc. Nat. Acad. Sci. USA, **59** (1968), 349–353.

[8] M. Christ and A. Kiselev, *Maximal functions associated to filtrations*, J. Funct. Anal., **179** (2001), 409–425.

[9] M. Christ and A. Kiselev, *WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials*, J. Funct. Anal., **179** (2001), 426–447.

[10] J.-P. Conze and E. Lesigne, *Théorèmes ergodiques pour des mesures diagonales*, Bull. Soc. Math. France, **112** (1984), no. 2, 143–175.

[11] M. Cotlar, *A unified theory of Hilbert transforms and ergodic theorems*, Rev. Mat. Cuyana, **1** (1955), 105–167.

[12] C. Demeter, *Pointwise convergence of the ergodic bilinear Hilbert transform* Illinois J. Math., **51** (2007), no. 4, 1123–1158.

[13] C. Demeter, M. Pramanik, C. Thiele, *Multilinear singular operators with fractional rank*, Pacific J. Math., **246** (2010), no. 2, 293–324.

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

[15] G. B. Folland, *A course in abstract harmonic analysis*, Stud. Adv. Math., CRC Press, Boca Raton, 1995.

[16] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal., **8** (1998), no. 3, 529–551.

[17] T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal., **11** (2001), no. 3, 465–588.

[18] T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2), **166** (2007), no. 3, 897–946.

[19] L. Grafakos and T. Tao, *Multilinear interpolation between adjoint operators*, J. Funct. Anal., **199** (2003), no. 2, 379–385.

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

[21] P. Jones, *Bilinear singular integrals and maximal functions*, in: Havin and Nikolski (Eds.), Linear and Complex Analysis Problem Book 3, Part I, LNM **1573**, Springer-Verlag, New York, 1994.

[22] R. L. Jones, A. Seeger, J. Wright, *Strong variational and jump inequalities in harmonic analysis*, Trans. Amer. Math. Soc., **360** (2008), no. 12, 6711–6742.

[23] V. Kovač, *Boundedness of the twisted paraproduct*, to appear in Rev. Mat. Iberoam., vol. 28, 2012.
arXiv:1011.6140v1 [math.CA]

[24] V. Kovač, *Uniform constants in Hausdorff-Young inequalities for the Cantor group model of the scattering transform*, Proc. Amer. Math. Soc., posted on July 13, 2011, S 0002-9939(2011)11078-5 (to appear in print).
arXiv:1012.3146v1 [math.CA]

[25] M. Lacey and C. Thiele, L^p *estimates on the bilinear Hilbert transform for* $2 < p < \infty$, Ann. of Math. (2), **146** (1997), no. 3, 693–724.

[26] M. Lacey and C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2), **149** (1999), no. 2, 475–496.

[27] D. Lépingle, *La variation dordre p des semi-martingales*, Z. Wahrscheinlichkeitstheor. Verw. Geb., **36** (1976), 295–316.

[28] Y. Meyer, *Ondelettes et Opérateurs, Volume I: Ondelettes*, Hermann, Paris, 1990.

[29] Y. Meyer, *Ondelettes et Opérateurs, Volume II: Opérateurs*, Hermann, Paris, 1990.

[30] D. S. Mitrinović and P. M. Vasić, *Analytic inequalities*, Springer-Verlag, New York, 1970.

[31] C. Muscalu, J. Pipher, T. Tao, C. Thiele, *Bi-parameter paraproducts*, Acta Math., **193** (2004), no. 2, 269–296.

[32] C. Muscalu, J. Pipher, T. Tao, C. Thiele, *Multi-parameter paraproducts*, Rev. Mat. Iberoam., **22** (2006), no. 3, 963–976.

[33] C. Muscalu, T. Tao, C. Thiele, *A Carleson type theorem for a Cantor group model of the scattering transform*, Nonlinearity **16** (2003), 219–246.

[34] C. Muscalu, T. Tao, C. Thiele, *A counterexample to a multilinear endpoint question of Christ and Kiselev*, Math. Res. Lett., **10** (2003), no. 2–3, 237–246.

[35] C. Muscalu, T. Tao, C. Thiele, L^p *estimates for the biest. I. The Walsh Case*, Math. Ann., **329** (2004), no. 3, 401–426.

[36] C. Muscalu, T. Tao, C. Thiele, L^p *estimates for the biest. II. The Fourier Case*, Math. Ann., **329** (2004), no. 3, 427–461.

[37] C. Muscalu, T. Tao, C. Thiele, *Multi-linear operators given by singular multipliers*, J. Amer. Math. Soc., **15** (2002), 469–496.

[38] C. Muscalu, T. Tao, C. Thiele, *Uniform estimates on paraproducts*, J. Anal. Math., **87** (2002), 369–384.

[39] F. Nazarov and S. Treil, *The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis*, Algebra i Analiz, Vol. 8 (1996), Issue 5, 32–162; English translation in St. Petersburg Math. J., **8** (1997), no. 5, 721–824.

[40] F. Nazarov, S. Treil, A. Volberg, *Bellman function in stochastic control and harmonic analysis*, Systems, approximation, singular integral operators, and related topics (Bordeaux, 2000), 393–423, Oper. Theory Adv. Appl., **129**, Birkhuser, Basel, 2001.

[41] F. Nazarov, S. Treil, A. Volberg, *The Bellman functions and two-weight inequalities for Haar multipliers*, J. Amer. Math. Soc., **12** (1999), no. 4, 909–928.

[42] R. Oberlin, A. Seeger, T. Tao, C. Thiele, J. Wright, *A variation norm Carleson theorem*, to appear in J. Eur. Math. Soc.

[43] C. Pommerenke, *Schlichte Funktionen und analytische Funktionen von beschränkter mittlerer Oszillation*, Comment. Math. Helv., **52** (1977), no. 4, 591–602.

[44] I. D. Shkredov, *On a problem of Gowers*, Izv. Ross. Akad. Nauk Ser. Mat., **70** (2006), no. 2, 179–221.

[45] E. M. Stein, *Harmonic analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton Univ. Press, Princeton, NJ, 1993.

[46] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems, **28** (2008), no. 2, 657–688.

[47] T. Tao, *The ergodic and combinatorial approaches to Szemerédi's theorem*, Centre de Recherches Mathématiques, CRM Proceedings and Lecture Notes, vol. 43 (2007), 145–193.

[48] T. Tao and C. Thiele, *Nonlinear Fourier analysis*, IAS/Park City Math. Institute, Graduate Summer School, 2003, Lecture Notes.

[49] C. Thiele, *Time-frequency analysis in the discrete phase plane*, Ph.D. thesis, Yale University, 1995, Topics in analysis and its applications, 99–152, World Sci. Publ., River Edge, NJ, 2000.

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

[51] S. Verblunsky, *On positive harmonic functions*, Proc. London Math. Soc., **34** (1934), 125–157.

[52] S. Verblunsky, *On positive harmonic functions II*, Proc. London Math. Soc., **40** (1935), 290–320.