The continuous analysis of entangled multilinear forms and applications

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Polona Durcik
aus
Postojna, Slowenien

Bonn 2017

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Christoph Thiele

2. Gutachter: Prof. Dr. Herbert Koch

Tag der Promotion: 14. Juli 2017

Erscheinungsjahr: 2017

Acknowledgements

I would like to express my deepest gratitude to my advisor Professor Christoph Thiele. This thesis would have been impossible without his guidance, continuous support and encouragement over the past several years. I am greatly indebted to him for always generously sharing his deep mathematical insight and brilliant ideas, and for sharing his immense energy, enthusiasm and strong passion for mathematics.

I would like to extend my sincerest thanks to Vjeko Kovač. It was a great pleasure to work with him on several projects that constitute parts of this thesis. I would like to thank him for suggesting very exciting directions of research, as well as for his endless patience and kindness, and his warm invitations to Zagreb, where I am always happy to return to.

It is my great pleasure to thank Kristina Ana Škreb and Luka Rimanić for collaborations that resulted in articles included in this thesis.

I am indebted to Frédéric Bernicot for always being happy to talk to me about my research and for spending a large amount of time discussing interesting math problems with me when kindly inviting me to Nantes. I thank Cristina Benea for many inspiring conversations during my visits in Nantes.

I am indebted to Stefanie Petermichl and Oliver Dragičević for their constant encouragement and invaluable advice.

Being in Bonn would not have been so pleasant without the friendly atmosphere in our research group. I thank all members of the Analysis and PDE group for creating a stimulating work environment and for many interesting discussions of all kinds that evolved during our coffee breaks. Further thanks goes to Bonn International Graduate School of Mathematics for the financial support.

I thank Joris Roos for many mathematical and non-mathematical adventures that we have been through together.

Finally, I thank my family and friends for the unconditional support that I have been receiving from them.

This thesis consists of the following articles.

P. Durcik, *An* L^4 *estimate for a singular entangled quadrilinear form*, Math. Res. Lett. **22** (2015), no. 5, 1317–1332.

P. Durcik, L^p *estimates for a singular entangled quadrilinear form* (2015), accepted for publication in Trans. Amer. Math. Soc., arXiv:1506.08150.

P. Durcik, V. Kovač, K. A. Škreb, C. Thiele, *Norm-variation of ergodic averages with respect to two commuting transformations* (2016), arXiv:1603.00631.

P. Durcik, V. Kovač, L. Rimanić, *On side lengths of corners in positive density subsets of the Euclidean space* (2016), arXiv:1609.09056.

P. Durcik, V. Kovač, C. Thiele, *Power-type cancellation for the simplex Hilbert transform* (2016), accepted for publication in J. Anal. Math., arXiv:1608.00156.

Abstract

The quadrilinear singular integral form

$$\Lambda(F_1, F_2, F_3, F_4) = \int_{\mathbb{R}^4} F_1(x,y)F_2(x,y')F_3(x',y')F_4(x',y)K(x-x',y-y')dxdydx'dy'$$

was motivated by the work of Kovač on the twisted paraproduct, who established boundedness in L^p spaces of a dyadic model of the quadrilinear form Λ. Here K is a smooth two-dimensional Calderón-Zygmund kernel.

In this thesis we introduce a continuous variant of Kovač's approach and address boundedness of the quadrilinear form Λ. Moreover, we study further related multilinear singular integral forms acting on two- and higher-dimensional functions, and discuss their applications to certain problems in ergodic theory and additive combinatorics.

The content of this thesis is organized into six chapters. Chapter 1 is an introductory chapter, stating the main results of Chapters 2–6.

In Chapter 2 we prove the estimate

$$|\Lambda(F_1, F_2, F_3, F_4)| \leq C_{p_1,p_2,p_3,p_4} \|F_1\|_{\mathrm{L}^{p_1}(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^{p_2}(\mathbb{R}^2)} \|F_3\|_{\mathrm{L}^{p_3}(\mathbb{R}^2)} \|F_4\|_{\mathrm{L}^{p_4}(\mathbb{R}^2)}$$

for the exponents $p_1 = p_2 = p_3 = p_4 = 4$.

In Chapter 3 we extend the range of exponents to $2 < p_1, p_2, p_3, p_4 \leq \infty$, whenever the exponents satisfy the scaling condition $\sum_{j=1}^{4} \frac{1}{p_j} = 1$.

In Chapter 4 we study double ergodic averages with respect to two general commuting transformations and establish a sharp quantitative result on their convergence in the norm, by counting their norm-jumps and bounding their norm-variation. This is a joint work with Vjekoslav Kovač, Kristina Ana Škreb and Christoph Thiele.

In Chapter 5 we study side-lengths of corners in subsets of positive upper Banach density of the Euclidean space. We show that if $p \in (1,2) \cup (2,\infty)$ and d is large enough, an arbitrary measurable set $A \subseteq \mathbb{R}^d \times \mathbb{R}^d$ of positive upper Banach density contains corners $(x,y), (x+s,y), (x,y+s)$ such that the ℓ^p norm of the side s attains all sufficiently large real values. This is a joint work with Vjekoslav Kovač and Luka Rimanić.

As a byproduct of the approach in Chapters 4 and 5 we obtain an $\mathrm{L}^4 \times \mathrm{L}^4 \to \mathrm{L}^2$ bound for a two-dimensional bilinear square function related to a singular integral called the triangular Hilbert transform. Boundedness of the triangular Hilbert transform is a major open problem in harmonic analysis.

Chapter 6 is devoted to the simplex Hilbert transform, a higher-dimensional multilinear variant of the triangular Hilbert transform. The content of this chapter is a joint

work with Vjekoslav Kovač and Christoph Thiele. We show that if the Hilbert kernel is truncated to the region $0 < r \leq |x| \leq R < \infty$ on the real line, then L^p bounds for the truncated simplex Hilbert transform grow with a power less than one of the truncation range in the logarithmic scale. Boundedness of the simplex Hilbert transform remains an open problem.

Contents

Chapter 1

Introduction and statement of the main results

Introduction and statement of the main results

A large class of multilinear singular integral forms considered in harmonic analysis can be schematically represented as

$$\int_{\mathbb{R}^n} \Big(\prod_{j=1}^{k} F_j(\rho_j(x))\Big) K(\rho(x)) dx \tag{1}$$

for some $k, n \geq 2$, surjective linear maps $\rho : \mathbb{R}^n \to \mathbb{R}^s$ and $\rho_j : \mathbb{R}^n \to \mathbb{R}^d$, $1 \leq j \leq k$, $d, s \geq 1$. Here K is a smooth s–dimensional Calderón-Zygmund kernel. That is, $\widehat{K}$ is a function on $\mathbb{R}^s$, which is smooth away from the origin and satisfies the standard symbol estimates: there exists a finite constant C_α such that

$$|\partial^\alpha \widehat{K}(\xi)| \leq C_\alpha \|\xi\|^{-|\alpha|} \tag{2}$$

for all multi-indices α and all $0 \neq \xi \in \mathbb{R}^s$. The form (1) is k-linear in the functions $F_j : \mathbb{R}^d \to \mathbb{C}$, $1 \leq j \leq k$.

Typically, one is given an object of the type (1) defined via the Fourier transform on an appropriate space of test functions, such as the Schwartz class. Then, the basic questions of interest are L^p estimates of the form

$$\Big| \int_{\mathbb{R}^n} \Big(\prod_{j=1}^{k} F_j(\rho_j(x))\Big) K(\rho(x)) dx \Big| \leq C \prod_{j=1}^{k} \|F_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^d)} \tag{3}$$

for some choice of exponents $1 \leq p_j \leq \infty$ and the constant C which may depend on $k, n, d, s, \rho_j, \rho, p_j$ and the constant C_α from (2), but not on the functions F_j. The symbol estimates (2) are invariant under isotropic dilations of $\widehat{K}$ in L^∞. A scaling argument shows that bounds of the type (3) are possible only if

$$n - s = d \sum_{j=1}^{k} \frac{1}{p_j}.$$

Let us turn our attention to some familiar instances of (1) and (3). For $x \in \mathbb{R}^n$ we will write $x = (x_1, \ldots, x_n)$.

Example 1 (Brascamp-Lieb)**.** If $\widehat{K}$ is a constant function, then the study of (3) falls under the theory of the *Brascamp-Lieb inequalities*. See the works by Bennett, Carbery, Christ and Tao [2], [3]. $\Diamond$

Example 2 (Coifman-Meyer)**.** Let $k \geq 3$, $n = k$, $d = 1$ and $s = k - 1$. Assume that the linear maps $\rho_j : \mathbb{R}^k \to \mathbb{R}$ and $\rho : \mathbb{R}^k \to \mathbb{R}^{k-1}$ are given by

$$\rho_j(x) = x_j \quad \text{for} \quad 1 \leq j \leq k, \quad \rho(x) = (x_k - x_1, \dots, x_k - x_{k-1}).$$

Then the L^p estimates (3) hold whenever the exponents satisfy $1 < p_j \leq \infty$ and the Hölder scaling condition $\sum_{j=1}^{k} \frac{1}{p_j} = 1$. This is the *multilinear Coifman-Meyer theorem.* We refer to [31] and the references contained therein. ♢

Example 3 (Bilinear Hilbert transform)**.** Let $k = 3, n = 2,\ d = 1,\ s = 1$. Let the maps $\rho_1,\ \rho_2,\ \rho_3 : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$\rho_1(x) = x_1, \quad \rho_2(x) = x_1 + x_2, \quad \rho_3(x) = x_1 + \beta x_2, \tag{4}$$

where $\beta \neq 0, 1$ is a real parameter. Let $\rho : \mathbb{R}^2 \to \mathbb{R}$ be given by $\rho(x) = x_2$ and let

$$\widehat{K}(\xi) = i\pi \mathrm{sgn}(\xi).$$

Then (1) is a trilinear form dual to the *bilinear Hilbert transform.* Lacey and Thiele proved boundedness of the bilinear Hilbert transform in the grounbreaking papers [26], [27]. The bilinear Hilbert transform satisfies $\mathrm{L}^{p_1} \times \mathrm{L}^{p_2} \to \mathrm{L}^{p_3'}$ bounds whenever $1 < p_1, p_2 \leq \infty$, $\frac{2}{3} < p_3' < \infty$, and $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p_3'}$. This in particular implies (3) for the associated trilinear form whenever the exponents satisfy $\sum_{j=1}^{3} \frac{1}{p_j} = 1$ and $1 < p_1, p_2, p_3 \leq \infty$.

In the proof, the authors of [26], [27] develop techniques that are known as time-frequency analysis, and are closely connected with the modulation symmetries that the bilinear Hilbert transform exhibits. ♢

Example 4 (Two-dimensional bilinear Hilbert transform)**.** Specifying $k = 3$, $n = 4$, $d = 2$, $s = 2$, the linear maps $\rho_1, \rho_2, \rho_3 : \mathbb{R}^4 \to \mathbb{R}^2$ by

$$\rho_1(x) = (x_1, x_2) + (x_3, x_4), \quad \rho_2(x) = (x_1, x_2) + B(x_3, x_4), \quad \rho_3(x) = (x_1, x_2), \tag{5}$$

where $B : \mathbb{R}^2 \to \mathbb{R}^2$ is linear, and $\rho : \mathbb{R}^4 \to \mathbb{R}^2$ by $\rho(x) = (x_3, x_4)$, one obtains the *two-dimensional bilinear Hilbert transform.* It was studied by Demeter and Thiele [11], who investigated its boundedness in L^p spaces in dependence on the map B. ♢

Example 5 (Twisted paraproduct)**.** Up to symmetries, the only case for which the time-frequency methods from [11] turned out to be insufficient was when

$$B(y_1, y_2) = (y_1, 0).$$

This case was later called the *twisted paraproduct*, and it was addressed by Kovač [23] by a completely different approach. Kovač proved that the twisted paraproduct satisfies L^p bounds whenever $\sum_{j=1}^{3} \frac{1}{p_j} = 1$ and $1 < p_1, p_3 < \infty$, $2 < p_2 \leq \infty$. ♢

The following observation will be used several times throughout the exposition. If we are interested in L^p estimates for a form associated with the maps ρ_j, $1 \leq j \leq k$, and ρ, then it suffices to bound a form associated with the maps $\tau_j \circ \rho_j \circ \sigma$, $1 \leq j \leq k$, and

$\tau \circ \rho \circ \sigma$ for arbitrary surjective linear maps $\tau_j : \mathbb{R}^d \to \mathbb{R}^d$, $1 \le j \le k$, $\tau : \mathbb{R}^s \to \mathbb{R}^s$, and $\sigma : \mathbb{R}^n \to \mathbb{R}^n$. This follows by changing variables $x \to \sigma(x)$ and observing

$$\|F_j \circ \tau_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^d)} = |\det \tau_j|^{-1/p_j} \|F_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^d)}$$

for each $1 \le j \le k$. The integral kernel $K \circ \tau$ remains Calderón-Zygmund, however, the estimates (2) are in general not uniform in τ.

If in (5) with $B(y_1, y_2) = (y_1, 0)$ we compose ρ_1, ρ_2, ρ_3 and ρ from the right with $\sigma : \mathbb{R}^4 \to \mathbb{R}^4$ given by

$$\sigma(y_1, y_2, y_3, y_4) = (y_3, y_4, y_1 - y_3, y_2 - y_4),$$

boundedness of the twisted paraproduct is equivalent to boundedness of a trilinear form associated with the maps

$$\rho_1(x) = (x_1, x_2), \quad \rho_2(x) = (x_1, x_4), \quad \rho_3(x) = (x_3, x_4)$$

and $\rho(x) = (x_1 - x_3, x_2 - x_4)$. Note that for each $x \in \mathbb{R}^2$ one has

$$\rho_1(x) \cdot e_1 = \rho_2(x) \cdot e_1 \quad \text{and} \quad \rho_2(x) \cdot e_2 = \rho_3(x) \cdot e_2,$$

where e_1 and e_2 are the standard unit vectors in $\mathbb{R}^2$. Because of this we refer to the twisted paraproduct as *twisted* or *entangled*. Informally, one can say that a form is entangled if it can be written in such a way that the functions involved share some one-dimensional variables. Such forms exhibit generalized modulation symmetries. For instance, replacing F_1 by $(g \otimes 1)F_1$ in the twisted paraproduct has the same effect as replacing F_2 by $(g \otimes 1)F_2$, for any $g \in \mathrm{L}^\infty(\mathbb{R})$. Here we have used the notation $(f_1 \otimes f_2)(x_1, x_2) := f_1(x_1) f_2(x_2)$.

The twisted paraproduct can be recognized as a more symmetric quadrilinear form

$$\int_{\mathbb{R}^4} F_1(x_1, x_2) F_2(x_1, x_4) F_3(x_3, x_4) F_4(x_3, x_2) K(x_1 - x_3, x_2 - x_4) dx_1 dx_2 dx_3 dx_4$$

with F_4 being the constant function 1. By localizing $\widehat{K}$ to cones in the frequency plane it suffices to consider the symbol

$$\widehat{K}(\xi_1, \xi_2) = \int_0^\infty c_t\, \widehat{\varphi}(t\xi_1) \widehat{\psi}(t\xi_2) \frac{dt}{t}, \tag{6}$$

where φ, ψ are Schwartz functions, $\widehat{\psi}$ is supported in $\{1/2 \le |\xi| \le 2\}$, and $|c_t| \le 1$ are measurable coefficients. That is, it suffices to consider the form

$$\Lambda(F_1, F_2, F_3, F_4) := \int_0^\infty c_t \int_{\mathbb{R}^4} F_1(x_1, x_2) F_2(x_1, x_4) F_3(x_3, x_4) F_4(x_3, x_2) \tag{7}$$
$$t^{-2} \varphi(t^{-1}(x_1 - x_3)) \psi(t^{-1}(x_2 - x_4)) dx_1 dx_2 dx_3 dx_4 \frac{dt}{t}.$$

To prove estimates for the twisted paraproduct, Kovač passed through a dyadic model of the quadrilinear form (7), given by

$$\Lambda_{\mathrm{d}}(F_1, F_2, F_3, F_4) := \sum_{|I|=|J|} \int_{\mathbb{R}^4} F_1(x_1, x_2) F_2(x_1, x_4) F_3(x_3, x_4) F_4(x_3, x_2) \tag{8}$$
$$|I|^{-2} \mathbf{1}_I(x_1) \mathbf{1}_I(x_3) (\mathbf{1}_{J_{\mathrm{l}}} - \mathbf{1}_{J_{\mathrm{r}}})(x_2) (\mathbf{1}_{J_{\mathrm{l}}} - \mathbf{1}_{J_{\mathrm{r}}})(x_4) dx_1 dx_2 dx_3 dx_4.$$

The sum runs over all dyadic intervals I and J of the same length, $\mathbf{1}_I$ denotes the characteristic function of I, and I_{l}, I_{r} denote the left and the right half of a dyadic interval I, respectively.

In [23], Kovač showed that the dyadic quadrilinear form satisfies the estimates

$$|\Lambda_{\mathrm{d}}(F_1, F_2, F_3, F_4)| \leq C\|F_1\|_{\mathrm{L}^{p_1}(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^{p_2}(\mathbb{R}^2)}\|F_3\|_{\mathrm{L}^{p_3}(\mathbb{R}^2)}\|F_4\|_{\mathrm{L}^{p_4}(\mathbb{R}^2)},$$

whenever $\sum_{j=1}^{4}\frac{1}{p_j} = 1$ and $2 < p_1, p_2, p_3, p_4 \leq \infty$. In [23], boundedness of the twisted paraproduct was then deduced by transferring from the dyadic quadrilinear form with $F_4 = 1$ to the continuous form using the square function of Jones, Seeger and Wright [22]. Using the fiber-wise Calderón-Zygmund decomposition by Bernicot [4] one was able to extend the range of exponents from $2 < p_1, p_2, p_3 < \infty$ to $1 < p_1, p_3 < \infty$ and $2 < p_2 \leq \infty$.

It is natural to ask if the bounds for the twisted paraproduct can be proven directly, without passing through a dyadic model. More generally, one can ask whether the continuous quadrilinear form (7) satisfies any L^p estimates. The question of obtaining estimates for the form (7) with F_4 not necessarily equal to 1 remained unresolved in [23]. The transference trick from the dyadic to continuous model does not apply in this case.

In Chapters 2 and 3 ([12] and [13]) we prove L^p estimates for the quadrilinear form (7). The results from Theorem 1 from [12] and Theorem 1 from [13] are stated in the following.

Theorem 1. *Let $2 < p_1, p_2, p_3, p_4 \leq \infty$ and $\sum_{j=1}^{4}\frac{1}{p_j} = 1$. There exists a finite constant C depending only on the exponents p_j and the Schwartz seminorms of φ, ψ, such that for any Schwartz functions F_1, F_2, F_3, F_4 on $\mathbb{R}^2$ one has the estimate*

$$|\Lambda(F_1, F_2, F_3, F_4)| \leq C\|F_1\|_{\mathrm{L}^{p_1}(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^{p_2}(\mathbb{R}^2)}\|F_3\|_{\mathrm{L}^{p_3}(\mathbb{R}^2)}\|F_4\|_{\mathrm{L}^{p_4}(\mathbb{R}^2)}.$$

To prove Theorem 1 we refine the technique from [23], which was used to show boundedness of the dyadic quadrilinear form (8), and apply it in the Euclidean setting. First by addressing the simpler L^4 case [12], and then the general L^p case [13]. The latter follows from certain generalized restricted weak-type estimates, and it can be obtained after working in a localized setting.

The approach in [23] relies on a structural induction scheme involving repeated applications of the Cauchy-Schwarz inequality, telescoping identity and positivity arguments. It is the easiest to adapt this approach to the Euclidean setting if the bump functions decomposing the kernel (6) are Gaussians. This is the situation which resembles the perfect dyadic model. The general case can then be reduced to the Gaussian case by carefully decomposing the kernel and intertwining the induction scheme with dominations of the bump functions by superpositions of Gaussians.

Multilinear singular integral forms such as (7) naturally appear in problems in ergodic theory when studying ergodic averages along orbits of measure preserving transformations. Let $(X, \mathcal{F}, \mu)$ be a probability space and $S : X \to X$ a measure preserving transformation, i.e. $\mu(S^{-1}E) = \mu(E)$. It is a classical result by von Neumann [33] that for any $f \in \mathrm{L}^2(X)$, the sequence of averages

$$M_n(f)(x) := \frac{1}{n}\sum_{i=0}^{n-1} f(S^i x) \tag{9}$$

converges in $\mathrm{L}^2(X)$ as $n \to \infty$. Birkhoff's pointwise ergodic theorem [7] yields convergence of these averages for almost every $x \in X$.

One can form a bilinear analogue of (9) by taking two commuting measure preserving transformations S, $T : X \to X$ and consider

$$M_n(f,g)(x) := \frac{1}{n}\sum_{i=0}^{n-1} f(S^i x)g(T^i x) \tag{10}$$

for functions f, $g \in \mathrm{L}^4(X)$. Such bilinear averages were motivated by Furstenberg and Katznelson [18] in their work on a multidimensional extension of Szemerédi's theorem. L^2 norm convergence of the sequence $(M_n(f,g))_{n=1}^{\infty}$ is due to Conze and Lesigne [8] and was generalized by Tao [36] to the case of several commuting transformations. Almost everywhere convergence of double ergodic averages is a major open problem in ergodic theory.

Conjecture 2. *Let $(X, \mathcal{F}, \mu)$ be a probability space, S, $T : X \to X$ commuting measure preserving transformations, and $f, g \in \mathrm{L}^\infty(X)$. Then the limit*

$$\lim_{n\to\infty} M_n(f,g)(x)$$

exists for a.e. $x \in X$.

This conjecture is only known to be true in very few special cases. Here we only mention the work of Bourgain [6] who verifies this conjecture in the case when $S = T^m$ for $m \in \mathbb{Z}$.

Classical proofs of norm convergence of ergodic averages give at most very little information on the rate of convergence. To quantify norm convergence of a sequence one typically asks for certain norm-variation estimates, which in turn control the number of jumps of the sequence of certain size. For an extensive treatment of variational estimates and jump inequalities we refer to Jones, Seeger and Wright [22], and Avigad and Rute [1]. In the case of single ergodic averages $M_n(f)$, norm-variation estimates were studied by Jones, Ostrovskii and Rosenblatt [21].

In Chapter 4 ([15]) we address quantitative norm convergence for the double ergodic averages in (10). We obtain a sharp quantitative result on the convergence of $(M_n(f,g))_{n=1}^{\infty}$ in norm, by counting the norm-jumps of this sequence and bounding its norm-variation. The following result is Theorem 1 from [15].

Theorem 3. *There is a finite constant C such that for any σ-finite measure space $(X, \mathcal{F}, \mu)$, any two commuting measure-preserving transformations S, T on that space, and all functions f, $g \in \mathrm{L}^4(X)$ one has*

$$\sum_{j=1}^{m} \|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|_{\mathrm{L}^2(X)}^2 \le C\, \|f\|_{\mathrm{L}^4(X)}^2 \|g\|_{\mathrm{L}^4(X)}^2$$

for each choice of positive integers m and $n_0 < n_1 < \cdots < n_m$.

In a certain model case, an analogue of Theorem 3 has been previously obtained by Kovač [24]. Due to perfect localization in both time and frequency, the model case avoids several of the difficulties arising in [15]. We approach Theorem 3 by transferring it to the Euclidean space via *Calderón's transference principle.* We first pass to the integer lattice $\mathbb{Z}^2$, and then to $\mathbb{R}^2$. For $F, G \in \mathrm{L}^4(\mathbb{R}^2)$, $r > 0$, and $(x_1, x_2) \in \mathbb{R}^2$ we introduce the "rough" bilinear averages on $\mathbb{R}^2$ given by

$$A_r(F,G)(x_1,x_2) := \int_{\mathbb{R}} F(x_1+s,x_2)G(x_1,x_2+s)\, r^{-1}\mathbf{1}_{[0,1)}(r^{-1}s)\, ds.$$

Theorem 3 is a consequence of the following norm-variation estimate in the Euclidean space, which is Theorem 2 from [15].

Theorem 4. *There exists a finite constant C such that for any F, $G \in \mathrm{L}^4(\mathbb{R}^2)$ one has*

$$\sum_{j=1}^{m} \|A_{r_j}(F,G) - A_{r_{j-1}}(F,G)\|^2_{\mathrm{L}^2(\mathbb{R}^2)} \leq C\, \|F\|^2_{\mathrm{L}^4(\mathbb{R}^2)} \|G\|^2_{\mathrm{L}^4(\mathbb{R}^2)} \tag{11}$$

for each choice of positive real numbers $r_0 < r_1 < \cdots < r_m$ and $m \in \mathbb{N}$.

The strategy of the proof of Theorem 4 is to approximate the rough characteristic function of the unit interval by a smooth bump function and expand out the L^2 norm in (11). This eventually leads to studying singular integral forms similar to

$$\int_{\mathbb{R}^4} F(x_1+s,x_2)G(x_1,x_2+s)F(x_1+t,x_2)G(x_1,x_2+t)K(s,t)dx_1dx_2dsdt \tag{12}$$

for a two-dimensional Calderón-Zygmund kernel K. This object falls under (1) with $F_1 = F_3 = F$, $F_2 = F_4 = G$, the maps $\rho_j : \mathbb{R}^4 \to \mathbb{R}^2$ given by

$$\rho_1(x) = (x_1+x_3, x_2), \quad \rho_2(x) = (x_1, x_2+x_3),$$
$$\rho_3(x) = (x_1+x_4, x_2), \quad \rho_4(x) = (x_1, x_2+x_4),$$

and $\rho : \mathbb{R}^4 \to \mathbb{R}^2$ by $\rho(x) = (x_3, x_4)$. Composing ρ_j and ρ from the left with

$$\tau_1(y_1,y_2) = (y_2, y_1+y_2), \quad \tau_2(y_1,y_2) = (y_1, y_1+y_2), \quad \tau_3 = \tau_1, \quad \tau_4 = \tau_2, \quad \tau = \mathrm{id},$$

and from the right with

$$\sigma(y_1,y_2,y_3,y_4) = (y_1, y_2, y_3-y_1-y_2, y_4-y_1-y_2),$$

it is equivalent to discuss bounds for the form

$$\int_{\mathbb{R}^4} F(x_2,x_3)G(x_1,x_3)F(x_2,x_4)G(x_1,x_4) \\ K(x_3-x_1-x_2, x_4-x_1-x_2)dx_1dx_2dx_3dx_4. \tag{13}$$

Structurally it resembles (7). Indeed, the maps ρ_j coincide with the ones in (7) after relabelling. However, the integral kernel is now singular along a different two-dimensional subspace of $\mathbb{R}^4$. Decomposing the kernel into bump functions which are well localized in frequency, one of the key points is to obtain estimates analogous to those in Theorem 1 with careful careful control of the operator norm in terms of the Schwartz seminorms of the bump functions. This in turn translates into the sharp variation-norm estimate for the rough averages on $\mathbb{R}^2$, and finally establishes Theorem 4.

A further source of motivation for studying entangled multilinear singular integral forms is provided by questions on distances in point configurations in "thick" subsets of the Euclidean space. The *upper Banach density* of a set $A \subseteq \mathbb{R}^d$ is defined as

$$\overline{\delta}_d(A) := \limsup_{N\to\infty} \sup_{x\in\mathbb{R}^d} \frac{\big|A \cap (x + [0,N]^d)\big|}{|x + [0,N]^d|}.$$

It is known that a set of positive upper Banach density in $\mathbb{R}^d$, $d \geq 2$, contains all large distances. More precisely, if $d \geq 2$ and $\overline{\delta}_d(A) > 0$, there exists $\lambda_0(A)$ such that for any $\lambda \geq \lambda_0(A)$ the set A contains points $x, x+s$ with $\|s\|_{\ell^2} = \lambda$. This was shown independently by Bourgain [5], Falconer and Marstrand [17], and Furstenberg, Katznelson and Weiss [19]. However, the same statement fails if $x, x+s$ is replaced by a three term arithmetic progression $x, x+s, x+2s$. A counterexample was constructed by Bourgain in [5], and it crucially uses the fact that the ℓ^2 norm satisfies the parallelogram identity.

Recently, Cook, Magyar, and Pramanik [9] investigated related questions on sizes of common differences of three term arithmetic progressions, but with differences measured in the ℓ^p norm for $p \neq 2$, rather than ℓ^2. They obtain the following result.

Theorem 5 (From [9])**.** *For any $p \in (1,2) \cup (2,\infty)$ there exists $d_p \geq 2$ such that for every integer $d \geq d_p$ the following holds. For any measurable set $A \subseteq \mathbb{R}^d$ with $\overline{\delta}_d(A) > 0$ one can find $\lambda_0(A) > 0$ such that for any real number $\lambda \geq \lambda_0(A)$, there exist $x, s \in \mathbb{R}^d$ such that $x, x+s, x+2s \in A$ and $\|s\|_{\ell^p} = \lambda$.*

Cook, Magyar and Pramanik reduce the proof of Theorem 5 to a harmonic analysis problem, which they solve by using bounds for certain modulation invariant multilinear singular integrals, similar to the bilinear Hilbert transform. See [32] and [10].

In Chapter 5 ([14]) we generalize Theorem 5 to *corners* in subsets of $\mathbb{R}^d \times \mathbb{R}^d$, i.e. patterns of the form (x,y), $(x+s,y)$, $(x,y+s)$. The following is Theorem 2 from [14].

Theorem 6. *For any $p \in (1,2) \cup (2,\infty)$ there exists $d_p \geq 2$ such that for every integer $d \geq d_p$ the following holds. For any measurable set $A \subseteq \mathbb{R}^d \times \mathbb{R}^d$ with $\overline{\delta}_d(A) > 0$ one can find $\lambda_0(A) > 0$ such that for any real number $\lambda \geq \lambda_0(A)$, there exist $x, y, s \in \mathbb{R}^d$ such that (x,y), $(x+s,y)$, $(x,y+s) \in A$ and $\|s\|_{\ell^p} = \lambda$.*

To obtain Theorem 6 we follow the outline from [9], but our proof differs in the harmonic analysis part. In this part we need to show an estimate for a higher-dimensional analogue of the form (13). More precisely, we prove an estimate for (13) with x_1, x_2, x_3, x_4 in $\mathbb{R}^d$, F, G functions on $\mathbb{R}^d \times \mathbb{R}^d$ and K a smooth $(2d)-$dimensional Calderón-Zygmund kernel.

As a byproduct of the approach in [15] and [14] we obtain an estimate for a two-dimensional bilinear square function. The following corollary is from [15].

Corollary 7. *Let ψ be a Schwartz function on $\mathbb{R}$ with $\widehat{\psi}(0) = 0$. For any Schwartz functions F, G on $\mathbb{R}^2$ one has*

$$\Big\|\Big(\sum_{i\in\mathbb{Z}}\Big|\int_{\mathbb{R}} F(x_1+s,x_2)G(x_1,x_2+s)2^{-i}\psi(2^{-i}s)ds\Big|^2\Big)^{1/2}\Big\|_{\mathrm{L}^2_{(x_1,x_2)}(\mathbb{R}^2)} \leq C_\psi \|F\|_{\mathrm{L}^4(\mathbb{R}^2)}\|G\|_{\mathrm{L}^4(\mathbb{R}^2)}$$

with a finite constant C_ψ depending on ψ alone.

This result follows after expanding the L^2 norm on the left hand-side, which immediately gives a form of the type (12). Bounds for the singular integral corresponding to the bilinear square function from Corollary 7 are a major open problem in harmonic analysis.

Conjecture 8. *For any Schwartz functions F, G on $\mathbb{R}^2$ one has*

$$\Big\|\mathrm{p.v.}\int_{\mathbb{R}} F(x_1+s,x_2)G(x_1,x_2+s)\frac{ds}{s}\Big\|_{\mathrm{L}^r_{(x_1,x_2)}(\mathbb{R}^2)} \leq C_{p,q}\|F\|_{\mathrm{L}^p(\mathbb{R}^2)}\|G\|_{\mathrm{L}^q(\mathbb{R}^2)} \tag{14}$$

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

This conjecture has been confirmed in a certain model case and when one of the functions takes a special form. See the work by Kovač, Thiele and Zorin-Kranich [25]. The operator in (14) was also called the *triangular Hilbert transform* in [25].

The triangular Hilbert transform controls issues related to pointwise convergence of double ergodic averages (18), as well as many known objects in harmonic analysis. Specifying the functions F, G properly, from the conjectured bounds for the triangular Hilbert transform one would obtain bounds for the *Carleson operator*

$$\mathrm{p.v.}\int_{\mathbb{R}} f(x-s)e^{iN(x)s}\frac{ds}{s},$$

which controls pointwise convergence of Fourier series. Here N is a measurable linearizing function. Bounds for the triangular Hilbert transform would also imply bounds for the one-dimensional bilinear Hilbert transform

$$\mathrm{p.v.}\int_{\mathbb{R}} f(x+s)g(x+\beta s)\frac{ds}{s}, \tag{15}$$

where $0, 1 \neq \beta \in \mathbb{R}$. Note that after dualizing (15) with a third function one obtains the trilinear form discussed in (4). Boundedness of the triangular Hilbert transform would even imply bounds uniform in the parameter β, which is a problem that has been studied extensively in recent years, see [37], [20], [28], [34]. Furthermore, by the method of rotations one could also deduce bounds for the two-dimensional bilinear Hilbert transform (5) with an odd integral kernel, uniformly in the choices of the map B, including bounds for the twisted paraproduct.

More generally, one can define the *simplex Hilbert transform* of degree $n \geq 1$ by

$$\text{p.v.} \int_{\mathbb{R}} \prod_{j=1}^{n} F_j(x + se_j) \frac{ds}{s}, \tag{16}$$

where $x \in \mathbb{R}^n$ and $e_1, \ldots, e_n$ are the standard unit vectors in $\mathbb{R}^n$. If $n = 1$, it coincides with the linear Hilbert transform, while the case $n = 2$ corresponds to the triangular Hilbert transform. No L^p bounds are known for the simplex Hilbert transform if $n \geq 2$.

Conjecture 9. *Let $n \geq 2$. For any Schwartz functions $F_1, \ldots, F_n$ on $\mathbb{R}^n$ one has*

$$\Big\| \text{p.v.} \int_{\mathbb{R}} \prod_{j=1}^{n} F_j(x + se_j) \frac{ds}{s} \Big\|_{\mathrm{L}^r_x(\mathbb{R}^n)} \leq C_{n,p_1,\ldots,p_n} \prod_{j=1}^{n} \|F_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^n)}$$

for some exponents $1 \leq p_1, \ldots, p_n, r \leq \infty$ satisfying $\sum_{j=1}^{n} \frac{1}{p_j} = \frac{1}{r}$.

Analogously to the case $n = 2$, by choosing the functions F_j properly, the simplex Hilbert transform specializes to the *polynomial Carleson operator*

$$\text{p.v.} \int_{\mathbb{R}} f(x - s) e^{i(N_1(x)s + N_2(x)s^2 + \cdots + N_{n-1}(x)s^{n-1})} \frac{ds}{s},$$

which was studied by Lie in [29] and [30]. It also specializes to the one-dimensional *multilinear Hilbert transform*, which is another major open problem in harmonic analysis.

Conjecture 10. *Let $n \geq 3$. For any Schwartz functions $f_1, \ldots, f_n$ on $\mathbb{R}$ one has*

$$\Big\| \text{p.v.} \int_{\mathbb{R}} \prod_{j=1}^{n} f_j(x + js) \frac{ds}{s} \Big\|_{\mathrm{L}^r_x(\mathbb{R})} \leq C_{n,p_1,\ldots,p_n} \prod_{j=1}^{n} \|f_j\|_{\mathrm{L}^{p_j}(\mathbb{R})}$$

for some exponents $1 \leq p_1, \ldots, p_n, r \leq \infty$ satisfying $\sum_{j=1}^{n} \frac{1}{p_j} = \frac{1}{r}$.

Dualizing (16) with an n-dimensional function F_0, interchanging the order of integration and composing the maps

$$(x, s) \mapsto x, \quad (x, s) \mapsto x + se_j \quad \text{for} \quad 1 \leq j \leq n, \quad (x, s) \mapsto s,$$

with suitable linear bijections, studying L^p bounds for the simplex Hilbert transform is equivalent to studying L^p bounds for the more symmetric $(n+1)-$linear form

$$\text{p.v.} \int_{\mathbb{R}^{n+1}} \prod_{j=0}^{n} F_j(x_0, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \frac{1}{x_0 + \cdots + x_n} dx_0 \ldots dx_n.$$

One may approach the multilinear and Hilbert simplex transform by truncating the Hilbert kernel and searching for bounds in terms of the truncation parameters, as initiated in [35], [38]. The *truncated simplex Hilbert transform* is defined by

$$\Lambda_{n,r,R} := \int_{r \leq |x_0 + \cdots + x_n| \leq R} \prod_{j=0}^{n} F_j(x_0, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \frac{1}{x_0 + \cdots + x_n} dx_0 \ldots dx_n,$$

where $0 < r < R < \infty$. We seek estimates of the form

$$|\Lambda_{n,r,R}| \leq C\Big(\log\frac{R}{r}\Big)^{\alpha}\prod_{j=0}^{n}\|F_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^n)} \tag{17}$$

for some $0 \leq \alpha \leq 1$ and exponents $1 \leq p_j \leq \infty$ satisfying $\sum_{j=0}^{n}\frac{1}{p_j} = 1$, with a finite constant C independent of the truncation parameters.

Conjecture 9 would follow from the bound with $\alpha = 0$. On the other hand, Hölder's inequality gives the trivial bound with $\alpha = 1$. Zorin-Kranich [38] improves this estimate to $o(\log\frac{R}{r})$. He builds on the approach by Tao [35], using techniques from additive combinatorics. Tao [35] obtains a $o(\log\frac{R}{r})$ bound for the truncated multilinear Hilbert transform.

In Chapter 6 ([16]) we strengthen these results by showing (17) with a power $\alpha = 1-\epsilon$ for some $\epsilon > 0$ depending only on n and the exponents p_j, which can be taken from the full open Banach range. The main work is spent in proving an estimate for a particular choice of exponents. The following is Theorem 1 from [16].

Theorem 11. *There exists a finite constant C depending only on n such that for any Schwartz functions $F_0,\ldots,F_n$ on $\mathbb{R}^n$ and any $0 < r < R$ we have*

$$|\Lambda_{n,r,R}| \leq C\Big(\log\frac{R}{r}\Big)^{1-2^{-n+1}}\|F_0\|_{\mathrm{L}^{2^n}(\mathbb{R}^n)}\prod_{j=1}^{n}\|F_j\|_{\mathrm{L}^{2^{n-j+1}}(\mathbb{R}^n)}.$$

For other exponents, an estimate with a power less than one then follows by interpolation with the trivial estimate for $\alpha = 1$. Our proof is a structural induction. The induction base is on the level of the forms (7), which are much easier to handle than the simplex Hilbert transform.

It is a natural question if Theorem 3 generalizes to the multiple ergodic averages

$$\frac{1}{n}\sum_{i=0}^{n-1} f_1(S_1^i x)f_2(S_2^i x)\cdots f_k(S_k^i x), \tag{18}$$

where $S_1, S_2, \ldots, S_k : X \to X$ are pairwise commuting measure preserving transformations, and if Theorem 6 generalizes to corners in $(\mathbb{R}^d)^k$

$$(x_1,x_2,\ldots,x_k),\ (x_1+s,x_2,\ldots,x_k),\ (x_1,x_2+s,\ldots,x_k),\ \ldots,\ (x_1,x_2,\ldots,x_k+s) \tag{19}$$

for $k \geq 3$. The main obstruction is that one faces quantities such as the L^2 norm of the k-linear square function corresponding to the simplex Hilbert transform of degree k, and its higher-dimensional analogues. If $k \geq 3$, no L^p bounds for this square function are known.

Conjecture 12. *Let $k \geq 3$. Let ψ be a Schwartz function on $\mathbb{R}$ with $\widehat{\psi}(0) = 0$. For any Schwartz functions $F_1,\ldots,F_k$ on $\mathbb{R}^k$ one has*

$$\Big\|\Big(\sum_{i\in\mathbb{Z}}\Big|\int_{\mathbb{R}}\prod_{j=1}^{k}F_j(x+se_j)2^{-i}\psi(2^{-i}s)ds\Big|^2\Big)^{1/2}\Big\|_{\mathrm{L}^r_x(\mathbb{R}^k)} \leq C_{\psi,k,p_1,\ldots,p_k}\prod_{j=1}^{k}\|F_j\|_{\mathrm{L}^{p_j}(\mathbb{R}^k)}$$

for some exponents $1 \leq p_1,\ldots,p_k,r \leq \infty$ satisfying $\sum_{j=1}^{k}\frac{1}{p_j} = \frac{1}{r}$.

Note that if $k = 2$, the left hand-side specializes to the bilinear square function which is bounded in Corollary 7.

Expanding out the L^2 norm of such a square function of degree $k \geq 3$ leads to problems of similar complexity as the simplex Hilbert transform of degree $k-1$. It is encouraging that [16] obtains estimates for the truncations of the simplex Hilbert transform with constants $J^{1-\epsilon}$ for some $0 < \epsilon < 1$, in the number J of consecutive dyadic scales. Estimates of this type could also be used to study (18) and (19). Furthermore, for the problem (19) also a $o(J)$ bound would suffice (see [35], [38]). However, one would need to consider arbitrary scales rather than consecutive scales as in [16], [35], [38].

Similarly, generalizing the result by Cook, Magyar, Pramanik [9] to $(k+1)$-term arithmetic progressions in $\mathbb{R}^d$,

$$x,\ x+s,\ x+2s,\ \ldots,\ x+ks,$$

is related to a d-dimensional version of the k-linear square function corresponding to the multilinear Hilbert transform. If $k \geq 3$, no L^p bounds for this square function are known.

Conjecture 13. *Let $k \geq 3$. Let ψ be a Schwartz function on $\mathbb{R}$ with $\widehat{\psi}(0) = 0$. For any Schwartz functions $f_1, \ldots, f_k$ on $\mathbb{R}$ one has*

$$\Big\| \Big(\sum_{i\in\mathbb{Z}} \Big| \int_{\mathbb{R}} \prod_{j=1}^{k} f_j(x+js) 2^{-i}\psi(2^{-i}s) ds \Big|^2 \Big)^{1/2} \Big\|_{L^r_x(\mathbb{R})} \leq C_{\psi,k,p_1,\ldots,p_k} \prod_{j=1}^{k} \|f_j\|_{L^{p_j}(\mathbb{R})}$$

for some exponents $1 \leq p_1, \ldots, p_k, r \leq \infty$ satisfying $\sum_{j=1}^{k} \frac{1}{p_j} = \frac{1}{r}$.

Conjecture 12 would imply Conjecture 13 by specifying the functions F_j properly. The analogue of Conjecture 13 in the case $k = 2$ can be deduced from the boundedness of the bilinear Hilbert transform. The case $k = 2$, $p_1 = p_2 = 4$, $r = 2$ can be alternatively deduced from Corollary 7.

References

[1] J. Avigad, J. Rute, *Oscillation and the mean ergodic theorem for uniformly convex Banach spaces*, Ergodic Theory Dynam. Systems **35** (2015), no. 4, 1009–1027.

[2] J. Bennett, A. Carbery, M. Christ, T. Tao, *Finite bounds for Hölder-Brascamp-Lieb multilinear inequalities*, Math. Res. Lett. **17** (2010), no. 4, 647666.

[3] J. Bennett, A. Carbery, M. Christ, T. Tao, *The Brascamp-Lieb inequalities: finiteness, structure, and extremals*, Geom. Funct. Anal. **17** (2008), no. 5, 1343-1415.

[4] F. Bernicot, *Fiber-wise Calderón-Zygmund decoposition and application to a bi-dimensional paraproduct*, Illinois J. Math. **56** (2012), no. 2, 415-422.

[5] J. Bourgain, *A Szemerédi type theorem for sets of positive density in* $\mathbf{R}^k$, Israel J. Math. **54** (1986), no. 3, 307–316.

[6] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161.

[7] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U.S.A. **17** (1931), no. 12, 656–660.

[8] J.-P. Conze, E. Lesigne, *Théorèmes ergodiques pour des mesures diagonales*, Bull. Soc. Math. France **112** (1984), no. 2, 143–175.

[9] B. Cook, Á. Magyar, M. Pramanik, *A Roth type theorem for dense subsets of* $\mathbb{R}^d$ (2015), to appear in Bull. London Math. Soc., available at `arXiv:1511.06010`.

[10] C. Demeter, M. Pramanik, C. Thiele, *Multilinear singular operators with fractional rank*, Pacific J. Math. **246** (2010), no. 2, 293324.

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

[12] P. Durcik, *An* L^4 *estimate for a singular entangled quadrilinear form*, Math. Res. Lett. **22** (2015), no. 5, 1317–1332.

[13] P. Durcik, L^p *estimates for a singular entangled quadrilinear form* (2015), to appear in Trans. Amer. Math. Soc., available at `arXiv:1506.08150`.

[14] P. Durcik, V. Kovač, L. Rimanić, *On side lengths of corners in positive density subsets of the Euclidean space* (2016), preprint, available at `arXiv:1609.09056`.

[15] P. Durcik, V. Kovač, K. A. Škreb, C. Thiele, *Norm-variation of ergodic averages with respect to two commuting transformations* (2016), preprint, available at `arXiv: 1603.00631`.

[16] P. Durcik, V. Kovač, C. Thiele, *Power-type cancellation for the simplex Hilbert transform* (2016), to appear in J. Anal. Math., available at `arXiv:1608.00156`.

[17] K. J. Falconer, J. M. Marstrand, *Plane sets with positive density at infinity contain all large distances*, Bull. London Math. Soc. **18** (1986), no. 5, 471–474.

[18] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Anal. Math. **38** (1978), no. 1, 275–291.

[19] H. Furstenberg, Y. Katznelson, B. Weiss, *Ergodic theory and configurations in sets of positive density. Mathematics of Ramsey theory*, pp. 184–198, Algorithms Combin. **5**, Springer, Berlin, 1990.

[20] L. Grafakos, X. Li, *Uniform bounds for the bilinear Hilbert transform I.*, Ann. of Math. (2), **159** (2004), no. 3, 889-933.

[21] R. L. Jones, I. V. Ostrovskii, J. M. Rosenblatt, *Square functions in ergodic theory*, Ergodic Theory Dynam. Systems **16** (1996), no. 2, 267–305.

[22] R. L. Jones, A. Seeger, J. Wright, *Strong variational and jump inequalities in harmonic analysis*, Trans. Amer. Math. Soc., **360** (2008), no. 12, 67116742.

[23] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. **28** (2012), no. 4, 1143–1164.

[24] V. Kovač, *Quantitative norm convergence of double ergodic averages associated with two commuting group actions*, Ergodic Theory Dynam. Systems **36** (2016), no. 3, 860–874.

[25] V. Kovač, C. Thiele, P. Zorin-Kranich, *Dyadic triangular Hilbert transform of two general and one not too general function*, Forum of Mathematics, Sigma **3** (2015), e25.

[26] M. Lacey, C. Thiele, L^p *estimates on the bilinear Hilbert transform for* $2 < p < \infty$, Ann. of Math. (2) **146** (1997), no. 3, 693–724.

[27] M. Lacey, C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2) **149** (1999), no. 2,

475–496.

[28] X. Li, *Uniform bounds for the bilinear Hilbert transform II.*, Rev. Mat. Iberoam., **22** (2006), no. 3, 1069-1126.

[29] V. Lie, *The (weak-L^2) boundedness of the quadratic Carleson operator*, Geom. Funct. Anal., 19.2 (2009), pp. 457-497.

[30] V. Lie, *The polynomial Carleson operator*, preprint, 2011, `arXiv:1105.4504`.

[31] C. Muscalu, W. Schlag, Classical and multilinear harmonic analysis. Vol. II, Cambridge Studies in Ad- vanced Mathematics **138**, Cambridge University Press, 2013.

[32] C. Muscalu, T. Tao, C. Thiele, *Multi-linear operators given by singular multipliers*, Journal of the AMS, **15** (2002), no.2, 469-496

[33] J. von Neumann, *Proof of the Quasi-Ergodic Hypothesis*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), no. 1, 70–82.

[34] R. Oberlin, C. Thiele, *New uniform bounds for a Walsh model of the bilinear Hilbert transform* Indiana Univ. Math. J., **60** (2011), no. 5, 1693-1712.

[35] T. Tao, *Cancellation for the multilinear Hilbert transform*, Collect. Math. **67** (2016), no. 2, 191–206.

[36] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems **28** (2008), no. 2, 657–688.

[37] C. Thiele, *A uniform estimate.*, Ann. Math. (2), **156** (2002), pp. 519–563.

[38] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform* (2015), to appear in Math. Res. Lett., available at `arXiv:1507.02436`.

Chapter 2

An L^4 estimate for a singular entangled quadrilinear form

An L^4 estimate for a singular entangled quadrilinear form

Polona Durcik

Abstract

The twisted paraproduct can be viewed as a two-dimensional trilinear form which appeared in the work by Demeter and Thiele on the two-dimensional bilinear Hilbert transform. L^p boundedness of the twisted paraproduct is due to Kovač, who in parallel established estimates for the dyadic model of a closely related quadrilinear form. We prove an $(\mathrm{L}^4, \mathrm{L}^4, \mathrm{L}^4, \mathrm{L}^4)$ bound for the continuous model of the latter by adapting the technique of Kovač to the continuous setting. The mentioned forms belong to a larger class of operators with general modulation invariance. Another instance of such is the triangular Hilbert transform, which controls issues related to two commuting transformations in ergodic theory, and for which L^p bounds remain an open problem.

1 Introduction

For four functions F_1, F_2, F_3, F_4 on $\mathbb{R}^2$ we denote their "entangled product"

$$\boldsymbol{F}_{(F_1,F_2,F_3,F_4)}(x, x', y, y') := F_1(x, y)F_2(x', y)F_3(x', y')F_4(x, y'). \tag{1.1}$$

Let m be a bounded function on $\mathbb{R}^2$, smooth away from the origin and satisfying[1]

$$|\partial^\alpha m(\xi, \eta)| \lesssim (|\xi| + |\eta|)^{-|\alpha|} \tag{1.2}$$

for all multi-indices α up to some large finite order. With any such m we associate a quadrilinear form $\Lambda = \Lambda_m$ defined as[2]

$$\Lambda(F_1, F_2, F_3, F_4) := \int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi, -\xi, \eta, -\eta)m(\xi, \eta)d\xi d\eta$$

for Schwartz functions $F_j \in \mathcal{S}(\mathbb{R}^2)$, where $\boldsymbol{F} := \boldsymbol{F}_{(F_1,F_2,F_3,F_4)}$. The object of this paper is to establish the following bound.

Theorem 1. *The quadrilinear form* Λ *satisfies the estimate*

$$|\Lambda(F_1, F_2, F_3, F_4)| \lesssim \|F_1\|_{\mathrm{L}^4(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^4(\mathbb{R}^2)}\|F_3\|_{\mathrm{L}^4(\mathbb{R}^2)}\|F_4\|_{\mathrm{L}^4(\mathbb{R}^2)}. \tag{1.3}$$

2010 *Mathematics Subject Classification.* Primary 42B15; Secondary 42B20.

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

[2] The Fourier transform we use is defined in (2.1).

First published in Mathematical Research Letters in Vol. 22, No. 5, published by International Press.
© 2015 International Press

When m is identically one, Λ corresponds to the pointwise product form

$$\Lambda(F_1, F_2, F_3, F_4) = \int_{\mathbb{R}^2} F_1(x,y)F_2(x,y)F_3(x,y)F_4(x,y)dxdy.$$

The bound (1.3) is then an immediate consequence of Hölder's inequality and holds in a larger range of exponents. In general, we can formally write $\Lambda(F_1, F_2, F_3, F_4)$ as

$$\int_{\mathbb{R}^4} F_1(x,y)F_2(x',y)F_3(x',y')F_4(x,y')\kappa(x'-x,y'-y)dxdx'dydy', \tag{1.4}$$

where κ is a two-dimensional Calderón-Zygmund kernel.

The motivation for these objects originates in the study of the twisted paraproduct [5]. We call the *twisted paraproduct* a trilinear form $T = T_m$ defined as

$$T(F_1, F_2, F_3) := \Lambda(F_1, F_2, F_3, 1).$$

That is, the fourth function in the entangled product $\boldsymbol{F}$ is the constant function one. The form T was proposed by Demeter and Thiele [2] as the dual of a particular case of the two-dimensional bilinear Hilbert transform. This was the only case which could not be treated with the time-frequency techniques in [2]. Lack of applicability of the latter is closely related with general modulation symmetries that the operators T and Λ exhibit. An example of such a symmetry is that for any $g \in \mathrm{L}^\infty(\mathbb{R})$ we have invariance

$$\Lambda((1 \otimes g)F_1, F_2, F_3, F_4) = \Lambda(F_1, (1 \otimes g)F_2, F_3, F_4),$$

where $(f \otimes g)(x,y) := f(x)g(y)$. This is evident from their entangled structure. One can informally say that the generalized modulation invariance is present since several functions depend on the same one-dimensional variable.

First bounds for T are due to Kovač [5], who established

$$|T(F_1, F_2, F_3)| \lesssim_{(p_j)} \|F_1\|_{\mathrm{L}^{p_1}(\mathbb{R}^2)} \|F_2\|_{\mathrm{L}^{p_2}(\mathbb{R}^2)} \|F_3\|_{\mathrm{L}^{p_3}(\mathbb{R}^2)} \tag{1.5}$$

whenever $1/p_1 + 1/p_2 + 1/p_3 = 1$ and $2 < p_1, p_2, p_3 < \infty$. His approach relied on the Bellman function technique. The fiber-wise Calderón-Zygmund decomposition of Bernicot [1] extended the range of exponents to $1 < p_1, p_3 < \infty$, $2 < p_2 \leq \infty$.

Kovač observed that adding the fourth function F_4 to T completes the cyclic structure of the form and results in an object with a high degree of symmetry. For instance, for even kernels κ one has $\Lambda(F_1, F_2, F_3, F_4) = \Lambda(F_3, F_4, F_1, F_2)$. Moreover, T and Λ can be seen as the smallest non-trivial examples of a family of entangled multilinear forms associated with bipartite graphs, whose dyadic models were studied in [4].

To prove (1.5), Kovač passed through a dyadic version of Λ, which we call Λ_{d}. He considered (1.4) with κ replaced by the perfect (dyadic) Calderón-Zygmund kernel

$$\sum_{I \times J} \varphi_I^{\mathrm{d}}(x)\varphi_I^{\mathrm{d}}(x')\psi_J^{\mathrm{d}}(y)\psi_J^{\mathrm{d}}(y'). \tag{1.6}$$

The sum in (1.6) runs over all dyadic squares[3] $I \times J$ in $\mathbb{R}^2$. For a dyadic interval I, the scaling function and the Haar function are defined as[4]

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

respectively. The large range of exponents in (1.5) was achieved by first proving a local bound for the variant of Λ_{d} with the summation in (1.6) running over a subset of dyadic squares called trees. Then, F_4 was set equal to 1 and contributions of a single tree were integrated into a global estimate. This established the desired estimate for the dyadic model T_{d} of T defined by the relation $T_{\mathrm{d}}(F_1, F_2, F_3) := \Lambda_{\mathrm{d}}(F_1, F_2, F_3, 1)$.

It remained to tackle T for continuous kernels. Via the cone decomposition, see [6], this problem was first reduced to the case

$$\kappa(s,t) = \sum_{k\in\mathbb{Z}} 2^k \varphi(2^k s) 2^k \psi(2^k t), \tag{1.7}$$

where $\varphi, \psi \in \mathcal{S}(\mathbb{R})$ are two Schwartz functions and $\widehat{\psi}$ is supported on $\{1 \le |\xi| \le 2\}$. Their dilations by 2^k can be seen as continuous analogues of φ_I, ψ_I. In [5], the bound (1.5) was finally established by relating the special case of T, associated with (1.7), to the dyadic T_{d}. This was done by rewriting T and T_{d} using convolutions and martingale averages in the respective cases. Then, the square functions of Jones, Seeger and Wright [3] were used to compare the continuous with the discrete averaging operator.

A natural question is what we can say about Λ if the function 1 is replaced by a completely general function F_4. For the dyadic model Λ_{d}, Kovač proved $(\mathrm{L}^{p_1}, \mathrm{L}^{p_2}, \mathrm{L}^{p_3}, \mathrm{L}^{p_4})$ estimates whenever p_j are Hölder-type exponents satisfying $2 < p_j < \infty$ for all j. See [5] and [4]. However, due to the more complex structure of the form, one cannot efficiently rewrite Λ and Λ_{d} in a similar way as T and T_{d} to exploit the mentioned square functions. Thus, the question about Λ associated with any m satisfying (1.2) remains.

In the present note we obtain an answer in this direction by adapting the technique used to treat Λ_{d} in [5] to the continuous setting. We address the simplest L^4 case only. It is expected that suitable tree decompositions will eventually enable us to prove (1.3) for a larger range of exponents. However, for the considered quadrilinear form we cannot make use of the fiber-wise Calderón-Zygmund decomposition by Bernicot.

The core argument in [5] intertwines two applications of the Cauchy-Schwarz inequality, which gradually separates the functions F_j, and two applications of an algebraic identity, which "interchanges" the functions φ^{d} and ψ^{d}. This identity, involving a telescoping argument in the dyadic case, is now replaced by a differential equality combining the fundamental theorem of calculus and the Leibniz rule. The main issue in the continuous setup is that the mentioned algebraic trick can be applied twice if the functions φ, ψ, decomposing the kernel, are sufficiently symmetric. For example, even functions would work. Moreover, they need to possess enough decay and have certain smoothness properties, which should be maintained throughout the process. Suitable candidates which fulfil the requirements are, for instance, the Gaussian exponential functions.

[3] A dyadic square is a product of two dyadic intervals of the same length. A dyadic interval is an interval of the form $[2^k m, 2^k(m+1))$, k, $m \in \mathbb{Z}$.

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

Although we cannot expect our functions φ, ψ to be even, much less the Gaussian exponential functions, we are able to overcome the mentioned restrictions as follows. First, the reduction to the case of a concrete kernel, such as (1.7), is done by a careful choice of the functions φ, ψ. This way we obtain some of the required symmetry and regularity. Second, after each application of the Cauchy-Schwarz inequality we dominate certain functions with a suitable superposition of dilated Gaussian exponential functions. This gradually reduces the two algebraic steps to the case of Gaussians, which most resembles the dyadic telescoping trick.

Besides extending the exponent range, it would be of interest to obtain boundedness results for the continuous models of the forms from [4], associated with bipartite graphs.

Let us briefly comment on another related open problem. There is a question of establishing L^p estimates for the akin trilinear form

$$\Lambda_{\triangle}(F_1, F_2, F_3) := \int_{\mathbb{R}} \widehat{\boldsymbol{F}}(\xi, \xi, \xi)\mathrm{sgn}(\xi)d\xi$$

where the entangled product $\boldsymbol{F}$ is now given by

$$\boldsymbol{F}(x, y, z) := F_1(x, y)F_2(y, z)F_3(z, x).$$

Passing to the spatial side, one has up to a constant

$$\Lambda_{\triangle}(F_1, F_2, F_3) = \int_{\mathbb{R}^3} F_1(x, y)F_2(y, z)F_3(z, x)\frac{-1}{x + y + z}dxdydz.$$

The structure of $\Lambda_{\triangle}$ corresponds to the three-cycle and for this reason it is called the *triangular Hilbert transform*. No L^p bounds for $\Lambda_{\triangle}$ or for its dyadic model are known. Lack of the bipartite structure prevents to approach it with the techniques from [4].

Boundedness of $\Lambda_{\triangle}$ would imply boundedness for certain instances of the two- dimensional bilinear Hilbert transform and the twisted paraproduct. Further interest in $\Lambda_{\triangle}$ arises from ergodic theory. It is proposed by Demeter and Thiele [2] to approach the open question of pointwise almost everywhere convergence for ergodic averages

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(S^n x),$$

where S, $T : X \to X$ are two commuting measure preserving transformations on a probability space X, via an examination of the triangular Hilbert transform.

Acknowledgement. I am grateful to my advisor Prof. Christoph Thiele for his constant support, valuable consultations on the problem and numerous suggestions on improving the text. I am thankful to Vjekoslav Kovač for useful discussions.

2 Decomposition of the symbol

To begin, we reduce the general symbol to a particular function by decomposing m into pieces which are supported on certain subsets of two double cones. We follow the main ideas discussed in [6]. However, we do not discretize, but rather keep continuum in the scale.

The Fourier transform we shall use throughout this note is defined as

$$\widehat{f}(\omega) := \int_{\mathbb{R}^n} f(\tau) e^{-2\pi i \tau\cdot\omega} d\tau. \tag{2.1}$$

By a smooth partition of unity and symmetry in ξ, η we may assume that m is supported on the double cone

$$\{(\xi,\eta) : |\xi| \le 1.001|\eta|\}$$

centered around the η-axis. Choosing double cones over single cones will allow us to use functions that are symmetric around the origin. We can choose the partition of unity such that (1.2) is preserved, possibly with a different constant.

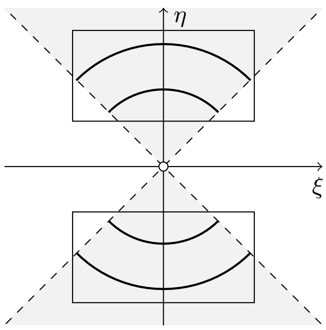

Figure 1: Decomposition of m.

By $B(0,R)$ we denote the ball of radius R centered at the origin in $\mathbb{R}^2$. Let θ be a function on $\mathbb{R}^2$ such that $\widehat{\theta}$ is smooth, real, radial and supported in the annulus $B(0,2.7)\setminus B(0,1.7)$. We normalize so that for every $(\xi,\eta)\neq 0$ we have

$$\int_0^\infty \widehat{\theta}(t\xi, t\eta)\frac{dt}{t} = 1.$$

This can be achieved, since $\widehat{\theta}$ is radial and supported away from 0. Then we can write

$$m(\xi,\eta) = \int_0^\infty m_t(\xi,\eta)\frac{dt}{t},$$

where $m_t(\xi,\eta) := m(\xi,\eta)\widehat{\theta}(t\xi,t\eta)$.

In what follows we will be working with certain smooth bump functions, for which we need the following technical lemma. Its proof can be found in the appendix.

Lemma 2. *Let* $\varepsilon := 0.001$. *There exists a non-negative real-valued function* $f \in C_0^\infty(\mathbb{R})$ *which is supported in* $[1,3]$, *even about* 2 *and constantly equal* 1 *on* $[1+\varepsilon, 3-\varepsilon]$, *such that* $f^{1/2}$ *and*

$$\Big(\int_x^\infty \frac{f(t)+f(-t)}{t}dt\Big)^{1/2} \tag{2.2}$$

belong to $C_0^\infty(\mathbb{R})$.

Now consider m_1. Its support is contained in the union of the rectangles

$$[-2,2]\times[-3,-1] \qquad \text{and} \qquad [-2,2]\times[1,3].$$

Let f be the function from Lemma 2 and let $\vartheta_1, \vartheta_2 \in \mathcal{S}(\mathbb{R})$ be such that $\widehat{\vartheta_1}(\xi) = f((\xi+4)/2)$ and $\widehat{\vartheta_2}(\xi) = f(\xi) + f(-\xi)$. Then $\widehat{\vartheta_1} \otimes \widehat{\vartheta_2}$ equals 1 on the support of m_1. Thus, by dilating $\widehat{\vartheta_1}, \widehat{\vartheta_2}$ in t, for every $t > 0$ we can write

$$m_t(\xi,\eta) = m_t(\xi,\eta)\widehat{\vartheta_1}(t\xi)\widehat{\vartheta_2}(t\eta).$$

This can be rewritten further using the Fourier inversion formula on m_t as

$$m_t(\xi,\eta) = \Big(\int_{\mathbb{R}^2} \mu_t(u,v) e^{2\pi i u t\xi} e^{2\pi i v t\eta} du dv\Big) \widehat{\vartheta_1}(t\xi)\widehat{\vartheta_2}(t\eta),$$

where $\mu_t := t^2 \widehat{m_t}(t\cdot, t\cdot)$. Integrating by parts sufficiently many times, using that (1.2) holds for $m(\xi/t, \eta/t)$ uniformly in t and considering the support of m_t we obtain

$$\begin{aligned}|\mu_t(u,v)| &= t^2 \Big|\int_{\mathbb{R}^2} m_t(\xi,\eta) e^{-2\pi i(ut\xi + vt\eta)} d\xi d\eta\Big| = \Big|\int_{\mathbb{R}^2} m_t\Big(\frac{\xi}{t}, \frac{\eta}{t}\Big) e^{-2\pi i(u\xi+v\eta)} d\xi d\eta\Big| \\ &\lesssim (1+|u|)^{-12}(1+|v|)^{-12}.\end{aligned}$$

Define $\varphi^{(u)}, \psi^{(v)}$ by

$$\begin{aligned}\widehat{\varphi^{(u)}}(\xi) &:= (1+|u|)^{-5}(\widehat{\vartheta_1}(\xi))^{1/2} e^{\pi i u\xi}, \\ \widehat{\psi^{(v)}}(\eta) &:= (\widehat{\vartheta_2}(\eta))^{1/2} e^{\pi i v\eta}.\end{aligned} \tag{2.3}$$

By Lemma 2 we have $(\widehat{\vartheta_1})^{1/2} \in C_0^\infty(\mathbb{R})$, so the function $\varphi^{(u)}$ satisfies the bound

$$|\varphi^{(u)}(x)| \lesssim (1+|x|)^{-5} \tag{2.4}$$

uniformly in u. We will apply this fact in the following section. Now we can write

$$m(\xi,\eta) = \int_0^\infty \int_{\mathbb{R}^2} \widetilde{\mu}_t(u,v) (\widehat{\varphi^{(u)}}(t\xi))^2 (\widehat{\psi^{(v)}}(t\eta))^2 du dv \frac{dt}{t},$$

where the coefficients $\widetilde{\mu}_t$ are defined as

$$\widetilde{\mu}_t(u,v) := (1+|u|)^{10}\mu_t(u,v).$$

Note that $(\widehat{\vartheta_1})^{1/2}$ and $(\widehat{\vartheta_2})^{1/2}$ are real-valued and even, so $\varphi^{(u)}, \psi^{(v)}$ are multiples of translates of real-valued functions and thus real-valued.

To summarize, on a double cone we have decomposed $\Lambda(F_1, F_2, F_3, F_4)$ into

$$\int_{\mathbb{R}^2}\int_0^\infty \widetilde{\mu}_t(u,v) \int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi,-\xi,\eta,-\eta) (\widehat{\varphi^{(u)}}(t\xi))^2 (\widehat{\psi^{(v)}}(t\eta))^2 d\xi d\eta \frac{dt}{t} du dv.$$

By the rapid decay of the coefficients $\widetilde{\mu}_t$ it will suffice to prove (1.3) for the form

$$\int_0^\infty \Big|\int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi,-\xi,\eta,-\eta) (\widehat{\varphi^{(u)}}(t\xi))^2 (\widehat{\psi^{(v)}}(t\eta))^2 d\xi d\eta\Big| \frac{dt}{t}, \tag{2.5}$$

provided that the estimate holds uniformly in the parameters u, v.

From now on we assume that the functions $F_j \in \mathcal{S}(\mathbb{R}^2)$ are real-valued, as otherwise we can split them into real and imaginary parts and use quadrisublinearity of (2.5).

3 Proof of Theorem 1

The proof proceeds with studying the special case (2.5). For $t > 0$ and four functions $\phi_i \in \mathcal{S}(\mathbb{R})$ we define

$$L^t_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4) := \int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi,-\xi,\eta,-\eta)\widehat{\phi_1}(t\xi)\widehat{\phi_2}(-t\xi)\widehat{\phi_3}(t\eta)\widehat{\phi_4}(-t\eta)d\xi d\eta.$$

For the rest of this note we will consider objects of the type

$$\Lambda_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4) := \int_0^\infty L^t_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4)\frac{dt}{t}$$

and

$$\widetilde{\Lambda}_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4) := \int_0^\infty \Big|L^t_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4)\Big|\frac{dt}{t}. \tag{3.1}$$

Observe that (2.5) is obtained from (3.1) by choosing

$$\begin{aligned} \phi_1 &= \varphi^{(u)}, & \phi_3 &= \psi^{(v)}, \\ \phi_2 &= \varphi^{(-u)}, & \phi_4 &= \psi^{(-v)}. \end{aligned}$$

This follows from (2.3) and from the functions $\widehat{\vartheta_1}, \widehat{\vartheta}_2$ being even.

We shall now express $L^t_{\phi_1,\phi_2,\phi_3,\phi_4}$ on the spatial side. Let us denote by $[f]_t$ the L^1-dilation of a function f by a parameter $t > 0$, i.e. $[f]_t(x) := t^{-1}f(t^{-1}x)$. Then, $\widehat{[f]_t}(\xi) = \widehat{f}(t\xi)$. Since the integral of the Fourier transform of a Schwartz function in $\mathbb{R}^4$ over the hyperplane

$$\{(\xi,-\xi,\eta,-\eta) : \xi,\eta \in \mathbb{R}\}$$

equals the integral of the function itself over the perpendicular hyperplane

$$\{(p,p,q,q) : p,q \in \mathbb{R}\},$$

we can write $L^t_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4)$ as

$$\int_{\mathbb{R}^2} \boldsymbol{F} * ([\phi_1]_t \otimes [\phi_2]_t \otimes [\phi_3]_t \otimes [\phi_4]_t)(p,p,q,q)dpdq.$$

Expanding the convolution, the last display can be identified as

$$\int_{\mathbb{R}^6} F_1(x,y)F_2(x',y)F_3(x',y')F_4(x,y') \\ [\phi_1]_t(p-x)[\phi_2]_t(p-x')[\phi_3]_t(q-y)[\phi_4]_t(q-y')dxdx'dydy'dpdq.$$

Now we are ready to start. The inequality (1.3), which we want to establish, is homogeneous, so we may normalize

$$\|F_j\|_{\mathrm{L}^4(\mathbb{R}^2)} = 1,$$

for $j = 1,2,3,4$. Thus, we are set to show

$$\widetilde{\Lambda}_{\varphi^{(u)},\varphi^{(-u)},\psi^{(v)},\psi^{(-v)}}(F_1,F_2,F_3,F_4) \lesssim 1.$$

The proof starts with an application of the Cauchy-Schwarz inequality. To preserve the mean zero property of $\psi^{(v)}, \psi^{(-v)}$ we separate the involved functions according to the variables y, y' and estimate $\widetilde{\Lambda}_{\varphi^{(u)},\varphi^{(-u)},\psi^{(v)},\psi^{(-v)}}(F_1,F_2,F_3,F_4)$ by

$$\int_0^\infty \int_{\mathbb{R}^4} \left| \int_{\mathbb{R}} F_1(x,y)F_2(x',y)[\psi^{(v)}]_t(q-y)dy \right| \left| \int_{\mathbb{R}} F_3(x',y')F_4(x,y')[\psi^{(-v)}]_t(q-y')dy' \right|$$
$$[|\varphi^{(u)}|]_t(p-x)[|\varphi^{(-u)}|]_t(p-x')dxdx'dpdq\frac{dt}{t}.$$

Applying the Cauchy-Schwarz inequality bounds this expression by the product

$$\Lambda_{|\varphi^{(u)}|,|\varphi^{(-u)}|,\psi^{(v)},\psi^{(v)}}(F_1,F_2,F_2,F_1)^{1/2}\Lambda_{|\varphi^{(u)}|,|\varphi^{(-u)}|,\psi^{(-v)},\psi^{(-v)}}(F_4,F_3,F_3,F_4)^{1/2}.$$

We estimate the first factor of the above display, the second is dealt with similarly.

To further separate the involved functions we would like to apply the Cauchy-Schwarz inequality again, which now needs to be done in the complementary variables. So we need to "switch" the functions $\varphi^{(u)}$ and $\psi^{(v)}$. This is where we make use of the following lemma, a continuous analogue of the telescoping identity from [5].

Lemma 3. *Assume that we have two pairs of real-valued Schwartz functions (ρ_i, σ_i), $i = 1, 2$, which satisfy*

$$-t\partial_t|\widehat{\rho_i}(t\tau)|^2 = |\widehat{\sigma_i}(t\tau)|^2 \quad \text{for} \quad i = 1, 2. \tag{3.2}$$

Then with $c := |\widehat{\rho_1}(0)|^2|\widehat{\rho_2}(0)|^2$ we have

$$\Lambda_{\sigma_1,\rho_2}(F_1,F_2,F_3,F_4) + \Lambda_{\rho_1,\sigma_2}(F_1,F_2,F_3,F_4) = c\int_{\mathbb{R}^2} F_1F_2F_3F_4, \tag{3.3}$$

where we have denoted $\Lambda_{\sigma,\rho} = \Lambda_{\sigma,\sigma,\rho,\rho}$.

Proof. By the fundamental theorem of calculus,

$$\int_0^\infty \partial_t(|\widehat{\rho_1}(t\xi)|^2|\widehat{\rho_2}(t\eta)|^2)dt = -|\widehat{\rho_1}(0)|^2|\widehat{\rho_2}(0)|^2. \tag{3.4}$$

The left hand-side of (3.4) equals

$$\int_0^\infty t\partial_t(|\widehat{\rho_1}(t\xi)|^2)|\widehat{\rho_2}(t\eta)|^2\frac{dt}{t} + \int_0^\infty |\widehat{\rho_1}(t\xi)|^2 t\partial_t(|\widehat{\rho_2}(t\eta)|^2)\frac{dt}{t}. \tag{3.5}$$

The functions ρ, σ are real-valued, so $\overline{\widehat{\rho}}(\eta) = \widehat{\rho}(-\eta)$, and analogously for σ. Together with (3.2) this shows that (3.5) can be written as

$$-\int_0^\infty \widehat{\sigma_1}(t\xi)\widehat{\sigma_1}(-t\xi)\widehat{\rho_2}(t\eta)\widehat{\rho_2}(-t\eta)\frac{dt}{t} - \int_0^\infty \widehat{\rho_1}(t\xi)\widehat{\rho_1}(-t\xi)\widehat{\sigma_2}(t\eta)\widehat{\sigma_2}(-t\eta)\frac{dt}{t}. \tag{3.6}$$

Now multiply (3.4) by $\widehat{\boldsymbol{F}}(\xi,-\xi,\eta,-\eta)$ and integrate in the variables ξ, η. It remains to use (3.6) and to evaluate the right hand-side of (3.4) as $-|\widehat{\rho_1}(0)|^2|\widehat{\rho_2}(0)|^2$ times

$$\int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi,-\xi,\eta,-\eta)d\xi d\eta = \int_{\mathbb{R}^2} \boldsymbol{F}(x,x,y,y)dxdy$$
$$= \int_{\mathbb{R}^2} F_1(x,y)F_2(x,y)F_3(x,y)F_4(x,y)dxdy.$$

This proves the claim. □

To apply Lemma 3 we would like to have $\varphi^{(u)} = \varphi^{(-u)}$, as then we would get

$$\Lambda_{|\varphi^{(u)}|,|\varphi^{(-u)}|,\psi^{(v)},\psi^{(v)}} = \Lambda_{|\varphi^{(u)}|,\psi^{(v)}}.$$

However, we do not have $\varphi^{(u)} = \varphi^{(-u)}$ in general. For this and to circumvent possible lack of smoothness of $|\varphi^{(\pm u)}|$, we dominate $|\varphi^{(\pm u)}|$ with a superposition of the Gaussian exponential functions. Consider

$$\Phi(x) := \int_1^\infty \frac{1}{\alpha^5} e^{-\left(\frac{x}{\alpha}\right)^2} d\alpha = \frac{1}{2x^4}(1 - e^{-x^2}(x^2+1)).$$

The function Φ is positive, continuous at zero and for large x comparable to x^{-4}. Let us denote the L^1-normalized Gaussian rescaled by a parameter $\alpha > 0$ by

$$g_\alpha(x) := \frac{1}{\sqrt{\pi}\alpha} e^{-\left(\frac{x}{\alpha}\right)^2}. \tag{3.7}$$

Then we can write

$$\Phi = \pi^{-1/2} \int_1^\infty \frac{1}{\alpha^4}\, g_\alpha \, d\alpha.$$

Since $|\varphi^{(\pm u)}|$ satisfies the decay estimate (2.4), we can bound it pointwise by Φ multiplied by some positive constant which is uniform in u. Positivity of the integrands in

$$\Lambda_{|\varphi^{(u)}|,|\varphi^{(-u)}|,\psi^{(v)},\psi^{(v)}}(F_1,F_2,F_2,F_1) = \int_0^\infty \int_{\mathbb{R}^4} \Big(\int_{\mathbb{R}} F_1(x,y) F_2(x',y) [\psi^{(v)}]_t (q-y) dy \Big)^2$$
$$[|\varphi^{(u)}|]_t(p-x)\, [|\varphi^{(-u)}|]_t (p-x') dx dx' dp dq \frac{dt}{t} \tag{3.8}$$

then allows us to dominate

$$\Lambda_{|\varphi^{(u)}|,|\varphi^{(-u)}|,\psi^{(v)},\psi^{(v)}}(F_1,F_2,F_2,F_1) \lesssim \int_1^\infty \int_1^\infty \Lambda_{g_\alpha, g_\beta, \psi^{(v)},\psi^{(v)}}(F_1,F_2,F_2,F_1) \frac{d\alpha}{\alpha^4}\frac{d\beta}{\beta^4}.$$

To reduce to only one scaling parameter in the last line we split the integration into the regions $\alpha \geq \beta$ and $\alpha < \beta$. By symmetry it suffices to estimate the region $\alpha \geq \beta$ only, on which we bound $\beta g_\beta \leq \alpha g_\alpha$ for $\alpha, \beta \geq 1$. This leaves us with having to estimate

$$\int_1^\infty \Lambda_{g_\alpha,\psi^{(v)}}(F_1,F_2,F_2,F_1) \frac{d\alpha}{\alpha^3}.$$

We shall now apply Lemma 3 with $(\rho_1,\sigma_1) = (g_\alpha, h_\alpha)$ and $(\rho_2,\sigma_2) = (\phi, \psi^{(v)})$, where we define $h_\alpha(x) := \alpha (g_\alpha)'(x)$ and ϕ is defined via

$$\widehat{\phi}(\xi) := \left(\int_\xi^\infty |\widehat{\psi^{(v)}}(\tau)|^2 \frac{d\tau}{\tau} \right)^{1/2}. \tag{3.9}$$

Since $|\widehat{\psi^{(v)}}|^2 = \widehat{\vartheta_2}$, by Lemma 2 the function $\widehat{\phi}$ belongs to $C_0^\infty(\mathbb{R})$. Note that the two pairs of functions (ρ_i, σ_i) satisfy (3.2), which follows by a straightforward calculation. Lemma 3 now yields

$$\Lambda_{g_\alpha,\psi^{(v)}}(F_1,F_2,F_2,F_1) = -\Lambda_{h_\alpha,\phi}(F_1,F_2,F_2,F_1) + \widehat{\phi}(0)^2 \int_{\mathbb{R}^2} F_1^2 F_2^2. \tag{3.10}$$

By the Cauchy-Schwarz inequality we have

$$\int_{\mathbb{R}^2} F_1^2 F_2^2 \leq \|F_1\|_{\mathrm{L}^4(\mathbb{R}^2)}^2 \|F_2\|_{\mathrm{L}^4(\mathbb{R}^2)}^2 = 1,$$

so it remains to consider the first term on the right hand-side of (3.10).

To estimate it we repeat the just performed steps, which will further separate the functions F_1, F_2. The role of $\varphi^{(\pm u)}$ is now taken over by ϕ and the role of $\psi^{(\pm v)}$ is assumed by h_α. Therefore we can group the integrals in $\Lambda_{h_\alpha,\phi}$ according to the variables x, x', and bound $|\Lambda_{h_\alpha,\phi}(F_1, F_2, F_2, F_1)|$ by

$$\int_0^\infty \int_{\mathbb{R}^4} \left| \int_{\mathbb{R}} F_1(x,y) F_1(x,y') [h_\alpha]_t(p-x) dx \right| \left| \int_{\mathbb{R}} F_2(x',y') F_2(x',y) [h_\alpha]_t(p-x') dx' \right| \\ [|\phi|]_t(q-y) [|\phi|]_t(q-y') dy dy' dp dq \frac{dt}{t}.$$

Applying the Cauchy-Schwarz inequality we obtain

$$|\Lambda_{h_\alpha,\phi}(F_1, F_2, F_2, F_1)| \leq \Lambda_{h_\alpha,|\phi|}(F_1, F_1, F_1, F_1)^{1/2} \Lambda_{h_\alpha,|\phi|}(F_2, F_2, F_2, F_2)^{1/2}.$$

Now we dominate the rapidly decaying function $|\phi|$ by a positive constant times Φ, which gives for the first factor

$$\Lambda_{h_\alpha,|\phi|}(F_1, F_1, F_1, F_1) \lesssim \int_1^\infty \int_1^\infty \Lambda_{h_\alpha,h_\alpha,g_\gamma,g_\delta}(F_1, F_1, F_1, F_1) \frac{d\gamma}{\gamma^4} \frac{d\delta}{\delta^4}. \tag{3.11}$$

By symmetry it again suffices to estimate

$$\int_1^\infty \Lambda_{h_\alpha,g_\gamma}(F_1, F_1, F_1, F_1) \frac{d\gamma}{\gamma^3}.$$

Lemma 3 with $(\rho_1, \sigma_1) = (g_\alpha, h_\alpha)$ and $(\rho_2, \sigma_2) = (g_\gamma, h_\gamma)$ gives

$$\Lambda_{h_\alpha,g_\gamma}(F_1, F_1, F_1, F_1) = -\Lambda_{g_\alpha,h_\gamma}(F_1, F_1, F_1, F_1) + \int_{\mathbb{R}^2} F_1^4.$$

The key gain we obtain from having reduced to a single function F_1 is that

$$\Lambda_{g_\alpha,h_\gamma}(F_1, F_1, F_1, F_1) \geq 0, \tag{3.12}$$

which can be seen by writing the form in (3.12) in an analogous way as in (3.8) and using positivity of g_α. By our normalization, $\int_{\mathbb{R}^2} F_1^4 = 1$. Thus,

$$\Lambda_{h_\alpha,g_\gamma}(F_1, F_1, F_1, F_1) \leq 1.$$

This establishes the desired estimate for $\widetilde{\Lambda}_{\varphi^{(u)},\varphi^{(-u)},\psi^{(v)},\psi^{(-v)}}$.

4 Appendix

In this appendix we give the following remaining proof.

Proof of Lemma 2. We construct a function f which has the prescribed behavior near the endpoints of its support, so that the considered square roots are evidently smooth. The construction essentially consists of algebraic manipulations of $\varphi(x) := e^{-\frac{1}{x}}\mathbf{1}_{(0,\infty)}(x)$.

Consider the function

$$g(x) := c\,\varphi_1(x)\varphi_2(2 - \frac{2}{\varepsilon}x),$$

where φ_1 and φ_2 are defined as

$$\varphi_1(x) := ((3-x)\varphi'(x))' \qquad \text{and} \qquad \varphi_2(x) := \frac{\varphi(x)}{\varphi(x)+\varphi(1-x)}.$$

The constant $c > 0$ is chosen such that $\int_{\mathbb{R}} g = 1$. The function g is smooth, non-negative and supported on $[0, \varepsilon]$. Since φ_2 equals 1 for $x \geq 1$, for $\delta := \varepsilon/2$ we have

$$g = c\,\varphi_1 \;\text{ on }\; (-\infty, \delta).$$

The factor $(3 - x)$ in the definition of φ_1 will be convenient when investigating (2.2).

We consider the antiderivative

$$f(x) := \int_{-\infty}^{x} g(t-1) - g(3-t)dt,$$

which is smooth and even about $x = 2$, i.e. $f(x) = f(4 - x)$. Moreover, it is supported on $[1, 3]$, positive on $(1, 3)$ and constantly equals 1 on $[1 + \varepsilon, 3 - \varepsilon]$. We have

$$f(x) = c\,(4-x)\varphi'(x-1) \;\text{ on }\; (-\infty, 1+\delta). \tag{4.1}$$

Thus, $f^{1/2}$ is smooth at $x = 1$. Smoothness at $x = 3$ follows by symmetry.

Consider the integral in (2.2), which is due to oddness of the integrand equal to

$$h(x) := \int_{-\infty}^{x} -\frac{f(t)+f(-t)}{t}dt.$$

The function h is even, supported on $[-3, 3]$ and positive on $(-3, 3)$. Using $f(-t) = f(t + 4)$ and (4.1) we see that

$$h(x) = c\,\varphi(x+3) \;\text{ on }\; (-\infty, -3+\delta).$$

This shows smoothness of $h^{1/2}$ at $x = -3$. By symmetry the same holds at $x = 3$, which establishes the claim of the lemma. □

References

[1] F. Bernicot, *Fiber-wise Calderón-Zygmund decoposition and application to a bi-dimensional paraproduct*, Illinois J. Math. **56** (2012), no. 2, 415-422.

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

[3] R. L. Jones, A. Seeger, J. Wright, *Strong variational and jump inequalities in harmonic analysis*, Trans. Amer. Math. Soc., **360** (2008), no. 12, 67116742.

[4] V. Kovač, *Bellman function technique for multilinear estimates and an application to generalized paraproducts*, Indiana Univ. Math. J., **60** (2011), no. 3, 813-846.
[5] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam., **28** (2012), no. 4, 1143-1164.
[6] C. Thiele, *Wave packet analysis*, CBMS Reg. Conf. Ser. Math., **105**, AMS, Providence, RI, 2006.

POLONA DURCIK, UNIVERSITÄT BONN, ENDENICHER ALLEE 60, 53115 BONN, GERMANY
E-mail address: durcik@math.uni-bonn.de

Chapter 3

L^p estimates for a singular entangled quadrilinear form

L^p estimates for a singular entangled quadrilinear form

Polona Durcik

Abstract

We prove L^p estimates for a continuous version of a dyadic quadrilinear form introduced by Kovač in [6]. This improves the range of exponents from the prequel [3] of the present paper.

1 Introduction

This article is a continuation of [3]. We are concerned with a quadrilinear singular integral form involving the *entangled* product of four functions on $\mathbb{R}^2$

$$\boldsymbol{F}(F_1, F_2, F_3, F_4)(x, y, x', y') := F_1(x, y)F_2(x', y)F_3(x', y')F_4(x, y').$$

For Schwartz functions $F_j \in \mathcal{S}(\mathbb{R}^2)$, the form is given by

$$\Lambda(F_1, F_2, F_3, F_4) := \int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi, \eta, -\xi, -\eta)m(\xi, \eta)d\xi d\eta,$$

where $\boldsymbol{F} := \boldsymbol{F}(F_1, F_2, F_3, F_4)$ and m is a bounded function on $\mathbb{R}^2$, smooth away from the origin. For all multi-indices α up to some large finite order it satisfies[1]

$$|\partial^\alpha m(\xi, \eta)| \lesssim (|\xi| + |\eta|)^{-|\alpha|}.$$

In [3] it is shown that

$$|\Lambda(F_1, F_2, F_3, F_4)| \lesssim \|F_1\|_{\mathrm{L}^4(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^4(\mathbb{R}^2)}\|F_3\|_{\mathrm{L}^4(\mathbb{R}^2)}\|F_4\|_{\mathrm{L}^4(\mathbb{R}^2)}. \tag{1.1}$$

Our present goal is to prove L^p estimates for Λ in a larger range of exponents.

Theorem 1. *For $F_1, F_2, F_3, F_4 \in \mathcal{S}(\mathbb{R}^2)$, the quadrilinear form Λ satisfies*

$$|\Lambda(F_1, F_2, F_3, F_4)| \lesssim_{(p_j)} \|F_1\|_{\mathrm{L}^{p_1}(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^{p_2}(\mathbb{R}^2)}\|F_3\|_{\mathrm{L}^{p_3}(\mathbb{R}^2)}\|F_4\|_{\mathrm{L}^{p_4}(\mathbb{R}^2)}$$

whenever $\sum_{j=1}^4 \frac{1}{p_j} = 1$ and $2 < p_j \leq \infty$ for all j.

This theorem is a consequence of the restricted type estimates given by Theorem 3 below. By the decomposition performed in [3], it suffices to prove Theorem 1 for m

2010 *Mathematics Subject Classification.* Primary 42B15; Secondary 42B20.

[1] We write $A \lesssim B$ if there is an absolute constant $C > 0$ such that $A \leq CB$. If P depends on a set of parameters P, we write $A \lesssim_P B$. We write $A \sim B$ if both $A \lesssim B$ and $B \lesssim A$.

reduced to a single cone in the frequency plane (ξ,η). More precisely, it is enough to consider the form

$$\int_0^\infty \mu_t \int_{\mathbb{R}^2} \widehat{\boldsymbol{F}}(\xi,\eta,-\xi,-\eta)\widehat{\varphi^{(u)}}(t\xi)\widehat{\psi^{(v)}}(t\eta)\widehat{\varphi^{(-u)}}(-t\xi)\widehat{\psi^{(-v)}}(-t\eta)d\xi d\eta\frac{dt}{t} \tag{1.2}$$

where $\varphi^{(u)}(x) = (1+|u|)^{-25}\varphi(x-u)$ and $\psi^{(v)}(x) = (1+|v|)^{-10}\psi(x-v)$. The functions $\varphi,\psi \in \mathcal{S}(\mathbb{R})$ are real-valued and ψ is such that $(\int_\eta^\infty |\widehat{\psi}(\tau)|^2 d\tau/\tau)^{1/2}$ belongs to $\mathcal{S}(\mathbb{R})$, $u,v\in\mathbb{R}$ and μ_t are measurable coefficients with $|\mu_t|\le 1$. We remark that the decomposition is not explicitly stated in this manner in [3], but it follows by a minor rephrasing of the arguments. The estimate for (1.2) will be uniform in the parameters u,v.

Since the integral of the Fourier transform of a Schwartz function over a hyperplane in $\mathbb{R}^4$ equals the integral of the function itself over the perpendicular hyperplane, we can express the form (1.2) as

$$\int_0^\infty \mu_t \int_{\mathbb{R}^2} \boldsymbol{F} * [\varphi^{(u)}\otimes\psi^{(v)}\otimes\varphi^{(-u)}\otimes\psi^{(-v)}]_t(p,q,p,q)dpdq\frac{dt}{t},$$

where $(f_1\otimes\cdots\otimes f_n)(x_1,\dots x_n) := f_1(x_1)\dots f_n(x_n)$ and $[f]_t(x_1,\dots,x_n) := t^{-n}f(t^{-1}x)$. We truncate in the scale t, that is, for $N>0$ we consider $\Lambda^N_{\varphi,\psi} = \Lambda^N_{\varphi,\psi,\mu,u,v}$ given by

$$\Lambda^N_{\varphi,\psi}(F_1,F_2,F_3,F_4) := \int_{2^{-N}}^{2^N} \mu_t \int_{\mathbb{R}^2} \boldsymbol{F} * [\varphi^{(u)}\otimes\psi^{(v)}\otimes\varphi^{(-u)}\otimes\psi^{(-v)}]_t(p,q,p,q)dpdq\frac{dt}{t},$$

which is well defined for bounded measurable functions F_j with finite measure support. We have the following analogue of Theorem 1 for $\Lambda^N_{\varphi,\psi}$.

Theorem 2. *For bounded measurable functions F_1,F_2,F_3,F_4 with finite measure support, the quadrilinear form $\Lambda^N_{\varphi,\psi}$ satisfies the estimate*

$$|\Lambda^N_{\varphi,\psi}(F_1,F_2,F_3,F_4)| \lesssim_{(p_j)} \|F_1\|_{\mathrm{L}^{p_1}(\mathbb{R}^2)}\|F_2\|_{\mathrm{L}^{p_2}(\mathbb{R}^2)}\|F_3\|_{\mathrm{L}^{p_3}(\mathbb{R}^2)}\|F_4\|_{\mathrm{L}^{p_4}(\mathbb{R}^2)} \tag{1.3}$$

whenever $\sum_{j=1}^4 \frac{1}{p_j} = 1$ and $2 < p_j \le \infty$ for all j.

The bound (1.3) is independent of N,u,v. Approximating $F_j\in\mathcal{S}$ in L^{p_j} with smooth compactly supported functions, Theorem 2 then implies Theorem 1. By the multilinear interpolation and the restricted type theory discussed in [10], Theorem 2 is a consequence of the following (generalized) restricted type estimates.

Theorem 3. *For $j=1,2,3,4$, let $E_j\subseteq\mathbb{R}^2$ be a set of finite measure. Let k be the largest index such that $|E_k|$ is maximal among the $|E_j|$. Then there exists a subset $E_k'\subseteq E_k$ with $2|E_k'|\ge|E_k|$, such that for any four measurable functions F_j with[2] $|F_j|\le \mathbf{1}_{E_j}$ for all j and $|F_k|\le\mathbf{1}_{E_k'}$ we have the estimate*

$$|\Lambda^N_{\varphi,\psi}(F_1,F_2,F_3,F_4)| \lesssim |E_1|^{\alpha_1}|E_2|^{\alpha_2}|E_3|^{\alpha_3}|E_4|^{\alpha_4}$$

whenever $\sum_{j=1}^4\alpha_j = 1$ and $-1/2\le\alpha_j\le 1/2$ for all j.

[2]By $\mathbf{1}_A$ we denote the characteristic function of a set $A\subseteq\mathbb{R}^2$.

Negative exponents α_j correspond to quasi-Banach space estimates for the dual operators of $\Lambda^N_{\varphi,\psi}$, for which one may consult [10].

Assuming Theorem 1, we now mention how to extend Λ to a bounded operator on $\mathrm{L}^{p_1} \times \mathrm{L}^{p_2} \times \mathrm{L}^{p_3} \times \mathrm{L}^{p_4}$ whenever p_j are as in Theorem 1. If $p_j < \infty$ for all j, this follows by density of $\mathcal{S}$ in L^{p_j}. If $p_j = \infty$ for some j, we argue by duality. Note that have at most one exponent equal to ∞. We sketch the argument when $p_4 = \infty$, the other instances following by symmetry of the form. We know that there is an operator T mapping $\mathrm{L}^4 \times \mathrm{L}^4 \times \mathrm{L}^4$ to $\mathrm{L}^{4/3}$ such that

$$\Lambda(F_1, F_2, F_3, F_4) = \int T(F_1, F_2, F_3)F_4.$$

We claim that for $F_j \in \mathcal{S}$, $\|T(F_1, F_2, F_3)\|_{\mathrm{L}^1} \lesssim \|F_1\|_{\mathrm{L}^{p_1}} \|F_2\|_{\mathrm{L}^{p_2}} \|F_3\|_{\mathrm{L}^{p_3}}$. Then Λ can be defined on $\mathcal{S} \times \mathcal{S} \times \mathcal{S} \times \mathrm{L}^\infty$ and density arguments yield a bounded extension on $\mathrm{L}^{p_1} \times \mathrm{L}^{p_2} \times \mathrm{L}^{p_3} \times \mathrm{L}^\infty$. To see the claim we write

$$\|T(F_1, F_2, F_3)\|_{\mathrm{L}^1([-M,M]^2)} = \int T(F_1, F_2, F_3)\vartheta$$

where ϑ is a modulation times $\mathbf{1}_{[-M,M]^2}$. Then we approximate ϑ weakly in L^4 with smooth compactly supported functions having L^∞ norms uniformly bounded by 1. Applying Theorem 1 for the tuple (p_1, p_2, p_3, ∞) yields the assertion.

Let us briefly comment on the form Λ. For more extensive motivation we refer to [3]. The instance of Λ which was first considered is the trilinear form[3] $\Lambda_1(F_1, F_2, F_3) := \Lambda(F_1, F_2, F_3, 1)$. It was introduced by Demeter and Thiele [2]. This trilinear form can also be seen as a simpler version of the twisted paraproduct proposed by Camil Muscalu and sometimes one refers to it with that name as well.

Boundedness of Λ_1 was established by Kovač [6], who first investigated a dyadic model of Λ for a general function F_4 by an induction on scales type argument. See also [5]. This led to an estimate for a dyadic version of Λ_1 whenever $2 < p_1, p_2, p_3 < \infty$ and $1/p_1 + 1/p_2 + 1/p_3 = 1$. Then Kovač passed to the bound for Λ_1 using the square functions of Jones, Seeger and Wright [4]. Bernicot's fiber-wise Calderón-Zygmund decomposition [1] extended the range of exponents to $1 < p_1, p_3 < \infty$, $2 < p_2 \leq \infty$. The transition to the continuous case and the extension of the exponent range both relied on the special structure arising from $F_4 = 1$.

For the quadrilinear form with a general fourth function, the L^4 estimate (1.1) was derived by adapting the induction of scales techique by Kovač to the continuous setting. In the present article we prove estimates in a larger range of exponents by extending his method to the continuous localized context.

By a classical stopping time argument, Theorem 3 is reduced to estimating entangled forms of the type

$$\int_\Omega |\boldsymbol{F} * [\varphi^{(u)} \otimes \psi^{(v)} \otimes \varphi^{(-u)} \otimes \psi^{(-v)}]_t(p, q, p, q)| dp dq \frac{dt}{t}.$$

Here Ω is a certain local region in the upper half space with "regular" boundary. Controlling such objects with the technique from [6] requires an algebraic telescoping identity.

[3]In [3] we called this form T, not to be interchanged with the dual operator introduced above.

In [3], its derivation relies on an identity involving the Fourier transform. The argument is of global nature and we cannot directly repeat it in the localized setting.

We obtain the desired telescoping element in Proposition 8 in Section 2. To overcome the mentioned difficulty, we first restrict the functions F_j to certain projections of the region Ω. This allows us to discard the spatial localization of the form and proceed in the manner of [3]. The issue in the described process is then in estimating boundary terms, representing differences between local and global objects. This requires certain control of the boundary and is carried out in Lemma 6 and Lemma 7 below. Our approach has been inspired by Muscalu, Tao and Thiele [7].

To conclude we remark that in general we do not know of any arguments which could extend the range of exponents from Theorem 1 to $p_j \leq 2$.

Acknowledgement. I would like to express my sincere gratitude to my advisor Prof. Christoph Thiele for his guidance and support throughout this project.

2 Local telescoping

First let us set up some notation. A *dyadic interval* is a interval of the form $[2^k m, 2^k(m+1)]$ for some $k, m \in \mathbb{Z}$. We denote the set of all dyadic intervals by $\mathcal{I}$ and the set of all dyadic intervals of length 2^k by $\mathcal{I}_k$. A *dyadic square* is the Cartesian product of two dyadic intervals of the same length. For a dyadic square S we denote by $\ell(S)$ its sidelength. We write $\mathcal{D}$ for the set of all dyadic squares and $\mathcal{D}_k$ for the set of all dyadic squares of sidelength 2^k. Each $S \in \mathcal{D}$ is divided into four congruent dyadic squares of half the sidelength, called the *children* of S. Conversely, each square in $\mathcal{D}$ has a unique *parent* in $\mathcal{D}$. Given any two dyadic squares, either one is contained in the other or they are almost disjoint, by which we mean that their intersection has Lebesgue measure zero.

As in [6], we collect the squares into units called *trees*. A finite collection $\mathcal{T} \subseteq \mathcal{D}$ is called a *tree* if there exists a square $R_{\mathcal{T}} \in \mathcal{T}$ called the *root*, satisfying $S \subseteq R_{\mathcal{T}}$ for every $S \in \mathcal{T}$. A tree is called *convex* if for all S_1, S_2, S_3 we have that $S_1 \subseteq S_2 \subseteq S_3$ and $S_1,\ S_3 \in \mathcal{T}$ imply $S_2 \in \mathcal{T}$. A *leaf* of $\mathcal{T}$ is a dyadic square which is not contained in $\mathcal{T}$, but its parent is. We denote the set of leaves of $\mathcal{T}$ by $\mathcal{L}(\mathcal{T})$. Note that the leaves of a convex tree partition its root. We split $\mathcal{T}$ into generations of squares of sidelength 2^k. For this we denote

$$\mathcal{T}_k := \mathcal{T} \cap \mathcal{D}_k \quad \text{and} \quad \mathcal{T}_k^c := \mathcal{D}_k \setminus \mathcal{T}_k.$$

For the union of all squares in $\mathcal{T}_k$ we write

$$T_k := \bigcup_{S \in \mathcal{T}_k} S.$$

Observe that for a convex tree $\mathcal{T}$ we have $T_k \subseteq T_{k'}$ if $k \leq k'$, $T_{k'} \neq \emptyset$.

The following lemma measures the size of the boundary of T_k. It is a variant of Lemma 4.8 from [7]. It estimates the cardinality of dyadic points

$$\Delta(\mathcal{T}_k) := \partial T_k \cap (2^k\mathbb{Z} \times 2^k\mathbb{Z}),$$

where ∂T_k denotes the topological boundary of $T_k \subseteq \mathbb{R}^2$. Note that ∂T_k is the union of all dyadic line segments in

$$\big\{[p, p+2^k] \times \{q\} \subseteq T_k : (p,q) \in 2^k\mathbb{Z} \times 2^k\mathbb{Z}, [p, p+2^k] \times [q-2^{k-1}, q+2^{k-1}] \not\subseteq T_k\big\}$$
$$\cup \big\{\{p\} \times [q, q+2^k] \subseteq T_k : (p,q) \in 2^k\mathbb{Z} \times 2^k\mathbb{Z}, [p-2^{k-1}, p+2^{k-1}] \times [q, q+2^k] \not\subseteq T_k\big\}.$$

Lemma 4. *For any convex tree $\mathcal{T}$ we have*

$$\sum_{k\in\mathbb{Z}} 2^{2k} \# \Delta(\mathcal{T}_k) \lesssim |R_{\mathcal{T}}|.$$

Proof. It suffices to prove the claim for all dyadic points $(p,q) \in \partial T_k$ such that $[p-2^k, p] \times [q-2^k, q] \notin \mathcal{T}_k$. For each such point consider the dyadic square

$$S(p,q,k) := [p-2^k, p-2^{k-1}] \times [q-2^k, q-2^{k-1}]$$

which has area $2^{2(k-1)}$. We claim that squares of this form are pairwise almost disjoint. This will prove the lemma, as they are contained in $3R_{\mathcal{T}}$.

To see the claim, suppose that $S(p,q,k)$ and $S(p',q',k')$ intersect in a set of positive measure. If $k = k'$, then they must coincide since they are dyadic and of the same scale. So suppose that $k < k'$, hence $S(p,q,k)$ is contained in $S(p',q',k')$. Then the point (p,q) is contained in the interior of $[p'-2^{k'}, p'] \times [q'-2^{k'}, q']$, which is disjoint from $T_{k'}$. This shows that $(p,q) \in T_k$ but $(p,q) \notin T_{k'}$, contradicting convexity of $\mathcal{T}$. $\square$

With any collection of dyadic squares $\mathcal{C} \subseteq \mathcal{D}$ we associate a region in the upper half space $\mathbb{R}^3_+$. The region consists of Whitney boxes associated with $S \in \mathcal{C}$ and is defined by

$$\Omega_{\mathcal{C}} := \bigcup_{S\in\mathcal{C}} S \times \Big[\frac{\ell(S)}{2}, \ell(S)\Big].$$

The case $\mathcal{C} = \mathcal{T}$ for a convex tree $\mathcal{T}$ is depicted in Figure 1. Observe that $\Omega_{\mathcal{T}} = \cup_{k\in\mathbb{Z}} \Omega_{\mathcal{T}_k} = \cup_{k\in\mathbb{Z}} T_k \times [2^{k-1}, 2^k]$.

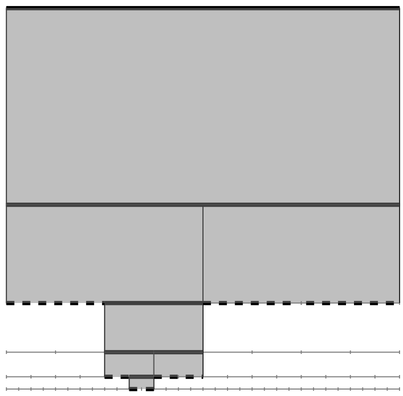

Figure 1: Projection of $\Omega_{\mathcal{T}}$ on $\mathbb{R}^2_+$. The bold lines represent $S \times \ell(S)$ for $S \in \mathcal{T}$, while the dotted lines correspond to $S \in \mathcal{L}(\mathcal{T})$.

Throughout the text, all two-dimensional functions will be measurable, bounded, with finite measure support and positive. Denote

$$\theta(x,y) := (1 + |(x,y)|^4)^{-1}.$$

For a function F on $\mathbb{R}^2$ and $\mathcal{C} \subseteq \mathcal{D}$ we define

$$M(F,\mathcal{C}) := \sup_{(p,q,t)\in\Omega_{\mathcal{C}}} (F^2 * [\theta]_t(p,q))^{1/2}.$$

Denote also

$$\vartheta(x) := (1 + |x|)^{-4}.$$

Now we consider a continuous variant of the Gowers box inner product used in [6]. The following estimate joins a version of the box Cauchy-Schwarz inequality and an estimate of the Gowers box norm by an L^2-type average. This is the reason for the restricted range of exponents in Theorem 1.

Lemma 5. *For $(p,q,t) \in \Omega_{\mathcal{C}}$ we have*

$$\boldsymbol{F} * [\vartheta \otimes \vartheta \otimes \vartheta \otimes \vartheta]_t(p,q,p,q) \le \prod_{j=1}^{4} M(F_j, \mathcal{C}). \tag{2.1}$$

Proof. Denote the left-hand side of (2.1) by $A^{(p,q,t)}(F_1,F_2,F_3,F_4)$ and rewrite it as

$$\int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}} F_1(x,y)F_2(x',y)[\vartheta]_t(q-y)dy \Big) \Big(\int_{\mathbb{R}} F_3(x',y')F_4(x,y')[\vartheta]_t(q-y')dy' \Big) \\ [\vartheta]_t(p-x)[\vartheta]_t(p-x')dxdx' \tag{2.2}$$

Now we apply the Cauchy-Schwarz inequality with respect to $[\vartheta]_t(p-x)dx$, $[\vartheta]_t(p-x')dx'$, which bounds this term by

$$A^{(p,q,t)}(F_1,F_2,F_2,F_1)^{1/2} A^{(p,q,t)}(F_4,F_3,F_3,F_4)^{1/2}.$$

By symmetry in (p,q) it follows that

$$A^{(p,q,t)}(F_1,F_2,F_2,F_1) \le A^{(p,q,t)}(F_1,F_1,F_1,F_1)^{1/2} A^{(p,q,t)}(F_2,F_2,F_2,F_2)^{1/2}.$$

Now we write $A^{(p,q,t)}(F_j,F_j,F_j,F_j)$ in the same way as in (2.2) and apply the Cauchy-Schwarz inequality with respect to dy, dy'. This yields

$$A^{(p,q,t)}(F_j,F_j,F_j,F_j) \le (F_j^2 * [\vartheta\otimes\vartheta]_t(p,q))^2 \le (F_j^2 * [\theta]_t(p,q))^2,$$

which proves the claim. □

With functions $\phi_j \in \mathrm{L}^1(\mathbb{R})$, $j = 1,2,3,4$, and $\mathcal{C} \subseteq \mathcal{D}$ we associate the local form

$$\Theta^{\mathcal{C}}_{\phi_1,\phi_2,\phi_3,\phi_4}(F_1,F_2,F_3,F_4) := \int_{\Omega_{\mathcal{C}}} \boldsymbol{F} * [\phi_1 \otimes \phi_2 \otimes \phi_3 \otimes \phi_4]_t(p,q,p,q)dpdq\frac{dt}{t}.$$

To shorten the notation we write $\Theta^{\mathcal{C}}_{\phi_1,\phi_3} := \Theta^{\mathcal{C}}_{\phi_1,\phi_3,\phi_1,\phi_3}$.

The following two complementary lemmas will be used to control error and boundary terms in Proposition 8.

Lemma 6. *For a convex tree $\mathcal{T}$ we have*

$$\sum_{k\in\mathbb{Z}} \Theta^{\mathcal{T}_k}_{\vartheta^2,\vartheta^2}(F_1\mathbf{1}_{T_k^c}, F_2, F_3, F_4) \lesssim |R_{\mathcal{T}}| \prod_{j=1}^{4} M(F_j, \mathcal{T}). \tag{2.3}$$

Observe that by symmetry of (2.3), the same result holds under any permutation of the arguments $F_1\mathbf{1}_{T_k^c}, F_2, F_3, F_4$.

Proof. For $k \in \mathbb{Z}$ and $t \in [2^{k-1}, 2^k]$ we consider

$$\int_{T_k}\int_{\mathbb{R}^4} \boldsymbol{F}(F_1\mathbf{1}_{T_k^c}, F_2, F_3, F_4)(x, y, x', y')[\vartheta\otimes\vartheta\otimes\vartheta\otimes\vartheta]_t(p-x, q-y, p-x', q-y')$$
$$\vartheta\otimes\vartheta(t^{-1}(p-x, q-y))\vartheta\otimes\vartheta(t^{-1}(p-x', q-y'))dxdydx'dy'dpdq. \tag{2.4}$$

Note that (2.3) is obtained by integrating this term in $t \in [2^{k-1}, 2^k]$ and summing over $k \in \mathbb{Z}$. We claim that for $(x, y) \in T_k^c$ and $(p, q) \in T_k$ there is a point (a, b) contained in

$$B(p, q) := \{(p', q') \in \partial T_k : p' = p \text{ or } q' = q\} \cup \Delta(\mathcal{T}_k) \tag{2.5}$$

such that $|(p, q) - (x, y)| \geq |(p, q) - (a, b)|$.

This can be seen as follows. By E we denote the intersection of ∂T_k and the line segment between (p, q) and (x, y). If E contains dyadic points from $\Delta(\mathcal{T}_k)$, we may set (a, b) to be any of these points. Otherwise, E must contain a point of the form $(p', q'+\alpha)$ or $(p'+\alpha, q')$ for some $p', q' \in 2^k\mathbb{Z}$, $\alpha \in (0, 2^k)$. Assume it contains at least one of the form $(p', q'+\alpha)$. For definiteness pick the one with the the least distance to (p, q). In case $q' < q < q' + 2^k$ we know that $(p', q) \in \partial T_k$ and we set $(a, b) = (p', q)$. If $q < q'$, we set $(a, b) = (p', q') \in \Delta(\mathcal{T}_k)$. In case $q > q' + 2^k$ we choose $(p', q' + 2^k) \in \Delta(\mathcal{T}_k)$. Analogously we proceed in the remaining case, that is, if E consists only of points $(p' + \alpha, q')$.

Since $\vartheta\otimes\vartheta \leq \theta$ and θ is radially decreasing, we have for $(p, q), (x, y), (a, b)$ as above

$$\vartheta\otimes\vartheta(t^{-1}(p-x, q-y)) \leq \theta(t^{-1}(p-a, q-b)) \leq \sum_{(a,b)\in B(p,q)} \theta(t^{-1}(p-a, q-b)).$$

Estimating $\vartheta\otimes\vartheta(t^{-1}(p-x', q-y')) \leq 1$, the term (2.4) is bounded by

$$\int_{T_k} \boldsymbol{F}(F_1\mathbf{1}_{T_k^c}, F_2, F_3, F_4) * [\vartheta\otimes\vartheta\otimes\vartheta\otimes\vartheta]_t(p, q, p, q) \sum_{(a,b)\in B(p,q)} \theta(t^{-1}(p-a, q-b))dpdq.$$

Applying Lemma 5, the last display is no greater than

$$\Big(M(F_1\mathbf{1}_{T_k^c}, \mathcal{T}_k)\prod_{j=2}^{4} M(F_j, \mathcal{T}_k)\Big)\int_{T_k}\sum_{(a,b)\in B(p,q)} \theta(t^{-1}(p-a, q-b))dpdq.$$

Observe that by homogeneity of the inequality (2.3) we may assume $M(F_j, \mathcal{T}) = 1$ for all j. Due to this fact and by symmetry in p, q, it suffices to further estimate

$$\sum_{Q\in\mathcal{I}_k}\int_Q \sum_{a:\{a\}\times Q\subseteq\partial T_k}\int_{\mathbb{R}} \theta(t^{-1}(p-a, 0))dpdq + \sum_{(a,b)\in\Delta(\mathcal{T}_k)}\int_{\mathbb{R}^2} \theta(t^{-1}(p-a, q-b))dpdq.$$

Integrating the function θ, the last display is estimated by a constant times

$$\sum_{Q\in\mathcal{I}_k}\int_Q t\,\#\{a:\{a\}\times Q\subseteq\partial T_k\}+t^2\#\Delta(\mathcal{T}_k)\lesssim 2^{2k}\#\Delta(\mathcal{T}_k).$$

Therefore, up to a constant, (2.3) is bounded by

$$\sum_{k\in\mathbb{Z}}\int_{2^{k-1}}^{2^k}2^{2k}\#\Delta(\mathcal{T}_k)\frac{dt}{t}\lesssim\sum_{k\in\mathbb{Z}}2^{2k}\#\Delta(\mathcal{T}_k)\lesssim|\mathcal{R}_{\mathcal{T}}|,$$

which is the desired result in view of the normalization $M(F_j,\mathcal{T})=1$. The last inequality follows from Lemma 4. □

Lemma 7. *For a convex tree $\mathcal{T}$ we have*

$$\sum_{k\in\mathbb{Z}}\Theta^{\mathcal{T}_k^c}_{\vartheta^2,\vartheta^2}(F_1\mathbf{1}_{T_k},F_2\mathbf{1}_{T_k},F_3\mathbf{1}_{T_k},F_4\mathbf{1}_{T_k})\lesssim|R_{\mathcal{T}}|\prod_{j=1}^4 M(F_j,\mathcal{T}).\tag{2.6}$$

Proof. Proceeding in the exact same way as in the proof of Lemma 6 we see that the left-hand side of (2.6) is bounded by

$$\begin{aligned}&\sum_{k\in\mathbb{Z}}\Big(\prod_{j=1}^4 M(F_j\mathbf{1}_{T_k},\mathcal{T}_k^c)\Big)\int_{T_k^c}\sum_{(a,b)\in B(p,q)}\theta(t^{-1}(p-a,q-b))dpdq\\ \lesssim&\sum_{k\in\mathbb{Z}}\Big(\prod_{j=1}^4 M(F_j\mathbf{1}_{T_k},\mathcal{T}_k^c)\Big)\,2^{2k}\#\Delta(\mathcal{T}_k),\end{aligned}$$

where $B(p,q)$ is defined as in (2.5). We claim that for each j we have

$$M(F_j\mathbf{1}_{T_k},\mathcal{T}_k^c)\lesssim M(F_j,\mathcal{T}).$$

Together with an application of Lemma 4 this will finish the proof.

The claim can be rephrased as follows: for each $(p,q)\in T_k^c$ we have

$$(F_j^2\mathbf{1}_{T_k}*[\theta]_t(p,q))^{1/2}\lesssim M(F_j,\mathcal{T}).$$

First we set $(p,q)=0$ without loss of generality. Also, we may assume that T_k is contained in the quadrant $\{(p,q):p\geq 0,\,q\geq 0\}$, as otherwise we restrict T_k to each of the four quadrants and all parts are treated in the same way. Denote

$$r:=\min_{(a,b)\in\partial T_k}|(a,b)|.$$

Take any point (a,b) which minimizes the distance and consider the closed cone C in $\mathbb{R}^2$ with vertex 0 and aperture $\pi/2$, its axis being the line spanned by (a,b). Observe that each $(x,y)\in T_k\cap C$ satisfies $|(x,y)|\geq|(x,y)-(a,b)|$ and thus $\theta(x,y)\leq\theta(x-a,y-b)$. If $T_k\setminus C\neq\emptyset$, then we iterate with T_k replaced by $T_k\setminus C$. We find a point $(a',b')\in\partial T_k\cap\partial(T_k\setminus C)$ and a cone C' such that for each $(x,y)\in(T_k\setminus C)\cap C'$ we have $|(x,y)|\geq|(x,y)-(a',b')|$ and so $\theta(x,y)\leq\theta(x-a',y-b')$. Since $C\cup C'$ covers T_k, for each $(x,y)\in T_k$ we have

$$\theta(x,y)\leq\theta(a-x,b-y)+\theta(a'-x,b'-y).$$

Therefore,

$$(F_j^2 \mathbf{1}_{T_k} * [\theta]_t(0))^{1/2} \lesssim (\sup_{(a,b,t)\in\Omega_{\mathcal{T}_k}} F_j^2 \mathbf{1}_{T_k} * [\theta]_t(a,b))^{1/2} \leq \sup_{(a,b,t)\in\Omega_{\mathcal{T}}} (F_j^2 * [\theta]_t(a,b))^{1/2}$$

as desired. □

For a function $f \in \mathcal{S}(\mathbb{R})$ we consider the Schwartz seminorm

$$\|f\| := \sup_{x\in\mathbb{R}} (1+|x|)^8 |f(x)| + (1+|x|)^9 |f'(x)|.$$

Now we are ready to state the estimate which will take the place of the telescoping identities used in [6], [3].

Proposition 8. *Let (ρ_i, σ_i) be two pairs of real-valued Schwartz functions which satisfy*

$$-t\partial_t |\widehat{\rho_i}(t\tau)|^2 = |\widehat{\sigma_i}(t\tau)|^2. \tag{2.7}$$

Then we have for any convex tree $\mathcal{T}$

$$\Theta^{\mathcal{T}}_{\rho_1,\sigma_2}(F_1,F_2,F_3,F_4) + \Theta^{\mathcal{T}}_{\sigma_1,\rho_2}(F_1,F_2,F_3,F_4) \lesssim_c |R_{\mathcal{T}}| \prod_{j=1}^{4} M(F_j,\mathcal{T}), \tag{2.8}$$

where $c = \|\rho_1\|^2\|\sigma_2\|^2 + \|\sigma_1\|^2\|\rho_2\|^2 + \|\rho_1\|^2\|\rho_2\|^2$.

Examples of functions satisfying (2.7) include Gaussian exponential functions and their derivatives such as $(\rho,\sigma) = (e^{-x^2}, -2xe^{-x^2})$. Another example is a pair (ρ,σ) where ρ is defined via $\widehat{\rho}(\xi) := (\int_\xi^\infty |\widehat{\sigma}(\tau)|^2 d\tau/\tau)^{1/2}$ and $|\widehat{\sigma}|^2 \in C_0^\infty(\mathbb{R})$ is even, vanishes at zero and is well behaved near the boundary of its support so that $\widehat{\rho}$ belongs to $C_0^\infty(\mathbb{R})$. An instance of such a pair has been explicitly constructed in [3].

Proof. By homogeneity of (2.8) we may assume $M(F_j,\mathcal{T}) = 1$ for all j. By scaling invariance we may suppose $|R_{\mathcal{T}}| = 1$. Thus, we are set to establish

$$\Theta^{\mathcal{T}}_{\rho_1,\sigma_2}(F_1,F_2,F_3,F_4) + \Theta^{\mathcal{T}}_{\sigma_1,\rho_2}(F_1,F_2,F_3,F_4) \lesssim_c 1. \tag{2.9}$$

Denote $\Psi := \rho_1 \otimes \rho_2 \otimes \rho_1 \otimes \rho_2$. By the fundamental theorem of calculus we have

$$[\Psi]_{2^{k-1}} - [\Psi]_{2^k} = \int_{2^{k-1}}^{2^k} (-t\partial_t[\Psi]_t) \frac{dt}{t}. \tag{2.10}$$

We convolve the equality (2.10) with $\boldsymbol{F}$ and evaluate the convolution at (p,q,p,q). Then we integrate in (p,q) over T_k and sum over $k \in \mathbb{Z}$. Writing T_k as the almost disjoint union of $S \in \mathcal{T}_k$, the left-hand side of (2.10) becomes

$$L := \sum_{k\in\mathbb{Z}} \sum_{S\in\mathcal{T}_k} \Big(\sum_{S' \text{ child of } S} \int_{S'} \boldsymbol{F} * [\Psi]_{\ell(S')}(p,q,p,q) dpdq - \int_S \boldsymbol{F} * [\Psi]_{\ell(S)}(p,q,p,q) dpdq \Big)$$

Since $\mathcal{T}$ is convex, each square $S \in \mathcal{T} \setminus \{R_\mathcal{T}\}$ has all four children S' in $\mathcal{T} \cup \mathcal{L}(\mathcal{T})$. Thus, the last display is a telescoping sum which equals

$$\sum_{S \in \mathcal{L}(\mathcal{T})} \int_S \boldsymbol{F} * [\Psi]_{\ell(S)}(p,q,p,q)dpdq - \int_{R_\mathcal{T}} \boldsymbol{F} * [\Psi]_{\ell(R_\mathcal{T})}(p,q,p,q)dpdq.$$

We bound $|\Psi| \lesssim_c \vartheta^2 \otimes \vartheta^2 \otimes \vartheta^2 \otimes \vartheta^2$ and apply Lemma 5. This yields

$$|L| \lesssim_c \Big(\sum_{S \in \mathcal{L}(\mathcal{T})} |S| + 1 \Big) \lesssim 1.$$

The last estimate follows since the leaves of $\mathcal{T}$ partition the root $R_\mathcal{T}$.

Now we consider the right-hand side of (2.10), which after convolving it with $\boldsymbol{F}$, integrating over T_k and summing in $k \in \mathbb{Z}$ results in

$$R := \sum_{k \in \mathbb{Z}} \int_{2^{k-1}}^{2^k} \int_{T_k} \boldsymbol{F}((F_j)_{j \in J}) * (-t\partial_t [\Psi]_t)(p,q,p,q)dpdq \frac{dt}{t},$$

where $J := \{1,2,3,4\}$. First we show that up to a controllable error, we may suppose that the functions F_j are supported on T_k. For $j \in J$ we write $F_j = F_j \mathbf{1}_{T_k} + F_j \mathbf{1}_{T_k^c}$. Then

$$R = M + E,$$

where the main term is defined as

$$M := \sum_{k \in \mathbb{Z}} \int_{2^{k-1}}^{2^k} \int_{T_k} \boldsymbol{F}((F_j \mathbf{1}_{T_k})_{j \in J}) * (-t\partial_t [\Psi]_t)(p,q,p,q)dpdq \frac{dt}{t}$$

and the error term is

$$E := \sum_{((X_{j,k})_{k \in \mathbb{Z}})_{j \in J}} \sum_{k \in \mathbb{Z}} \int_{2^{k-1}}^{2^k} \int_{T_k} \boldsymbol{F}((F_j \mathbf{1}_{X_{j,k}})_{j \in J}) * (-t\partial_t [\Psi]_t)(p,q,p,q)dpdq \frac{dt}{t},$$

where the outer summation is over $((X_{j,k})_{k \in \mathbb{Z}})_{j \in J} \in \{T, T^c\}^4 \setminus \{(T,T,T,T)\}$ for $T := (T_k)_{k \in \mathbb{Z}}$, $T^c := (T_k^c)_{k \in \mathbb{Z}}$.

To treat E we expand $-t\partial_t [\Psi]_t = -t\partial_t ([\rho_1]_t \otimes [\rho_2]_t \otimes [\rho_1]_t \otimes [\rho_2]_t)$ and use the chain rule, which results in four terms. By symmetry we consider only $-t\partial_t ([\rho_1]_t) \otimes [\rho_2]_t \otimes [\rho_1]_t \otimes [\rho_2]_t$, on which we use the identity

$$-t\partial_t [\rho_1]_t = -t\partial_t \Big(\frac{1}{t} \rho_1 \Big(\frac{x}{t} \Big) \Big) = \frac{1}{t} \rho_1 \Big(\frac{x}{t} \Big) + \frac{1}{t} \frac{x}{t} \rho_1' \Big(\frac{x}{t} \Big). \tag{2.11}$$

and bound the right-hand side of (2.11) by $\lesssim_c [\vartheta^2]_t$. This gives $|t\partial_t [\Psi]_t| \lesssim_c [\vartheta^2 \otimes \vartheta^2 \otimes \vartheta^2 \otimes \vartheta^2]_t$. By Lemma 6 we then have $|E| \lesssim_c 1$.

To estimate M we expand the convolution and interchange the order of integration such that the integration in (p,q) becomes the innermost. For now we consider only this innermost integral, which we write in the form

$$\int_{T_k} -t\partial_t \Big(\Big([\rho_1]_t(p-x)[\rho_1]_t(p-x') \Big) \Big([\rho_2]_t(q-y)[\rho_2]_t(q-y') \Big) \Big) dpdq.$$

Deriving the product of $[\rho_1]_t(p-x)[\rho_1]_t(p-x')$ and $[\rho_2]_t(q-y)[\rho_2]_t(q-y')$ yields two terms. Using Fubini and moving differentiation outside the integral we arrive at

$$\sum_{Q\in\mathcal{I}_k}\Big(-t\partial_t\int_{T_{Q,1}}[\rho_1]_t(p-x)[\rho_1]_t(p-x')dp\Big)\int_Q[\rho_2]_t(q-y)[\rho_2]_t(q-y')dq \tag{2.12}$$

$$+\sum_{P\in\mathcal{I}_k}\int_P[\rho_1]_t(p-x)[\rho_1]_t(p-x')dp\Big(-t\partial_t\int_{T_{P,2}}[\rho_2]_t(q-y)[\rho_2]_t(q-y')dq\Big), \tag{2.13}$$

where for a dyadic interval Q we denote $T_{Q,1}:=\cup_{P:P\times Q\in\mathcal{T}}P$ and $T_{P,2}$ is defined analogously. As both parts are treated in the same way, we further investigate only (2.12).

The identity (2.7) implies

$$-t\partial_t\int_{\mathbb{R}}[\rho_1]_t(p-x)[\rho_1]_t(p-x')dp=\int_{\mathbb{R}}[\sigma_1]_t(p-x)[\sigma_1]_t(p-x')dp,$$

which can be seen by an application of the inverse Fourier transform on (2.7). Hence,

$$-t\partial_t\int_{T_{Q,1}}[\rho_1]_t(p-x)[\rho_1]_t(p-x')dp=\int_{T_{Q,1}}[\sigma_1]_t(p-x)[\sigma_1]_t(p-x')dp+b_1,$$

where b_1 is the boundary portion

$$b_1:=\int_{\mathbb{R}\setminus T_{Q,1}}[\sigma_1]_t(p-x)[\sigma_1]_t(p-x')dp+t\partial_t\int_{\mathbb{R}\setminus T_{Q,1}}[\rho_1]_t(p-x)[\rho_1]_t(p-x')dp.$$

Therefore we have

$$M=\Big(\sum_{k\in\mathbb{Z}}\Theta^{\mathcal{T}_k}_{\sigma_1,\rho_2}((F_j\mathbf{1}_{T_k})_{j\in J})+\Theta^{\mathcal{T}_k}_{\rho_1,\sigma_2}((F_j\mathbf{1}_{T_k})_{j\in J})\Big)+B_1+B_2, \tag{2.14}$$

where the boundary term B_1 emerges from b_1 and equals

$$\begin{aligned}B_1:=&\sum_{k\in\mathbb{Z}}\int_{2^{k-1}}^{2^k}\sum_{Q\in\mathcal{I}_k}\int_Q\int_{\mathbb{R}\setminus T_{Q,1}}\boldsymbol{F}((F_j\mathbf{1}_{T_k})_{j\in J})(x,y,x',y')\\&\Big([\sigma_1]_t(p-x)[\sigma_1]_t(p-x'))+t\partial_t\big([\rho_1]_t(p-x)[\rho_1]_t(p-x')\big)\Big)[\rho_2]_t(q-x)[\rho_2]_t(q-x')\\&\qquad dxdydx'dy'dpdq\frac{dt}{t}.\end{aligned}$$

The boundary term B_2 arises from the treatment of (2.13) and is analogous to B_1 with (σ_1,ρ_2) replaced by (ρ_1,σ_2). For B_1,B_2 we derive by t using (2.11) and dominate the resulting functions by $\lesssim_c\vartheta^2$. Note that

$$|B_1+B_2|\lesssim_c\sum_{k\in\mathbb{Z}}\Theta^{\mathcal{T}_k^c}_{\vartheta^2,\vartheta^2}((F_j\mathbf{1}_{T_k})_{j\in J})\lesssim 1,$$

where the last inequality follows by Lemma 7.

Summarizing, since $L=R=M+E$, using (2.14) yields the identity

$$L=\Big(\sum_{k\in\mathbb{Z}}\Theta^{\mathcal{T}_k}_{\sigma_1,\rho_2}((F_j\mathbf{1}_{T_k})_{j\in J})+\Theta^{\mathcal{T}_k}_{\rho_1,\sigma_2}((F_j\mathbf{1}_{T_k})_{j\in J})\Big)+B_1+B_2+E. \tag{2.15}$$

Proposition 8 now follows by writing

$$\Theta^{\mathcal{T}}_{\rho_1,\sigma_2}((F_j)_{j\in J}) + \Theta^{\mathcal{T}}_{\sigma_1,\rho_2}((F_j)_{j\in J})$$

in the form

$$\sum_{k\in\mathbb{Z}} \Theta^{\mathcal{T}_k}_{\sigma_1,\rho_2}((F_j\mathbf{1}_{T_k})_{j\in J}) + \Theta^{\mathcal{T}_k}_{\rho_1,\sigma_2}((F_j\mathbf{1}_{T_k})_{j\in J})$$
$$+ \sum_{((X_{j,k})_{k\in\mathbb{Z}})_{j\in J}} \sum_{k\in\mathbb{Z}} \Theta^{\mathcal{T}_k}_{\rho_1,\sigma_2}((F_j\mathbf{1}_{X_{j,k}})_{j\in J}) + \Theta^{\mathcal{T}_k}_{\sigma_1,\rho_2}((F_j\mathbf{1}_{X_{j,k}})_{j\in J}),$$

where in the second line, the outer sum runs over $((X_{j,k})_{k\in\mathbb{Z}})_{j\in J} \in \{T, T^c\}^4 \backslash \{(T,T,T,T)\}$ for T as above. Using (2.15) together with

$$|L - B_1 - B_2 - E| \lesssim_c 1$$

and evoking Lemma 6 two more times finally yields (2.9). □

3 Tree estimate

In this section we derive an estimate for a quadrisublinear variant of $\Lambda^N_{\varphi,\psi}$ restricted to $\Omega_{\mathcal{T}}$ for a convex tree $\mathcal{T}$. This form is given by

$$\widetilde{\Theta}^{\mathcal{T}}_{\varphi,\psi}(F_1,F_2,F_3,F_4) := \int_{\Omega_{\mathcal{T}}} \left|\boldsymbol{F} * [\varphi^{(u)} \otimes \psi^{(v)} \otimes \varphi^{(-u)} \otimes \psi^{(-v)}]_t(p,q,p,q)\right| dpdq\frac{dt}{t}.$$

It can also be recognized as a quadrisublinear version of $\Theta^{\mathcal{T}}_{\varphi^{(u)},\psi^{(v)},\varphi^{(-u)},\psi^{(-v)}}$.

Proposition 9. *We have the estimate*

$$\widetilde{\Theta}^{\mathcal{T}}_{\varphi,\psi}(F_1,F_2,F_3,F_4) \lesssim |R_{\mathcal{T}}| \prod_{j=1}^{4} M(F_j,\mathcal{T}). \tag{3.1}$$

The proof of Proposition 9 proceeds in a very similar way as the proof of the L^4 bound (1.1). Besides replacing [3, Lemma 3] with Proposition 8, the only modification is the choice of a faster decaying superposition of the Gaussian exponential functions (3.2). For completeness we summarize all steps of the proof, interested readers are referred to [3].

Proof. By homogeneity and scale-invariance we may suppose $M(F_j,\mathcal{T}) = 1$ and $|R_{\mathcal{T}}| = 1$. First we expand the left-hand side of (3.1) and use the triangle inequality to arrive at

$$\int_{\Omega_{\mathcal{T}}} \int_{\mathbb{R}^2} \left| \int_{\mathbb{R}} F_1(x,y)F_2(x',y)[\psi^{(v)}]_t(q-y)dy \int_{\mathbb{R}} F_3(x',y')F_4(x,y')[\psi^{(-v)}]_t(q-y')dy' \right|$$
$$[|\varphi^{(u)}|]_t(p-x)[|\varphi^{(-u)}|]_t(p-x')dxdx'dpdq\frac{dt}{t}.$$

By an application of the Cauchy-Schwarz inequality, this is bounded by

$$\Theta^{\mathcal{T}}_{|\varphi^{(u)}|,\psi^{(v)},|\varphi^{(-u)}|,\psi^{(v)}}(F_1,F_2,F_2,F_1)^{1/2}\Theta^{\mathcal{T}}_{|\varphi^{(u)}|,\psi^{(-v)},|\varphi^{(-u)}|,\psi^{(-v)}}(F_4,F_3,F_3,F_4)^{1/2}.$$

As both terms are treated analogously, we consider the first one only. We shall now apply the telescoping identity, for which we dominate $\varphi^{(\pm u)}$ with a superposition of Gaussians. Denote the L^1-normalized Gaussian exponential function rescaled by $\alpha > 0$ by

$$g_\alpha(x) := \frac{1}{\sqrt{\pi}\alpha} e^{-\left(\frac{x}{\alpha}\right)^2}.$$

Consider the superposition of the functions g_α given by

$$\Phi(x) := \int_1^\infty \frac{1}{\alpha^{21}} e^{-\left(\frac{x}{\alpha}\right)^2} d\alpha = \frac{1}{\sqrt{\pi}} \int_1^\infty \frac{1}{\alpha^{20}} g_\alpha(x) d\alpha. \tag{3.2}$$

For large x we have $\Phi(x) \sim x^{20}$, which can be seen by the change of variables $\alpha' = (x/\alpha)^2$ and by inductive integration by parts. The power of α is now larger as in [3], as due to Proposition 8 we need control over higher order Schwartz seminorms of g_α.

Since $\varphi^{(\pm u)} \in \mathcal{S}(\mathbb{R}^2)$, we can bound it by Φ times a positive constant, which is uniform in u. By positivity of

$$\Theta^{\mathcal{T}}_{|\varphi^{(u)}|,\psi^{(v)},|\varphi^{(-u)}|,\psi^{(v)}}(F_1, F_2, F_2, F_1) = \int_{\Omega_{\mathcal{T}}} \int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}} F_1(x,y) F_2(x',y) [\psi^{(v)}]_t(q-y) dy \Big)^2$$
$$[|\varphi^{(u)}|]_t(p-x)\, [|\varphi^{(u)}|]_t(p-x') dx dx' dp dq \frac{dt}{t},$$

we can estimate this term up to a constant by

$$\int_1^\infty \int_1^\infty \Theta^{\mathcal{T}}_{g_\alpha,\psi^{(v)},g_\beta,\psi^{(v)}}(F_1, F_2, F_2, F_1) \frac{d\alpha}{\alpha^{20}} \frac{d\beta}{\beta^{20}}.$$

We split the integration into the regions $\alpha \geq \beta$ and $\alpha < \beta$. By symmetry it suffices to estimate the region $\alpha \geq \beta$ only, on which $\beta g_\beta \leq \alpha g_\alpha$ for $\alpha, \beta \geq 1$. This leaves us with

$$\int_1^\infty \Theta^{\mathcal{T}}_{g_\alpha,\psi^{(v)}}(F_1, F_2, F_2, F_1) \frac{d\alpha}{\alpha^{19}}.$$

Now we are ready to apply Proposition 8 with $(\rho_1, \sigma_1) = (g_\alpha, h_\alpha)$ and $(\rho_1, \sigma_2) = (\phi, \psi^{(v)})$, where $h_\alpha(x) := \alpha (g_\alpha)'(x)$ and

$$\widehat{\phi}(\xi) := \left(\int_\xi^\infty |\widehat{\psi^{(v)}}(\tau)|^2 \frac{d\tau}{\tau} \right)^{1/2},$$

which is a Schwartz function by our condition on ψ. Proposition 8 yields

$$\Theta^{\mathcal{T}}_{g_\alpha,\psi^{(v)}}(F_1, F_2, F_2, F_1) \lesssim -\Theta^{\mathcal{T}}_{h_\alpha,\phi}(F_1, F_2, F_2, F_1) + c \tag{3.3}$$

with $c = \|g_\alpha\|^2 \|\psi^{(v)}\|^2 + \|\phi\|^2 \|h_\alpha\|^2 + \|g_\alpha\|^2 \|\phi\|^2 \lesssim \alpha^{16}$. Thus it remains to estimate the form on the right-hand side of (3.3).

In the second iteration of the procedure we bound $|\Theta^{\mathcal{T}}_{h_\alpha,\phi}(F_1, F_2, F_2, F_1)|$ by

$$\int_{\Omega_{\mathcal{T}}} \int_{\mathbb{R}^2} \Big| \int_{\mathbb{R}} F_1(x,y) F_1(x,y') [h_\alpha]_t(p-x) dx \int_{\mathbb{R}} F_2(x',y') F_2(x',y) [h_\alpha]_t(p-x') dx' \Big|$$
$$[|\phi|]_t(q-y) [|\phi|]_t(q-y') dy dy' dp dq \frac{dt}{t}.$$

Again we apply the Cauchy-Schwarz inequality and arrive to

$$|\Theta^{\mathcal{T}}_{h_\alpha,\phi}(F_1,F_2,F_2,F_1)| \leq \Theta^{\mathcal{T}}_{h_\alpha,|\phi|}(F_1,F_1,F_1,F_1)^{1/2}\Theta^{\mathcal{T}}_{h_\alpha,|\phi|}(F_2,F_2,F_2,F_2)^{1/2}$$

Dominating the rapidly decaying $|\phi|$ by a positive constant times Φ gives

$$\Theta^{\mathcal{T}}_{h_\alpha,|\phi|}(F_1,F_1,F_1,F_1) \lesssim \int_1^\infty \int_1^\infty \Theta^{\mathcal{T}}_{h_\alpha,g_\gamma,h_\alpha,g_\delta}(F_1,F_1,F_1,F_1)\frac{d\gamma}{\gamma^{20}}\frac{d\delta}{\delta^{20}}.$$

As before, by symmetry this reduces to having to estimate

$$\int_1^\infty \Theta^{\mathcal{T}}_{h_\alpha,g_\gamma}(F_1,F_1,F_1,F_1)\frac{d\gamma}{\gamma^{19}}.$$

Now we apply Proposition 8 to the pairs $(\rho_1,\sigma_1)=(g_\alpha,h_\alpha)$ and $(\rho_2,\sigma_2)=(g_\gamma,h_\gamma)$, giving

$$\Theta^{\mathcal{T}}_{h_\alpha,g_\gamma}(F_1,F_1,F_1,F_1) \lesssim -\Theta^{\mathcal{T}}_{g_\alpha,h_\gamma}(F_1,F_1,F_1,F_1)+c$$

with $c=\|g_\alpha\|^2\|h_\gamma\|^2+\|g_\gamma\|^2\|h_\alpha\|^2+\|g_\alpha\|^2\|g_\gamma\|^2 \lesssim \alpha^{16}\gamma^{16}$. Finally observe that

$$\Theta^{\mathcal{T}}_{g_\alpha,h_\gamma}(F_1,F_1,F_1,F_1) \geq 0,$$

which can be seen by writing it as an integral of a square multiplied with $g_\alpha \geq 0$. Thus,

$$\Theta^{\mathcal{T}}_{h_\alpha,g_\gamma}(F_1,F_1,F_1,F_1) \leq 1.$$

This concludes the proof in view of our normalization. □

4 Completing the proof of Theorem 3

Now we are ready to establish the restricted type estimate from Theorem 3. We adapt the approach of [10] and we also rely on [9].

Proof of Theorem 3. First note that by quadrilinearity of $\Lambda^N_{\varphi,\psi}$ it suffices to prove the theorem for positive functions F_j, as otherwise we split them into real and imaginary, positive and negative parts.

For $j=1,2,3,4$ let α_j be such that $-1/2 \leq \alpha_j \leq 1/2$ and $\alpha_1+\alpha_2+\alpha_3+\alpha_4=1$. For each j let $E_j \subseteq \mathbb{R}^2$ be measurable. Without loss of generality we may assume $|E_1|$ is maximal among the $|E_j|$. Note that for $a=2^k$ we have the scaling identity

$$\Lambda^N_{\varphi,\psi}(F_1,F_2,F_3,F_4)=a^2\Lambda^{N/a}_{\varphi,\psi}(F_1(a\cdot),F_2(a\cdot),F_3(a\cdot),F_4(a\cdot)).$$

Since our bound will be independent of N, by $\sum_j \alpha_j=1$ we may then suppose $1 \leq |E_1| \leq 4$. All squares which we consider in this section are assumed to have their side-lengths in the interval $[2^{-N},2^N]$.

For F on $\mathbb{R}^2$ we denote the quadratic Hardy-Littlewood maximal function by

$$\mathcal{M}(F) := \sup_S \Big(\frac{1}{|S|}\int_S F^2\Big)^{1/2}\mathbf{1}_S,$$

where the supremum is taken over all (not necessarily dyadic) squares in $\mathbb{R}^2$ with sides parallel to the coordinate axes. From now on, by the word "average" we will always mean the second power average as in the definition of $\mathcal{M}(F)$. Define the exceptional set

$$H := \bigcup_{j=1}^{4} \{\mathcal{M}(|E_j|^{-1/2}\mathbf{1}_{E_j}) > 2^{10}\}.$$

By the Hardy-Littlewood maximal theorem we have $|H| \leq 1/18$. Let $\mathcal{R}$ be the set of all dyadic squares $R \subseteq H$ which are maximal with respect to set inclusion. Denote by $3R$ the square with the same center as R but with three times the sidelength of R. We set $E_1' := E_1 \setminus \cup_{R\in\mathcal{R}} 3R$. Then $2|E_1'| \geq |E_1|$.

Suppose we are given four functions F_j with $|F_j| \leq \mathbf{1}_{E_j}$ for all j and $|F_1| \leq \mathbf{1}_{E_1'}$. Since $\alpha_j \leq 1/2$ and $|E_1| \leq 4$, it suffices to prove

$$|\Lambda^N_{\varphi,\psi}(F_1, F_2, F_3, F_4)| \lesssim |E_1|^{1/2}|E_2|^{1/2}|E_3|^{1/2}|E_4|^{1/2}.$$

If we set $G_j := |E_j|^{-1/2}F_j$, then the inequality we need to establish reads

$$|\Lambda^N_{\varphi,\psi}(G_1, G_2, G_3, G_4)| \lesssim 1.$$

Observe that $\|G_j\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq 1$ for all j.

We split $\mathbb{R}^2 \times [2^{-N}, 2^N]$ into the regions $\Omega_{\{S\}} = S \times [\ell(S)/2, \ell(S)]$, $S \in \mathcal{D}$, and consider the cases $S \subseteq H$ and $S \not\subseteq H$. By the triangle inequality we estimate

$$|\Lambda^N_{\varphi,\psi}| \leq \sum_{S\subseteq H} \widetilde{\Theta}^{\{S\}}_{\varphi,\psi} + \sum_{S\not\subseteq H} \widetilde{\Theta}^{\{S\}}_{\varphi,\psi}.$$

First we consider the sum over $S \not\subseteq H$. For $k \in \mathbb{Z}$ let $\mathcal{S}_k$ be the set of all dyadic squares S for which

$$2^{k-1} < \max_{j\in\{1,2,3,4\}} \sup_{S'\supseteq S} \Big(\frac{1}{|S'|}\int_{S'} G_j^2\Big)^{1/2} \leq 2^k.$$

The supremum is taken over all (not necessarily dyadic) squares $S' \supseteq S$ in $\mathbb{R}^2$ with sides parallel to the coordinate axes. Denote by $\mathcal{R}_k$ the collection of the maximal squares in $\mathcal{S}_k$ with respect to set inclusion. For $R \in \mathcal{R}_k$ we define

$$\mathcal{T}_R := \{S \in \mathcal{S}_k : S \subseteq R\},$$

which is a convex tree with the root R. Convexity follows from monotonicity of the supremum. By construction, if $S \not\subseteq H$, for each j the average of $|E_j|^{-1/2}\mathbf{1}_{E_j}$ over S is no greater than 2^{10}. Thus, the same holds for the average of G_j over S. Therefore,

$$\{S : S \not\subseteq H\} \subseteq \bigcup_{k\leq 10} \mathcal{S}_k$$

and we can split the summation as

$$\sum_{S\not\subseteq H} \widetilde{\Theta}^{\{S\}}_{\varphi,\psi} \leq \sum_{k\leq 10}\sum_{R\in\mathcal{R}_k}\sum_{S\in\mathcal{T}_R} \widetilde{\Theta}^{\{S\}}_{\varphi,\psi} = \sum_{k\leq 10}\sum_{R\in\mathcal{R}_k} \widetilde{\Theta}^{\mathcal{T}_R}_{\varphi,\psi}.$$

For the forms on the right-hand side we have by Proposition 9 that

$$\widetilde{\Theta}_{\varphi,\psi}^{\mathcal{T}_R}(G_1, G_2, G_3, G_4) \lesssim |R| \prod_{j=1}^{4} M(G_j, \mathcal{T}_R). \tag{4.1}$$

To estimate the right-hand side of (4.1) we discretize the function θ by a standard approximation with characteristic functions of balls of radius at least 1. We now sketch the required argument. Denote by B_r the ball of radius r centered at 0 in $\mathbb{R}^2$. We write

$$G_j^2 * [\theta]_t = G_j^2 * [\theta \mathbf{1}_{B_1}]_t + G_j^2 * [\theta \mathbf{1}_{B_1^c}]_t.$$

Let $(p, q, t) \in S \times [\ell(S)/2, \ell(S)] \subseteq \Omega_{\mathcal{T}_R}$ and assume $(p, q) = 0$. On B_1 we have

$$G_j^2 * [\theta \mathbf{1}_{B_1}]_t(0) \lesssim \|\theta\|_{\mathrm{L}^\infty(\mathbb{R}^2)} \frac{1}{(2t)^2} \int_{[-t,t]^2} G_j^2 \lesssim \frac{1}{(2\ell(S))^2} \int_{[-\ell(S),\ell(S)]^2} G_j^2 \lesssim 2^{2k}. \tag{4.2}$$

For the part on B_1^c we consider the function $\theta \mathbf{1}_{B_1^c} + \frac{1}{2}\mathbf{1}_{B_1}$. It dominates $\theta \mathbf{1}_{B_1^c}$, is positive and radially decreasing. Therefore it can be approximated pointwise by a monotonously increasing sequence of simple functions of the form

$$E = \sum_{i=1}^{n} a_i \mathbf{1}_{B_{r_i}}, \ \ r_i \geq 1, \ a_i > 0.$$

For E we have, using $t \sim \ell(S)$, that

$$G_j^2 * [E]_t(0) \lesssim \sum_{i=1}^{n} a_i |B_{r_i}| \frac{1}{(r_i \ell(S))^2} \int_{[-r_i\ell(S), r_i\ell(S)]^2} G_j^2 \lesssim \|\theta\|_{\mathrm{L}^1(\mathbb{R}^2)} 2^{2k}.$$

This implies the estimate

$$G_j^2 * [\theta \mathbf{1}_{B_1^c}]_t(0) \lesssim 2^{2k}. \tag{4.3}$$

By a translation argument, the same bound holds at any $(p, q, t) \in \Omega_{\mathcal{T}_R}$. Therefore, by (4.2) and (4.3), we have $M(G_j, \mathcal{T}_R) \lesssim 2^k$ for each j and hence

$$\sum_{S \not\subseteq H} \widetilde{\Theta}_{\varphi,\psi}^{\{S\}}(G_1, G_2, G_3, G_4) \lesssim \sum_{k \leq 10} 2^{4k} \sum_{R \in \mathcal{R}_k} |R|. \tag{4.4}$$

It remains to sum up the right-hand side of the last display. Since for $R \in \mathcal{R}_k$ there is an index j such that on R we have $\mathcal{M}(G_j) > 2^{k-1}$, by maximality of the squares in $\mathcal{R}_k$

$$\sum_{R \in \mathcal{R}_k} |R| = \Big| \bigcup_{R \in \mathcal{R}_k} R \Big| \leq \sum_{j=1}^{4} |\{\mathcal{M}(G_j) > 2^{k-1}\}|.$$

By the Hardy-Littlewood maximal theorem and $\|G_j\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq 1$, for each j we have $|\{\mathcal{M}(G_j) > 2^{k-1}\}| \lesssim 2^{-2k}$. Thus, (4.4) is up to a constant dominated by

$$\sum_{k \leq 10} 2^{2k} \lesssim 1.$$

This establishes the desired estimate for $S \not\subseteq H$.

Now consider the sum over all dyadic squares S contained in H. Every $S \subseteq H$ is contained in one maximal dyadic square $R \in \mathcal{R}$. Let $\mathcal{S}_{R,k}$ be the set of dyadic squares S which are k generations below $R \in \mathcal{R}$. That is, $2^k \ell(S) = \ell(R)$. We split

$$\sum_{S \subseteq H} \widetilde{\Theta}_{\varphi,\psi}^{\{S\}} = \sum_{R \in \mathcal{R}} \sum_{k \geq 0} \sum_{S \in \mathcal{S}_{R,k}} \widetilde{\Theta}_{\varphi,\psi}^{\{S\}}.$$

For $S \in \mathcal{S}_{R,k}$ we expand $\widetilde{\Theta}_{\varphi,\psi}^{\{S\}}(G_1, G_2, G_3, G_4)$ and estimate $|\varphi^{(u)}|, |\psi^{(v)}| \lesssim \vartheta^4$ to arrive at

$$\int_{\ell(S)/2}^{\ell(S)} \int_S \int_{\mathbb{R}^4} \boldsymbol{F}(G_1, G_2, G_3, G_4)(x, y, x', y')[\vartheta \otimes \vartheta \otimes \vartheta \otimes \vartheta]_t(p - x, q - y, p - x', q - y')$$
$$\theta^2(t^{-1}(p - x, q - y))\, dx dy dx' dy' dp dq \frac{dt}{t}. \tag{4.5}$$

Since G_1 is supported on the complement of $3R$, we have $|(p, q) - (x, y)| \geq \ell(R)$ for $(p, q) \in S$. We also have $\ell(R) = 2^k \ell(S) \sim 2^k t$, therefore $\theta^2(t^{-1}(p - x, q - y)) \lesssim 2^{-8k}$. Applying Lemma 5, the term (4.5) is then up to a constant dominated by

$$2^{-8k} |S| \prod_{j=1}^{4} M(G_j, \{S\}).$$

Denote by R' the parent of R. For each j we have

$$M(G_j, \{S\}) \lesssim 2^k M(G_j, \{R'\}) \lesssim 2^k.$$

The last inequality follows by the same approximation argument as before and using that the averages of G_j over squares containing R' are less than 2^{10}, which is true by maximality of R. This establishes

$$\sum_{S \subseteq H} \widetilde{\Theta}_{\varphi,\psi}^{\{S\}}(G_1, G_2, G_3, G_4) \lesssim \sum_{R \in \mathcal{R}} \sum_{k \geq 0} \sum_{S \in \mathcal{S}_{R,k}} 2^{-4k} |S|.$$

Since $\sum_{S \in \mathcal{S}_{R,k}} |S| \leq |R|$, the last display is estimated by

$$\sum_{R \in \mathcal{R}} |R| \sum_{k \geq 0} 2^{-4k} \lesssim |H| \lesssim 1.$$

For the second to last inequality we summed the geometric series and used disjointness of $R \in \mathcal{R}$. In the last step we used $|H| \leq 1/2$. □

References

[1] F. Bernicot, *Fiber-wise Calderón-Zygmund decoposition and application to a bi-dimensional paraproduct*, Illinois J. Math. **56** (2012), no. 2, 415–422.

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

[3] P. Durcik, *An L^4 estimate for a singular entangled quadrilinear form*, to appear in Math. Res. Lett., available at arXiv:1412.2384.

[4] 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.

[5] V. Kovač, *Bellman function technique for multilinear estimates and an application to generalized paraproducts*, Indiana Univ. Math. J., **60** (2011), no. 3, 813–846.

[6] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam., **28** (2012), no. 4, 1143–1164.

[7] C. Muscalu, T. Tao, C. Thiele, *Uniform estimates on multi-linear operators with modulation symmetry*, J. Anal. Math. **88** (2002), 255–309.

[8] E. M. Stein, *Harmonic Analysis: Real Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, Princeton, 1993.

[9] C. Thiele, *Time-frequency analysis in the discrete phase plane*, Ph.D. thesis, Yale University, 1995, Topics in analysis and its applications, 99152, World Sci. Publ., River Edge, NJ, 2000.

[10] C. Thiele, *Wave packet analysis*, CBMS Reg. Conf. Ser. Math., **105**, AMS, Providence, RI, 2006.

Polona Durcik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany

E-mail address: `durcik@math.uni-bonn.de`

Chapter 4

Norm-variation of ergodic averages with respect to two commuting transformations

Norm-variation of ergodic averages with respect to two commuting transformations

Polona Durcik, Vjekoslav Kovač, Kristina Ana Škreb, and Christoph Thiele

Abstract

We study double ergodic averages with respect to two general commuting transformations and establish a sharp quantitative result on their convergence in the norm. We approach the problem via real harmonic analysis, using recently developed methods for bounding multilinear singular integrals with certain entangled structure. A byproduct of our proof is a bound for a two-dimensional bilinear square function related to the so-called triangular Hilbert transform.

1 Introduction

Many problems in ergodic theory are related to the convergence of certain averages along the orbits with respect to one or several transformations. Let $(X, \mathcal{F}, \mu)$ be a σ-finite measure space and let $S\colon X \to X$ be a measure-preserving transformation, i.e. for any $E \in \mathcal{F}$ we have $S^{-1}E \in \mathcal{F}$ and $\mu(S^{-1}E) = \mu(E)$. The most classical result in this direction is von Neumann's *mean ergodic theorem* [38], which guarantees convergence of the *single ergodic averages*

$$M_n f(x) := \frac{1}{n} \sum_{i=0}^{n-1} f(S^i x) \tag{1.1}$$

in the $\mathrm{L}^2(X)$ norm for any $f \in \mathrm{L}^2(X)$. Classical proofs of this fact do not provide any information on the rate of this convergence. With the aid of the spectral theorem, Jones, Ostrovskii, and Rosenblatt [20] have observed the quantitative variant of this result in the form of the norm-variation estimate

$$\sum_{j=1}^{m} \|M_{n_j} f - M_{n_{j-1}} f\|_{\mathrm{L}^2(X)}^2 \le C\, \|f\|_{\mathrm{L}^2(X)}^2 \tag{1.2}$$

for any positive integers $n_0 < n_1 < \cdots < n_m$ and with an absolute finite constant C. The work of Bourgain [9] prequels (1.2) and his pointwise variation estimates imply the same inequality albeit with the power 2 replaced by an arbitrary $\varrho > 2$. Calderón's transference principle, a version of which we discuss in Section 6, reduces (1.2) to studying operators in harmonic analysis that are well-understood by now.

Multiple ergodic averages were motivated by the work of Furstenberg and others [16], [17], [18] connecting ergodic theory with arithmetic combinatorics. In this paper we

2010 *Mathematics Subject Classification.* Primary 37A30; Secondary 42B15, 42B20.

are concerned with the bilinear case. Let $S, T\colon X \to X$ be two measure-μ-preserving transformations such that $ST = TS$. For any two complex-valued measurable functions f, g on X and any positive integer n one can define the *double ergodic average* $M_n(f,g)$ as a function on X given by

$$M_n(f,g)(x) := \frac{1}{n} \sum_{i=0}^{n-1} f(S^i x) g(T^i x) \tag{1.3}$$

for each $x \in X$. It is a classical result by Conze and Lesigne [10] that for any two functions $f, g \in \mathrm{L}^\infty(X)$ on a probability space the sequence of averages $(M_n(f,g))_{n=1}^\infty$ converges in the L^2 norm. Standard density arguments combined with log-convexity of L^p norms extend this result to functions $f \in \mathrm{L}^{p_1}(X)$, $g \in \mathrm{L}^{p_2}(X)$, with convergence in the L^p norm, as long as the exponents satisfy $p < \infty$ and $1/p \geq 1/p_1 + 1/p_2$. However, no explicitly quantitative variant of this fact for completely general commuting transformations S, T exists in the literature and this is the topic of the present paper.

Our main result is the following estimate for the averages (1.3).

Theorem 1. *There is a finite constant C such that for any σ-finite measure space $(X, \mathcal{F}, \mu)$, any two commuting measure-preserving transformations S, T on that space, and all functions $f, g \in \mathrm{L}^4(X)$ we have*

$$\sum_{j=1}^{m} \|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|_{\mathrm{L}^2(X)}^2 \leq C \, \|f\|_{\mathrm{L}^4(X)}^2 \|g\|_{\mathrm{L}^4(X)}^2 \tag{1.4}$$

for each choice of positive integers m and $n_0 < n_1 < \cdots < n_m$.

Such quantitative estimate for multiple ergodic averages was stated as an open problem by Avigad and Rute in the closing section of [3], after the question had already circulated in the community for a while. A result analogous to Theorem 1 was previously established by the second author in [24], but only for a simplified model, where the actions of $\mathbb{Z}$ are replaced by actions of infinite powers $\mathbb{A}^\omega$ of a fixed finite abelian group $\mathbb{A}$, and which avoided challenges we address in this paper.

Unlike for (1.2), Calderón's transference of (1.4) leads to a non-classical problem in harmonic analysis, whose solution is the main point of our paper. We do not know of a martingale approach to (1.4), even for particular cases of indices n_j. This is in contrast with the powerful martingale techniques for handling the single ergodic averages (1.1); compare with [3], [9], [21].

The techniques of this paper do not immediately generalize to the multiple variants of (1.3), i.e. to the analogous ergodic averages with respect to several commuting transformations. However, such averages are also known to converge in the norm, as was first shown by Tao [36], with a different proof given by Austin [2]. More generally, norm convergence of multiple averages was established by Walsh [39] in the case when the transformations generate a nilpotent group.

Almost everywhere convergence of the averages (1.3) is a longstanding open problem. In the single average case (1.1), almost everywhere convergence is Birkhoff's classical *pointwise ergodic theorem* [6], with quantitative estimates discussed in Bourgain [9] and

Jones, Kaufman, Rosenblatt, and Wierdl [19]. For two transformations S, T the task simplifies if T is assumed to be a power of S, for instance S is invertible and $T = S^{-1}$, and was successfully studied by the analytic approach and an almost everywhere convergence result was established by Bourgain [8]. Subsequently, a pointwise variation estimate was established by Do, Oberlin, and Palsson [12]. The result from [12] also implies a variant of our Theorem 1 with exponent $\varrho > 2$ in the special case $T = S^{-1}$. For further partial progress on a.e. convergence for general commuting transformations we refer to the preprint by Donoso and Sun [13] and references therein. In [13] the a.e. convergence is verified under the additional assumption that $(X, \mathcal{F}, \mu, S, T)$ forms a so-called distal system, i.e. a certain iterated topological extension of the trivial system.

Recall that the number of *ε-jumps* or *ε-fluctuations* of a sequence $(a_n)_{n=1}^{\infty}$ in a Banach space B, in our case $\mathrm{L}^2(X)$, is defined as the supremum of the set of integers J for which there exist indices

$$m_1 < n_1 \leq m_2 < n_2 \leq \cdots \leq m_J < n_J$$

such that $\|a_{n_j} - a_{m_j}\|_B \geq \varepsilon$ for $j = 1, 2, \ldots, J$. A direct consequence of our main theorem is that for all functions f, g of norm one in $\mathrm{L}^4(X)$ the number of ε-jumps of the averages (1.3) is at most $C\varepsilon^{-2}$. In particular, the number of ε-jumps is finite for each $\varepsilon > 0$, which implies norm convergence, i.e. it reproves the result by Conze and Lesigne [10]. It follows further that for any $\varepsilon > 0$ the sequence $(M_n(f,g))_{n=1}^{\infty}$ can be covered by at most $C\varepsilon^{-2} + 1$ balls of radius ε in the Hilbert space $\mathrm{L}^2(X)$. Such a result is sometimes called a uniform bound for the metric entropy. It was shown by Bourgain [7] that a.e. convergence of certain sequences of functions, including the single ergodic averages (1.1), necessarily implies the uniform bound on their metric entropy. In that light Theorem 1 can also be thought of as a partial progress towards the conjecture on a.e. convergence of (1.3), even though the bilinear analogue of [7] does not appear in the literature.

Our main inequality may be reformulated as

$$\|M_n(f,g)\|_{\mathrm{V}_n^{\varrho}(\mathbb{N},\mathrm{L}^p(X))} \leq C^{1/2}\, \|f\|_{\mathrm{L}^{p_1}(X)} \|g\|_{\mathrm{L}^{p_2}(X)},$$

with $\varrho = p = 2$ and $p_1 = p_2 = 4$, where for $1 \leq \varrho < \infty$ the *ϱ-variation* of a Banach-space-valued function $a\colon \mathcal{U} \to B$ with $\mathcal{U} \subseteq \mathbb{R}$ is defined as

$$\|a\|_{\mathrm{V}^{\varrho}(\mathcal{U},B)} := \|a(t)\|_{\mathrm{V}_t^{\varrho}(\mathcal{U},B)} := \sup_{\substack{m\in\mathbb{N}\cup\{0\}\\ t_0,t_1,\ldots,t_m\in\mathcal{U}\\ t_0<t_1<\cdots<t_m}} \Big(\sum_{j=1}^{m} \|a(t_j) - a(t_{j-1})\|_B^{\varrho}\Big)^{1/\varrho}.$$

If $(X, \mathcal{F}, \mu)$ is a probability space, then for any $f, g \in \mathrm{L}^{\infty}(X)$, $1 \leq p < \infty$, and $\varrho \geq \max\{p, 2\}$ we have

$$\|M_n(f,g)\|_{\mathrm{V}_n^{\varrho}(\mathbb{N},\mathrm{L}^p(X))} \leq C_{p,\varrho}\, \|f\|_{\mathrm{L}^{\infty}(X)} \|g\|_{\mathrm{L}^{\infty}(X)}$$

for some finite constant $C_{p,\varrho}$ depending only on p and ϱ. In order to see this, by the monotonicity of L^p norms on a probability space in the case $p < 2$ we can use

$$\|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|_{\mathrm{L}^p(X)} \leq \|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|_{\mathrm{L}^2(X)}$$

and by their log-convexity for $p > 2$ we have

$$\begin{aligned}&\|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|_{\mathrm{L}^p(X)} \\ &\leq (2\|f\|_{\mathrm{L}^\infty(X)}\|g\|_{\mathrm{L}^\infty(X)})^{1-2/p}\|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|_{\mathrm{L}^2(X)}^{2/p}.\end{aligned}$$

We then apply (1.4) and for that purpose in the latter case we need $2\varrho/p \geq 2$.

The variation exponent 2 is best possible in Theorem 1. To see this, it suffices to consider the special case $|f| = |g|$ and $S = T$ and notice that this special case is tantamount to estimate (1.2), where the exponent 2 is well known to be sharp. The range of exponents p_1, p_2, p, ϱ in the above discussion is likely not exhausted as the analogous work [24] in the simplified setting suggests.

This paper, while self-contained, builds on a technique for bounding multi-linear and multi-scale singular integral operators gradually developed by the authors in [14], [15], [22], [23], [24], [25], [26]. We consider the present application to quantitative norm convergence for double ergodic averages a milestone in these efforts. A notable difference from the almost everywhere result by Do, Oberlin, and Palsson [12] is that we do not use wave packet analysis or time-frequency analysis, as these tools are not well-adapted to our problem.

The technique we use resembles energy methods in partial differential equations. The main ingredients are integration by parts, positivity arguments, and the Cauchy-Schwarz inequality. The idea is to set up a partial integration scheme to produce positive terms, similar to energies, and then use upper bounds on a sum of positive terms to control each term individually. Unlike for most energy arguments in partial differential equations, here the partial integration happens in the scale parameter, which is typical for the singular integral theory. The structural complexity of the problem requires to iterate these steps, with the Cauchy-Schwarz inequality used inbetween to reduce the complexity of the expressions.

We elaborate more on the harmonic analysis part of the paper. For a one-dimensional integrable function φ and two-dimensional functions $F, G \in \mathrm{L}^4(\mathbb{R}^2)$, for $t > 0$, and for $(x, y) \in \mathbb{R}^2$ we introduce the bilinear averages

$$A_t^\varphi(F,G)(x,y) := \int_{\mathbb{R}} F(x+s,y)G(x,y+s)\, t^{-1}\varphi(t^{-1}s)\, ds.$$

Theorem 1 will be a consequence of the following bilinear estimate where $\varphi = \mathbb{1}_{[0,1)}$ is the characteristic function of the interval $[0, 1)$.

Theorem 2. *There exists a finite constant C such that for any $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ we have*

$$\big\|A_t^{\mathbb{1}_{[0,1)}}(F,G)\big\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \leq C\, \|F\|_{\mathrm{L}^4(\mathbb{R}^2)}\|G\|_{\mathrm{L}^4(\mathbb{R}^2)}.$$

By invariance of the left hand side under rescaling in t and by superposition, the theorem implies an inequality independent of the choice of positive numbers $t_0 < \cdots < t_m$:

$$\sum_{j=1}^m \|A_{t_j}^\varphi(F,G) - A_{t_{j-1}}^\varphi(F,G)\|_{\mathrm{L}^2(\mathbb{R}^2)}^2 \leq C_\varphi^2\, \|F\|_{\mathrm{L}^4(\mathbb{R}^2)}^2\|G\|_{\mathrm{L}^4(\mathbb{R}^2)}^2, \tag{1.5}$$

where

$$\varphi(s) = \int_{(-\infty,0)} \mathbb{1}_{[\alpha,0)}(s) \frac{d\mu(\alpha)}{-\alpha} + \int_{(0,\infty)} \mathbb{1}_{[0,\alpha)}(s) \frac{d\mu(\alpha)}{\alpha}$$

for some finite complex Radon measure μ on $(-\infty, 0) \cup (0, \infty)$. In particular, we get (1.5) for compactly supported functions φ of bounded variation and the constant C_φ is then a universal multiple of the total mass of the measure μ. Moreover, by choosing $d\mu(\alpha) = -\alpha\varphi'(\alpha)d\alpha$ we can recover an arbitrary Schwartz function φ and in that case the constant C_φ in (1.5) is a multiple of $\int_{\mathbb{R}} |s\varphi'(s)|ds$.

In the proof of Theorem 2 we gradually consider various classes of functions φ and carefully control C_φ for these classes. Indeed, we begin by showing that (1.5) holds for an arbitrary Schwartz function. However, we will actually need to apply the theorem with $\varphi = \mathbb{1}_{[0,1)}$, and this case is more subtle and requires more precise decay conditions in the auxiliary estimates. Prior to our paper, inequality (1.5) was not known even for a single nonzero function φ.

While deriving Theorem 1 from Theorem 2, the following discrete estimate will appear along the way. It is worth stating as a separate corollary due to its elegant formulation. For any two double sequences $\widetilde{F}, \widetilde{G}\colon \mathbb{Z}^2 \to \mathbb{R}$, for $n \in \mathbb{N}$, and for $(k,l) \in \mathbb{Z}^2$ we define the discrete averages $\widetilde{A}_n$ by

$$\widetilde{A}_n(\widetilde{F}, \widetilde{G})(k,l) := \frac{1}{n} \sum_{i=0}^{n-1} \widetilde{F}(k+i,l)\, \widetilde{G}(k,l+i). \tag{1.6}$$

Corollary 3. *There exists a finite constant C such that for any $\widetilde{F}, \widetilde{G} \in \ell^4(\mathbb{Z}^2)$ we have*

$$\big\|\widetilde{A}_n(\widetilde{F}, \widetilde{G})\big\|_{\mathrm{V}^2_n(\mathbb{N},\ell^2(\mathbb{Z}^2))} \leq C\, \|\widetilde{F}\|_{\ell^4(\mathbb{Z}^2)} \|\widetilde{G}\|_{\ell^4(\mathbb{Z}^2)}.$$

Inequality (1.5), even for Schwartz functions φ, is already new in the special case $t_j = 2^j$. In this case we set $\psi(s) := \varphi(s) - 2\varphi(2s)$ and define the square function

$$S(F,G)(x,y) := \Big(\sum_{j\in\mathbb{Z}} \Big| \int_{\mathbb{R}} F(x+s,y)\, G(x,y+s)\, 2^{-j}\psi(2^{-j}s)\, ds \Big|^2 \Big)^{1/2}.$$

A simple limiting argument as $m \to \infty$ in (1.5) yields the following corollary.

Corollary 4. *For any $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ we have*

$$\|S(F,G)\|_{\mathrm{L}^2(\mathbb{R}^2)} \leq C_\psi \|F\|_{\mathrm{L}^4(\mathbb{R}^2)} \|G\|_{\mathrm{L}^4(\mathbb{R}^2)},$$

with a finite constant C_ψ depending on ψ alone.

Indeed, square function estimates of this type are a stepping stone towards the proof of Theorem 2; for example compare with Proposition 7 stated in Section 2.

In contrast with Corollary 4, no bounds are known for the corresponding bilinear singular integral

$$T(F,G)(x,y) := \text{p.v.} \int_{\mathbb{R}} F(x+s,y)\, G(x,y+s) \frac{ds}{s},$$

which was introduced in [11] and later named the *triangular Hilbert transform*. Only partial results in this direction exist; see [27] for a particular case when one of the functions takes a special form. Moreover, Zorin-Kranich showed in [41], building on the approach of Tao [35], that the truncations to m consecutive scales,

$$T_m(F,G)(x,y) := \sum_{j=1}^{m} \int_{\mathbb{R}} F(x+s,y)\, G(x,y+s)\, 2^{-j}\psi(2^{-j}s)\, ds,$$

have norms from $\mathrm{L}^{p_1}(\mathbb{R}^2) \times \mathrm{L}^{p_2}(\mathbb{R}^2)$ to $\mathrm{L}^p(\mathbb{R}^2)$ that grow like $o(m)$ as $m \to \infty$, for any fixed choice of exponents $1 < p, p_1, p_2 < \infty$ such that $1/p = 1/p_1 + 1/p_2$. Using Corollary 4 and the Cauchy-Schwarz inequality we improve this growth to $O(m^{1/2})$ for $p = 2$, $p_1 = p_2 = 4$, and then the interpolation with the trivial estimates coming from Hölder's inequality gives the growth $O(m^{1-\epsilon})$ for general exponents p, p_1, p_2 as before and for some $\epsilon > 0$ depending on them.

Furthermore, for given $f, g \in \mathrm{L}^4(\mathbb{R})$ let us take

$$F(x,y) := f(x-y)R^{-1/4}\vartheta(R^{-1}y), \quad G(x,y) := g(x-y)R^{-1/4}\vartheta(R^{-1}x),$$

where $R > 0$ and ϑ is a smooth compactly supported nonnegative function on $\mathbb{R}$ that is constantly 1 on the interval $[-1,1]$. By substituting $z = x - y$, observing

$$\begin{aligned} &\int_{-R}^{R}\int_{-R}^{R} \Big| \int_{\mathbb{R}} F(x+s,y)\, G(x,y+s)\, 2^{-j}\psi(2^{-j}s)\, ds \Big|^2 dxdy \\ &\geq \int_{-R}^{R} \Big| \int_{\mathbb{R}} f(z+s)\, g(z-s)\, 2^{-j}\psi(2^{-j}s)\, ds \Big|^2 dz, \end{aligned}$$

applying Corollary 4, and letting $R \to \infty$ we recover the $\mathrm{L}^4(\mathbb{R}) \times \mathrm{L}^4(\mathbb{R}) \to \mathrm{L}^2(\mathbb{R})$ estimate for the one-dimensional bilinear square function

$$\widetilde{S}(f,g)(x) := \Big(\sum_{j\in\mathbb{Z}} \Big| \int_{\mathbb{R}} f(x+s)\, g(x-s)\, 2^{-j}\psi(2^{-j}s)\, ds \Big|^2 \Big)^{1/2}.$$

The only previously known proof of an L^p bound for $\widetilde{S}$ employs wave-packet analysis, i.e. it uses Khintchine's inequality to reduce to an average of a family of bilinear singular integrals parametrized by random signs and then recognizes these operators in the proof of boundedness of the bilinear Hilbert transform [29], [30].

Somewhat related, there is an open problem stated in the introductory section of the paper by Bernicot [4] to show L^p bounds for the bilinear square function

$$S_\Omega(f,g)(x) := \Big(\sum_{\omega\in\Omega} \Big| \int_{\mathbb{R}} f(x+s)\, g(x-s)\, \check{\mathbb{1}}_\omega(s)\, ds \Big|^2 \Big)^{1/2}$$

for an arbitrary collection of disjoint intervals Ω, which would be a bilinear variant of the well-known result by Rubio de Francia [34]. Here $\check{\mathbb{1}}_\omega$ denotes the inverse Fourier transform of $\mathbb{1}_\omega$. Bernicot [4] has verified this conjecture for a particular case of equidistant intervals of the same length, such as $\Omega = \{[j, j+1) : j \in \mathbb{Z}\}$. The problem becomes simpler if we replace $\mathbb{1}_\omega$ with a smooth bump function adapted to ω, as was already observed by Lacey [28] in the case of the intervals $[j, j+1)$, see also [5], [32], [33]. The above bilinear square function $\widetilde{S}$ is associated with smooth truncations of the lacunary intervals $\Omega = \{[2^j, 2^{j+1}) : j \in \mathbb{Z}\}$.

This paper is organized as follows: In Section 2 we begin the proof of Theorem 2 by splitting the jumps into the "long ones" (i.e. those corresponding to the scales t_j that are dyadic numbers 2^k, $k \in \mathbb{Z}$) discussed in Lemma 8 and the "short ones" (i.e. those corresponding to t_j from a fixed interval $[2^k, 2^{k+1}]$) discussed in Lemmata 9 and 10. Propositions 5–7 are the key results here. Their proofs are postponed to Sections 3–5 and these three sections contain the main novelties of our approach. Finally, the somewhat standard transition from Theorem 2 to Corollary 3 and then to Theorem 1 is presented in details in Section 6.

2 Averages on $\mathbb{R}^2$, long and short variations

In this section we split Theorem 2 into long and short variation estimates and show how to deduce these from Propositions 5, 6, and 7 below.

For two non-negative quantities A and B we write $A \lesssim B$ if there exists a constant $C > 0$ such that $A \leq CB$. When we want to emphasize dependence of the constant on some parameters $p, q, \ldots$, we denote them in the subscript, i.e. we write $\lesssim_{p,q,\ldots}$. Occasionally we may omit writing down parameters that are understood. We write $A \sim B$ if both $A \lesssim B$ and $B \lesssim A$ are satisfied.

For a function φ on $\mathbb{R}^d$ and $t > 0$ we set $\varphi_t(x) := t^{-d}\varphi(t^{-1}x)$. Consequently, $A_t^\varphi = A_1^{\varphi_t}$. By $\mathcal{S}(\mathbb{R}^d)$ we denote the class of all Schwartz functions on $\mathbb{R}^d$, while the word "smooth" will always mean C^∞. The Fourier transform of an integrable function φ on $\mathbb{R}^d$ is defined as

$$\widehat{\varphi}(\xi) := \int_{\mathbb{R}^d} \varphi(x) e^{-2\pi i x\cdot\xi} dx,$$

so the Fourier inversion formula takes form

$$\varphi(x) = \int_{\mathbb{R}^d} \widehat{\varphi}(\xi) e^{2\pi i x\cdot\xi} d\xi,$$

whenever $\varphi, \widehat{\varphi} \in \mathrm{L}^1(\mathbb{R}^d)$. Derivatives of a single-variable function φ will be denoted φ', φ'', etc. or $D\varphi$, $D^2\varphi$, etc., while we write $\partial^n\varphi$ for the partial derivatives. Let us remark that we reserve the notation $\varphi^{(n)}$ for the upper indices.

Now we can formulate the three propositions that will be the key ingredients in the proof of Theorem 2. Their own proofs will be postponed to the subsequent sections.

Proposition 5. *Let $\lambda > 1$ and let $\vartheta, \varphi \in \mathcal{S}(\mathbb{R})$ be such that*

$$|\vartheta(s)| \leq (1+|s|)^{-\lambda}, \quad |\varphi(s)| \leq (1+|s|)^{-\lambda}$$

for all $s \in \mathbb{R}$. Moreover, assume that $\widehat{\vartheta}$ is supported in $[-2^{-4}, 2^{-4}]$, while $\widehat{\varphi}$ is supported in $[-1,1]$ and constant on $[-2^{-2}, 2^{-2}]$. Then for any $m \in \mathbb{N}$, $k_0, \ldots, k_m \in \mathbb{Z}$, and for any real-valued $F, G \in \mathcal{S}(\mathbb{R}^2)$ normalized by

$$\|F\|_{\mathrm{L}^4(\mathbb{R}^2)} = \|G\|_{\mathrm{L}^4(\mathbb{R}^2)} = 1 \tag{2.1}$$

we have

$$\Big|\sum_{j=1}^{m}\int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v)$$
$$\vartheta_{2^{k_j}}(u)(\varphi_{2^{k_j}}-\varphi_{2^{k_{j-1}}})(v)\,dxdydudv\Big| \lesssim_\lambda 1. \tag{2.2}$$

Proposition 6. *Let $\lambda > 1$, $t > 0$ and let $\Phi \in \mathcal{S}(\mathbb{R}^2)$ be such that*

$$|\Phi(u,v)| \le (1+|u+v|)^{-\lambda} t(1+t|u-v|)^{-\lambda}. \tag{2.3}$$

for all $u, v \in \mathbb{R}$. Moreover, assume that $2^{-2} \le |\xi+\eta| \le 1$ for all (ξ,η) in the support of $\widehat{\Phi}$. Then for any real-valued $F, G \in \mathcal{S}(\mathbb{R}^2)$ normalized as in (2.1) *and for any $N \in \mathbb{N}$ we have*

$$\Big|\sum_{j=-N}^{N}\int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v)\Phi_{2^j}(u,v)\,dxdydudv\Big| \lesssim_\lambda 1. \tag{2.4}$$

Proposition 7. *Let $\lambda > 1$ and let $\Phi \in \mathcal{S}(\mathbb{R}^2)$ be such that*

$$|\Phi(u,v)| \le (1+|u+v|)^{-\lambda}(1+|u-v|)^{-2\lambda} \tag{2.5}$$

for all $u, v \in \mathbb{R}$. Moreover, assume that $\widehat{\Phi}$ is supported in $([-2,-2^{-5}] \cup [2^{-5},2])^2$. Then for any real-valued $F, G \in \mathcal{S}(\mathbb{R}^2)$ normalized as in (2.1) *and for any $N \in \mathbb{N}$ we have* (2.4)*.*

Note that for $\nu = 3\lambda$, the estimate

$$|\Phi(u,v)| \le (1+|u|)^{-\lambda/2}(1+|v|)^{-\lambda/2}(1+|u-v|)^{-\nu} \tag{2.6}$$

implies (2.5) within an absolute constant. Moreover, (2.5) implies (2.6) with $\nu = \lambda$, modulo a constant. We will pass between the two formulations in the subsequent sections.

We also remark that the bump functions in (2.2) do not satisfy any estimates of the type (2.3) or (2.5) within an absolute constant since there is no control on $k_j - k_{j-1}$. However, the form in Proposition 5 has better cancellation properties than the one in Proposition 7. The support of its multiplier symbol does not intersect the antidiagonal $\eta = -\xi$, which is the key property we need in the proof. This is also the case in Proposition 6, which will be the main ingredient in the proof of Proposition 7.

In the rest of this section we concentrate on deducing Theorem 2 from these propositions. Throughout the text, χ will denote a fixed smooth frequency cutoff. More precisely, we fix a function χ such that its Fourier transform $\widehat{\chi}$ is smooth, even, nonnegative, supported in $[-1,1]$, constantly equal to 1 on $[-2^{-1},2^{-1}]$, and monotone on $[2^{-1},1]$. Moreover, we can achieve that $\widehat{\chi}$ is the square of some nonnegative smooth function. Any constants are allowed to depend on χ and this dependence will not be mentioned explicitly.

2.1 Long variation

The following lemma is derived from Propositions 5 and 7.

Lemma 8. *Let $\phi \in \mathcal{S}(\mathbb{R})$ and assume that for some $\lambda > 1$ and constants C_0, C_1 one has*

$$|\phi * \chi_{2^4}(s)| \leq C_0(1+|s|)^{-\lambda}, \quad |\phi(s)| \leq C_1(1+|s|)^{-\lambda} \tag{2.7}$$

for all $s \in \mathbb{R}$, and that for some $\lambda > 1$ and a constant C_2 one has

$$|\phi(u)\overline{\phi(v)}| \leq C_2(1+|u+v|)^{-\lambda}(1+|u-v|)^{-2\lambda} \tag{2.8}$$

for all $u, v \in \mathbb{R}$. Moreover, assume that $\widehat{\phi}$ is supported in $[-1,1]$ and constant on $[-2^{-2}, 2^{-2}]$. If $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ are normalized by (2.1), then

$$\|A^{\phi}_{2^k}(F,G)\|_{\mathrm{V}^2_k(\mathbb{Z},\mathrm{L}^2(\mathbb{R}^2))} \lesssim_\lambda C_0^{1/2}C_1^{1/2} + C_2^{1/2}. \tag{2.9}$$

Observe that if $\widehat{\phi}$ vanishes on $[-2^{-2}, 2^{-2}]$, then the first estimate in (2.7) holds with $C_0 = 0$. In this case Lemma 8 yields

$$\|A^{\phi}_{2^k}(F,G)\|_{\mathrm{V}^2_k(\mathbb{Z},\mathrm{L}^2(\mathbb{R}^2))} \lesssim_\lambda C_2^{1/2}. \tag{2.10}$$

Proof of Lemma 8. Standard limiting arguments reduce the estimate (2.9) for each fixed choice of the integers $k_0 < \cdots < k_m$ to the case of Schwartz functions F and G. By splitting into real and imaginary parts and using Minkowski's inequality, we may assume that F, G, and ϕ take only real values.

Fix integers $k_0 < k_1 < \cdots < k_m$ and denote

$$V(F,G) := \sum_{j=1}^{m} \big\|A^{\phi}_{2^{k_j}}(F,G) - A^{\phi}_{2^{k_{j-1}}}(F,G)\big\|^2_{\mathrm{L}^2(\mathbb{R}^2)}.$$

Expanding the L^2 norm gives

$$\begin{aligned} V(F,G) = \sum_{j=1}^{m} \int_{\mathbb{R}^4} & F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v) \\ & (\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v)\, dxdydudv. \end{aligned}$$

We have the identity

$$\begin{aligned} (\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v) = {} & \big(\phi_{2^{k_{j-1}}}(u)\phi_{2^{k_{j-1}}}(v) - \phi_{2^{k_j}}(u)\phi_{2^{k_j}}(v)\big) \\ & + \phi_{2^{k_j}}(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v) \\ & + (\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(u)\phi_{2^{k_j}}(v). \end{aligned} \tag{2.11}$$

Summing (2.11) over $1 \leq j \leq m$, the first term on the right hand-side telescopes into

$$\phi_{2^{k_0}}(u)\phi_{2^{k_0}}(v) - \phi_{2^{k_m}}(u)\phi_{2^{k_m}}(v).$$

Applying Hölder's inequality in (x,y) for the exponents $(4,4,4,4)$ and using that (2.8) implies $\int_{\mathbb{R}^2} |\phi(u)\phi(v)|dudv \lesssim_\lambda C_2$ we obtain

$$\Big| \int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v) \\ (\phi_{2^{k_0}}(u)\phi_{2^{k_0}}(v) - \phi_{2^{k_m}}(u)\phi_{2^{k_m}}(v))\, dxdydudv\Big| \lesssim_\lambda C_2 \|F\|^2_{\mathrm{L}^4(\mathbb{R}^2)}\|G\|^2_{\mathrm{L}^4(\mathbb{R}^2)} = C_2. \tag{2.12}$$

By symmetry of the second and the third term on the right hand side of (2.11), it then suffices to bound

$$\Lambda(F,G) := \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v) \\ \phi_{2^{k_j}}(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v)\, dxdydudv.$$

Now we localize the multiplier symbol associated with this form. Let ω be defined by $\omega := \chi_{2^{-1}} - \chi_{2^4}$. Note that $\widehat{\omega}$ is supported in $[-2,-2^{-5}] \cup [2^{-5},2]$ and that $\widehat{\chi_{2^4}} + \widehat{\omega}$ equals 1 on $[-1,1]$, and in particular also on the support of $\widehat{\phi}$. Then we can write

$$\phi = \phi * \chi_{2^4} + \phi * \omega.$$

Using this decomposition we split $\Lambda = \Lambda_{\chi_{2^4}} + \Lambda_\omega$, where for a function ρ, the form Λ_ρ is defined by

$$\Lambda_\rho(F,G) := \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v) \\ (\phi*\rho)_{2^{k_j}}(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v)\, dxdydudv.$$

By the assumptions (2.7) on ϕ, Proposition 5 gives

$$|\Lambda_{\chi_{2^4}}(F,G)| \lesssim_\lambda C_0 C_1. \tag{2.13}$$

Rewrite Λ_ω by separating the functions in u and v as

$$\Lambda_\omega(F,G) = \sum_{j=1}^{m} \int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}} F(x+u,y)G(x,y+u)(\phi*\omega)_{2^{k_j}}(u)du \Big) \\ \Big(\int_{\mathbb{R}} F(x+v,y)G(x,y+v)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v)dv \Big)\, dxdy.$$

Applying the Cauchy-Schwarz inequality in x, y, and j gives

$$|\Lambda_\omega(F,G)| \leq \Lambda_{\omega,\omega}(F,G)^{1/2} V(F,G)^{1/2}, \tag{2.14}$$

where for a function ρ we have set

$$\Lambda_{\rho,\rho}(F,G) := \sum_{j=1}^{m} \int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}} F(x+u,y)G(x,y+u)(\phi*\rho)_{2^{k_j}}(u)\, du \Big)^2 dxdy.$$

Note that, up to increasing the quantity $\Lambda_{\omega,\omega}(F,G)$ by adding nonnegative terms, we may assume that $k_j = j$ and that the summation is taken over all integers j from a sufficiently large interval $[-N,N]$. Expanding the square in $\Lambda_{\omega,\omega}(F,G)$ we can write this form as

$$\sum_{j=1}^{m} \int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v)(\phi*\omega)_{2^{k_j}}(u)(\phi*\omega)_{2^{k_j}}(v)\, dxdydudv.$$

By the assumption (2.8), Proposition 7 implies

$$\Lambda_{\omega,\omega}(F,G) \lesssim_\lambda C_2. \tag{2.15}$$

Inequalities (2.12), (2.13), (2.14), and (2.15) together give a bootstrapping estimate

$$V(F,G) \lesssim_\lambda C_2 + C_0C_1 + C_2^{1/2} V(F,G)^{1/2}.$$

This shows $V(F,G) \lesssim_\lambda C_0C_1 + C_2$ and hence proves (2.9). □

2.2 Short variation

The following two closely related lemmata are derived from Proposition 7.

Lemma 9. *Let $\phi \in \mathcal{S}(\mathbb{R})$ and assume that for some $\lambda > 1$ and a constant C_3 one has*

$$\Big| \int_1^2 t\partial_t(\phi_t(u)) t\partial_t(\overline{\phi_t(v)}) \frac{dt}{t} \Big| \le C_3(1+|u+v|)^{-\lambda}(1+|u-v|)^{-2\lambda} \tag{2.16}$$

for all $u,v \in \mathbb{R}$. Moreover, assume that $\widehat{\phi}$ is supported in $[-1,1]$ and constant on $[-2^{-4},2^{-4}]$. If $F,G \in \mathrm{L}^4(\mathbb{R}^2)$ are normalized by (2.1), then for each $N \in \mathbb{N}$ one has

$$\Big(\sum_{i=-N}^{N} \|A_t^\phi(F,G)\|^2_{\mathrm{V}_t^2([2^i,2^{i+1}],\mathrm{L}^2(\mathbb{R}^2))} \Big)^{1/2} \lesssim_\lambda C_3^{1/2}, \tag{2.17}$$

with the implicit constant independent of N.

Lemma 10. *Let ϕ, F, G be as in the previous lemma. If in addition for some $\lambda > 1$ and a constant C_2 the function ϕ satisfies (2.8) for all $u,v \in \mathbb{R}$ and if $\widehat{\phi}$ vanishes on $[-2^{-4},2^{-4}]$, then for each $N \in \mathbb{N}$ we have the estimate*

$$\Big(\sum_{i=-N}^{N} \|A_t^\phi(F,G)\|^2_{\mathrm{V}_t^2([2^i,2^{i+1}],\mathrm{L}^2(\mathbb{R}^2))} \Big)^{1/2} \lesssim_\lambda C_2^{1/4} C_3^{1/4}, \tag{2.18}$$

with the implicit constant independent of N.

Proof of Lemma 9. As in the proof of Lemma 8 we may assume that F, $G \in \mathcal{S}(\mathbb{R}^2)$ and that F, G, and ϕ are real-valued.

Denote $\psi(s) := (s\phi(s))'$, so that one has $\psi_t(s) = -t\partial_t(\phi_t(s))$. By Lemma 12 (in the Appendix) applied with $a(t) = A_t^\phi(F,G)(x,y)$ for each fixed (x,y), for any $2^i \le t_0 < \cdots < t_m \le 2^{i+1}$ we have

$$\sum_{j=1}^{m} \|A_{t_j}^\phi(F,G) - A_{t_{j-1}}^\phi(F,G)\|^2_{\mathrm{L}^2(\mathbb{R}^2)} \le \int_{\mathbb{R}^2} \int_1^2 \big(A_{2^i t}^\psi(F,G)(x,y)\big)^2 \frac{dt}{t} dxdy.$$

Indeed, this follows from $A_t^{\psi}(F,G) = -t\partial_t(A_t^{\phi}(F,G))$ and by rescaling in t. Taking the supremum over all choices of t_j and summing over $-N \le i \le N$ we obtain

$$\sum_{i=-N}^{N} \|A_t^{\phi}(F,G)\|^2_{\mathrm{V}^2_t([2^i,2^{i+1}],\mathrm{L}^2(\mathbb{R}^2))} \le \sum_{i=-N}^{N} \int_{\mathbb{R}^2} \int_1^2 \big(A^{\psi}_{2^i t}(F,G)(x,y)\big)^2 \frac{dt}{t} dxdy.$$

Expanding the square on the right hand-side, in order to finish the proof of Lemma 9 we need to bound

$$\sum_{i=-N}^{N} \int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v) \Big(\int_1^2 \psi_{2^i t}(u)\psi_{2^i t}(v)\frac{dt}{t}\Big)\, dxdydudv. \tag{2.19}$$

Observe that $\widehat{\psi}(\xi) = -\xi\widehat{\phi}'(\xi)$ is supported in $[-1,-2^{-4}] \cup [2^{-4},1]$, so

$$\Phi(u,v) := \int_1^2 \psi_t(u)\psi_t(v)\frac{dt}{t}$$

has its frequency support in $([-1,-2^{-5}] \cup [2^{-5},1])^2$, and recall that we assume (2.16). Proposition 7 implies boundedness of (2.19) within an absolute constant times C_3, which yields (2.17). □

Proof of Lemma 10. Let all the notation and the assumptions be as in the proof of the previous lemma. By Lemma 12 and the Cauchy-Schwarz inequality in (x,y) this time we deduce

$$\sum_{j=1}^{m} \|A^{\phi}_{t_j}(F,G) - A^{\phi}_{t_{j-1}}(F,G)\|^2_{\mathrm{L}^2(\mathbb{R}^2)} \lesssim \prod_{\rho\in\{\phi,\psi\}} \Big(\int_{\mathbb{R}^2} \int_1^2 \big(A^{\rho}_{2^i t}(F,G)(x,y)\big)^2 \frac{dt}{t} dxdy\Big)^{1/2}.$$

Taking the supremum over t_j, summing over $-N \le i \le N$, and applying the Cauchy-Schwarz inequality in i we obtain

$$\sum_{i=-N}^{N} \|A_t^{\phi}(F,G)\|^2_{\mathrm{V}^2_t([2^i,2^{i+1}],\mathrm{L}^2(\mathbb{R}^2))} \le \prod_{\rho\in\{\phi,\psi\}} \Big(\sum_{i=-N}^{N} \int_{\mathbb{R}^2} \int_1^2 \big(A^{\rho}_{2^i t}(F,G)(x,y)\big)^2 \frac{dt}{t} dxdy\Big)^{1/2}.$$

By the support assumptions on ϕ, (2.8), and (2.16), Proposition 7 applied twice concludes that the right hand-side is no greater than an absolute constant times $C_2^{1/2}C_3^{1/2}$, which in turn implies (2.18). □

Finally, we are ready to deduce Theorem 2 from these lemmata. The first step is to show the estimate (1.5) for a general Schwartz function φ.

2.3 Deriving Theorem 2 for a Schwartz function φ

Let $F,G \in \mathcal{S}(\mathbb{R}^2)$ be normalized by (2.1). If $\varphi \in \mathcal{S}(\mathbb{R})$ is such that $\widehat{\varphi}$ is supported in $[-1,1]$ and constant on $[-2^{-2},2^{-2}]$, then Lemmata 8 and 9 combined with the standard separation into long and short jumps imply

$$\big\|A_t^{\varphi}(F,G)\big\|_{\mathrm{V}^2_t((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim_{\lambda} C_0^{1/2}C_1^{1/2} + C_2^{1/2} + C_3^{1/2} \lesssim_{\varphi} 1. \tag{2.20}$$

The details can be found for instance in [21] or [12]. Note that the constants C_i depend only on some Schwartz norm of φ of a sufficiently large degree. This gives (1.5) in the particular case.

Now we show (2.20) for a general Schwartz function φ. Take $\varphi \in \mathcal{S}(\mathbb{R})$ and denote $\theta := \chi - \chi_2$. Observe that $\widehat{\theta}$ is supported in $[-1,-2^{-2}] \cup [2^{-2},1]$ and that

$$\sum_{k\in\mathbb{Z}} \widehat{\theta}(2^k\xi) = 1 \tag{2.21}$$

for all $0 \neq \xi \in \mathbb{R}$. Then we can write

$$\varphi = c\chi + (\varphi - c\chi) = c\chi + \sum_{k\in\mathbb{Z}} (\varphi - c\chi) * \theta_{2^k}, \tag{2.22}$$

where the number c is chosen such that $\widehat{\varphi}(0) - c\widehat{\chi}(0) = 0$, i.e. $c = \widehat{\varphi}(0)$. Note that the series in (2.22) converges pointwise (in any summation order) since $\varphi - c\chi$ and θ are Schwartz and θ has mean zero.

We proceed by bounding norm-variation of bilinear averages corresponding to the individual terms in the expansion (2.22). For the part associated with $c\chi$ boundedness follows from (2.20) since χ is Schwartz and $\widehat{\chi}$ is constant near the origin:

$$\|A_t^{c\chi}(F,G)\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 1. \tag{2.23}$$

For the part associated with $(\varphi - c\chi) * \theta_{2^k}$ we show that the function $\vartheta = \vartheta^{(k)}$ defined by

$$\vartheta := (\varphi - c\chi)_{2^{-k}} * \theta$$

satisfies the estimate

$$\big\|A_t^{\vartheta}(F,G)\big\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 2^{-|k|} \tag{2.24}$$

for any $k \in \mathbb{Z}$. By scaling invariance of the left hand-side of (2.24) the same estimate remains to hold for $\vartheta_{2^k} = (\varphi - c\chi) * \theta_{2^k}$, i.e. for each term in the series expansion (2.22). Then, from (2.22), (2.23), (2.24), Minkowski's inequality, and Fatou's lemma we obtain

$$\|A_t^{\varphi}(F,G)\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 1 + \sum_{k\in\mathbb{Z}} 2^{-|k|} \lesssim 1,$$

which finishes the proof.

In order to verify (2.24), observe that $\widehat{\vartheta}$ is supported in $[-1,-2^{-2}] \cup [2^{-2},1]$, so in particular it is constant on $[-2^{-2},2^{-2}]$. Since $\widehat{\varphi} - c\widehat{\chi}$ vanishes at zero, we have $|\widehat{\varphi}(\xi) - c\widehat{\chi}(\xi)| \lesssim_{\varphi} \min\{|\xi|,|\xi|^{-1}\}$ and hence, by $\widehat{\vartheta}(\xi) = (\widehat{\varphi} - c\widehat{\chi})(2^{-k}\xi)\widehat{\theta}(\xi)$ and the product rule,

$$\big\||\xi|^{\alpha} D^{\beta}\widehat{\vartheta}(\xi)\big\|_{\mathrm{L}_{\xi}^{\infty}(\mathbb{R})} \lesssim_{\alpha,\beta} 2^{-|k|}$$

for any $\alpha,\beta \geq 0$. Therefore, $2^{|k|}\vartheta$ satisfies (2.7), (2.8), and (2.16) with the constants independent of k. The estimate (2.24) then follows from (2.20) applied with $\varphi = 2^{|k|}\vartheta$ and by homogeneity.

2.4 Deriving Theorem 2 for $\varphi = \mathbb{1}_{[0,1)}$

Once again we can work with Schwartz functions F and G only. Let $F, G \in \mathcal{S}(\mathbb{R}^2)$ be normalized by (2.1) and let χ, θ be as in the previous subsection. We have

$$\mathbb{1}_{[0,1)} = \mathbb{1}_{[0,1)} * \chi + \sum_{k=-\infty}^{-1} \mathbb{1}_{[0,1)} * \theta_{2^k}. \tag{2.25}$$

By the Plancherel identity the series in (2.25) converges in the L^2 norm. However, the same series also converges a.e., which follows from the weak L^2 boundedness of the maximally truncated convolution-type singular integrals. Alternatively, we can pass to an a.e. convergent subsequence of partial sums, as taking the limit over a subsequence is enough for our intended application.

By the discussion in Subsection 2.3 we obtain

$$\big\|A_t^{\mathbb{1}_{[0,1)}*\chi}(F,G)\big\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 1. \tag{2.26}$$

Now we concentrate on the individual terms in (2.25) for negative values of k. By $\widetilde{\theta}$ we denote the primitive of θ, i.e. $\widetilde{\theta}(s) := \int_{-\infty}^{s} \theta(u)du$. Observe that, since θ has integral zero, its primitive $\widetilde{\theta}$ decays rapidly. The arguments from the previous subsection give

$$\|A_t^{\widetilde{\theta}}(F,G)\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 1. \tag{2.27}$$

By scaling invariance of the left hand-side, (2.27) also holds with $\widetilde{\theta}$ replaced by $\widetilde{\theta}_{2^k}$. We will show that for each $k < 0$ and for the function $\vartheta = \vartheta^{(k)}$ defined by

$$\vartheta(s) := 2^k \widetilde{\theta}(s - 2^{-k})$$

we have the variational inequality

$$\|A_t^{\vartheta}(F,G)\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 2^{k/8}. \tag{2.28}$$

Once this is shown, by scaling invariance of the left hand-side, the estimate (2.28) remains to hold with ϑ replaced by ϑ_{2^k}. Then we need to observe that

$$\mathbb{1}_{[0,1)} * \theta_{2^k} = 2^k \widetilde{\theta}_{2^k} - \vartheta_{2^k}.$$

From (2.25), (2.26), (2.27), (2.28), Minkowski's inequality, and Fatou's lemma we finally obtain

$$\|A_t^{\mathbb{1}_{[0,1)}}(F,G)\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim 1 + \sum_{k\leq -1} (2^k + 2^{k/8}) \lesssim 1.$$

In order to see (2.28), note that the Fourier support of ϑ is contained in $[-1,-2^{-2}] \cup [2^{-2},1]$. For any $\lambda > 0$, $\nu > 0$, and $k < 0$ we claim that

$$|\vartheta(u)\vartheta(v)| \lesssim_{\lambda,\nu} 2^{k(2-\lambda)}(1+|u|)^{-\lambda/2}(1+|v|)^{-\lambda/2}(1+|u-v|)^{-\nu}, \tag{2.29}$$

$$\Big|\int_1^2 t\partial_t(\vartheta_t(u))t\partial_t(\vartheta_t(v))\frac{dt}{t}\Big| \lesssim_{\lambda,\nu} 2^{k(1-\lambda)}(1+|u|)^{-\lambda/2}(1+|v|)^{-\lambda/2}(1+|u-v|)^{-\nu}. \tag{2.30}$$

We have already commented how bounds of this form with $\nu = 3\lambda$ transform into bounds (2.8) and (2.16). Once these two estimates are verified, the separation into short and long jumps together with (2.10) and Lemma 10, which require (2.29) and (2.30) to hold with $\lambda > 1$, give

$$\big\| A_t^{\vartheta}(F,G) \big\|_{\mathrm{V}_t^2((0,\infty),\mathrm{L}^2(\mathbb{R}^2))} \lesssim_\lambda C_2^{1/2} + C_2^{1/4} C_3^{1/4}$$

with $C_2 \sim 2^{k(2-\lambda)}$ and $C_3 \sim 2^{k(1-\lambda)}$. Choosing $\lambda = 5/4$ we obtain (2.28).

Proof of (2.29). By the rapid decay of $\widetilde{\theta}$ we have

$$\begin{aligned} |\tilde{\theta}(u-2^{-k})\tilde{\theta}(v-2^{-k})| &\lesssim_{\lambda,\nu} (1+|u-2^{-k}|)^{-\lambda/2-\nu}(1+|v-2^{-k}|)^{-\lambda/2-\nu} \\ &\le (1+|u-2^{-k}|)^{-\lambda/2}(1+|v-2^{-k}|)^{-\lambda/2}(1+|u-v|)^{-\nu}, \end{aligned}$$

where we used $|u-v| \le |u-2^{-k}| + |v-2^{-k}|$. From

$$(1+|u-2^{-k}|)^{-\lambda/2} \le (1+|u|)^{-\lambda/2}(1+2^{-k})^{\lambda/2} \lesssim_\lambda (1+|u|)^{-\lambda/2} 2^{-k\lambda/2}$$

we then conclude (2.29). □

Proof of (2.30). Observe that $-t\partial_t(\vartheta_t(s)) = \vartheta_t(s) + (s\vartheta'(s))_t$. Thus, $t\partial_t(\vartheta_t(u)) t\partial_t(\vartheta_t(v))$ consist of four terms. We will show (2.30) corresponding to $(s\vartheta'(s))_t$, that is,

$$\Big| \int_1^2 (u\vartheta'(u))_t (v\vartheta'(v))_t \frac{dt}{t} \Big| \lesssim_{\lambda,\nu} 2^{k(1-\lambda)}(1+|u|)^{-\lambda/2}(1+|v|)^{-\lambda/2}(1+|u-v|)^{-\nu}. \quad (2.31)$$

The analogous inequalities corresponding to the other terms are treated in the same manner. To see (2.31) we first observe

$$(s\vartheta'(s))_t = (s2^k\theta(s-2^{-k}))_t = st^{-1}2^k\theta_t(s-t2^{-k})$$

and bound $|\theta_t(s)| \lesssim_{\lambda,\nu} (1+|s|)^{-\lambda/2-\nu-1}$ using $t \in [1,2]$. Then we estimate

$$\begin{aligned} &\Big| \int_1^2 u\theta_t(u-t2^{-k}) v\theta_t(v-t2^{-k}) \frac{dt}{t^3} \Big| \\ &\lesssim_{\lambda,\nu} |uv| \int_1^2 \big((1+|u-t2^{-k}|)(1+|v-t2^{-k}|)\big)^{-\lambda/2-\nu-1} dt. \end{aligned}$$

By the triangle inequality $|u-v| \le |u-t2^{-k}| + |v-t2^{-k}|$ and the Cauchy-Schwarz inequality in t, this is bounded by

$$(1+|u-v|)^{-\nu}|u|\Big(\int_1^2 (1+|u-t2^{-k}|)^{-\lambda-2} dt\Big)^{1/2} |v|\Big(\int_1^2 (1+|v-t2^{-k}|)^{-\lambda-2} dt\Big)^{1/2}.$$

Now, if $|u| \le 2^{-k+2}$, then we estimate

$$\begin{aligned} &(1+|u|)^{\lambda/2}|u|\Big(\int_1^2 (1+|u-t2^{-k}|)^{-\lambda-2} dt\Big)^{1/2} \\ &\le (1+|u|)^{\lambda/2+1}\Big(\int_{-\infty}^{\infty} (1+|u-t2^{-k}|)^{-\lambda-2} dt\Big)^{1/2} \\ &\lesssim_\lambda 2^{k/2}(1+|u|)^{\lambda/2+1} \lesssim_\lambda 2^{k(-1-\lambda)/2}, \end{aligned}$$

where the second inequality follows by integrating in t. If $|u| \geq 2^{-k+2}$, then we have $|u - t2^{-k}| \geq |u|/2$ and hence

$$\begin{aligned}&(1+|u|)^{\lambda/2}|u|\Big(\int_1^2 (1+|u-t2^{-k}|)^{-\lambda-2}dt\Big)^{1/2}\\ &\lesssim_\lambda (1+|u|)^{\lambda/2+1}(1+|u|)^{-\lambda/2-1} = 1 \leq 2^{k(-1-\lambda)/2}.\end{aligned}$$

The same estimates hold for the terms with v. After multiplication by 2^{2k} and division by $(1+|u|)^{\lambda/2}(1+|v|)^{\lambda/2}$ this shows (2.31). □

3 Proof of Proposition 5

Let us rewrite the form (2.2) from Proposition 5 in a more convenient way. Denote $\psi := \varphi - \varphi_2$. Then we have the telescoping identity

$$\varphi_{2^{k_{j-1}}} - \varphi_{2^{k_j}} = \sum_{l=k_{j-1}}^{k_j-1} \psi_{2^l}. \tag{3.1}$$

We insert (3.1) into (2.2) and substitute

$$x' = x+y+u, \quad y' = x+y+v, \quad \widetilde{F}(y,x') := F(x'-y,y), \quad \widetilde{G}(x,x') := G(x,x'-x).$$

Note that we still have $\|\widetilde{F}\|_{\mathrm{L}^4(\mathbb{R}^2)} = \|\widetilde{G}\|_{\mathrm{L}^4(\mathbb{R}^2)} = 1$. Omitting the tildas for notational simplicity, it then suffices to show the inequality

$$\begin{aligned}\Big|\sum_{j=1}^{m}\sum_{l=k_{j-1}}^{k_j-1}\int_{\mathbb{R}^4} &F(y,x')G(x,x')F(y,y')G(x,y')\\ &\vartheta_{2^{k_j}}(x'-x-y)\psi_{2^l}(y'-x-y)\,dxdydx'dy'\Big| \lesssim 1.\end{aligned}$$

First, we would like to write the kernel as a superposition of elementary tensors in the four variables x, y, x', y'. Using the Fourier inversion formula we write

$$\vartheta_{2^{k_j}}(x'-x-y)\psi_{2^l}(y'-x-y) = \int_{\mathbb{R}^2} \widehat{\vartheta}(2^{k_j}\xi)\widehat{\psi}(2^l\eta)e^{2\pi i\xi(x'-x-y)}e^{2\pi i\eta(y'-x-y)}d\xi d\eta.$$

Since $\widehat{\varphi}$ is supported in $[-1,1]$ and constant on $[-2^{-2},2^{-2}]$, the function $\widehat{\psi}$ is supported in $[-1,-2^{-3}] \cup [2^{-3},1]$. If $2^{k_j}\xi \in \operatorname{supp}(\widehat{\vartheta})$ and $2^l\eta \in \operatorname{supp}(\widehat{\psi})$, then

$$2^l(\xi+\eta) = 2^{l-k_j}2^{k_j}\xi + 2^l\eta \in [-2,-2^{-4}] \cup [2^{-4},2].$$

Let χ be as before, which guarantees that there exists a smooth nonnegative even function $\widehat{\omega}$, being the Fourier transform of some $\omega \in \mathcal{S}(\mathbb{R})$, satisfying

$$\widehat{\omega}(\xi)^2 = \widehat{\chi}(2^{-2}\xi) - \widehat{\chi}(2^4\xi).$$

The function $\widehat{\omega}$ is supported in $[-2^2,-2^{-5}] \cup [2^{-5},2^2]$ and equal to 1 on $[-2,-2^{-4}] \cup [2^{-4},2]$. For each $(\xi,\eta) \in \mathbb{R}^2$ we have

$$\widehat{\vartheta}(2^{k_j}\xi)\widehat{\psi}(2^l\eta) = \widehat{\vartheta}(2^{k_j}\xi)\widehat{\psi}(2^l\eta)\widehat{\omega}(2^l(\xi+\eta))^2 \tag{3.2}$$

and hence

$$\vartheta_{2^{k_j}}(x'-x-y)\psi_{2^l}(y'-x-y)$$
$$=\int_{\mathbb{R}^2}\widehat{\vartheta}(2^{k_j}\xi)e^{2\pi i x'\xi}\widehat{\psi}(2^l\eta)e^{2\pi i y'\eta}\widehat{\omega}(2^l(-\xi-\eta))e^{2\pi i x(-\xi-\eta)}\widehat{\omega}(2^l(-\xi-\eta))e^{2\pi i y(-\xi-\eta)}d\xi d\eta.$$

The last expression can be viewed as the integral of the Fourier transform of the function

$$\mathcal{H}(x_1,x_2,x_3,x_4):=\vartheta_{2^{k_j}}(x_1+x')\psi_{2^l}(x_2+y')\omega_{2^l}(x_3+x)\omega_{2^l}(x_4+y)$$

over the hyperplane

$$\{(\xi,\eta,-\xi-\eta,-\xi-\eta):\xi,\eta\in\mathbb{R}\}.$$

It equals the integral of $\mathcal{H}$ itself over the perpendicular hyperplane

$$\{(p+q,p+q,p,q):p,q\in\mathbb{R}\}.$$

Therefore, $\vartheta_{2^{k_j}}(x'-x-y)\psi_{2^l}(y'-x-y)$ can be written as

$$\int_{\mathbb{R}^2}\vartheta_{2^{k_j}}(x'-p-q)\psi_{2^l}(y'-p-q)\omega_{2^l}(x-p)\omega_{2^l}(y-q)\,dpdq$$

and the object we need to bound is

$$\sum_{j=1}^{m}\sum_{l=k_{j-1}}^{k_j-1}\int_{\mathbb{R}^6}F(y,x')G(x,x')F(y,y')G(x,y')$$
$$\vartheta_{2^{k_j}}(x'-p-q)\psi_{2^l}(y'-p-q)\omega_{2^l}(x-p)\omega_{2^l}(y-q)\,dxdydx'dy'dpdq. \tag{3.3}$$

In order to estimate this form we adapt the arguments from [24] to the Euclidean setting. First we apply the Cauchy-Schwarz inequality, which will reduce the complexity of the form. To preserve the mean zero property of ω we rewrite (3.3) as

$$\sum_{j=1}^{m}\sum_{l=k_{j-1}}^{k_j-1}\int_{\mathbb{R}^4}\Big(\int_{\mathbb{R}}F(y,x')F(y,y')\omega_{2^l}(y-q)\,dy\Big)\Big(\int_{\mathbb{R}}G(x,x')G(x,y')\omega_{2^l}(x-p)\,dx\Big)$$
$$\vartheta_{2^{k_j}}(x'-p-q)\psi_{2^l}(y'-p-q)\,dx'dy'dpdq.$$

Taking absolute values, using the triangle inequality, and applying the Cauchy-Schwarz inequality in the variables x', y', p, q, and t, we bound this expression by

$$\Gamma(F)^{1/2}\Gamma(G)^{1/2}, \tag{3.4}$$

where we have denoted

$$\Gamma(F):=\sum_{j=1}^{m}\sum_{l=k_{j-1}}^{k_j-1}\int_{\mathbb{R}^4}\Big(\int_{\mathbb{R}}F(y,x')F(y,y')\omega_{2^l}(y-q)dy\Big)^2$$
$$|\vartheta|_{2^{k_j}}(x'-p)|\psi|_{2^l}(y'-p)\,dx'dy'dpdq.$$

Here the two appearances of the function ω have been separated, which allowed us to change variables $p \to p - q$ in the last expression. Integrating in p, using $l \le k_j$ and the normalization of ϑ and φ, we get

$$\int_{\mathbb{R}} |\vartheta|_{2^{k_j}}(x' - p)|\psi|_{2^l}(y' - p)dp \lesssim_\lambda 2^{-k_j}(1 + 2^{-k_j}|x' - y'|)^{-\lambda}. \tag{3.5}$$

This fact can be shown along the lines of [37, Lemma 2.1]. For completeness and to keep track of the constants we now give a detailed proof.

If $|x' - y'| \le 2^{k_j+1}/(\lambda - 1)$, then we can bound the left hand-side of (3.5) by

$$\|\vartheta_{2^{k_j}}\|_{\mathrm{L}^\infty(\mathbb{R})}\|\psi_{2^l}\|_{\mathrm{L}^1(\mathbb{R})} \lesssim_\lambda \|\vartheta\|_{\mathrm{L}^\infty(\mathbb{R})}\|\psi\|_{\mathrm{L}^1(\mathbb{R})}2^{-k_j}(1 + 2^{-k_j}|x' - y'|)^{-\lambda}.$$

If $|x' - y'| \ge 2^{k_j+1}/(\lambda - 1)$, then let us denote by c the midpoint of x' and y'. Without loss of generality we may assume $x' < c < y'$. We split the integral as $\int_{\mathbb{R}} = \int_{-\infty}^{c} + \int_{c}^{\infty}$ and estimate it by

$$\|\vartheta\|_{\mathrm{L}^1(\mathbb{R})}2^{-l}(1 + 2^{-l}|y' - c|)^{-\lambda} + 2^{-k_j}(1 + 2^{-k_j}|x' - c|)^{-\lambda}\|\psi\|_{\mathrm{L}^1(\mathbb{R})}. \tag{3.6}$$

Since $|x' - c| = |y' - c| = |x' - y'|/2,\ l \le k_j$,

$$2^{-l-1}|x' - y'| \ge 2^{-k_j-1}|x' - y'| \ge (\lambda - 1)^{-1},$$

and the function $s \mapsto s(1+s)^{-\lambda}$ is decreasing on the interval $[(\lambda-1)^{-1}, \infty)$, the expression (3.6) is at most

$$(\|\vartheta\|_{\mathrm{L}^1(\mathbb{R})} + \|\psi\|_{\mathrm{L}^1(\mathbb{R})})\, 2^{-k_j}(1 + 2^{-k_j-1}|x' - y'|)^{-\lambda}.$$

It remains to note $\|\vartheta\|_{\mathrm{L}^\infty(\mathbb{R})} \le 1$, $\|\vartheta\|_{\mathrm{L}^1(\mathbb{R})} \lesssim_\lambda 1$, and $\|\psi\|_{\mathrm{L}^1(\mathbb{R})} \lesssim_\lambda 1$, which shows the claim.

Our inequality did not preserve the tensor structure in the variables x' and y' which will be needed later in (3.13). For that purpose we further estimate (3.5) by a superposition of Gaussians as it was done in [14]. Denote

$$g(s) := e^{-\pi s^2} \quad \text{and} \quad \sigma(s) := \int_1^\infty g_\alpha(s)\alpha^{-\lambda}d\alpha, \tag{3.7}$$

where $g_\alpha(s) = \alpha^{-1}g(\alpha^{-1}s)$, as before. Observe that $\sigma(0) = (\lambda - 1)^{-1}$ and the change of variables $\beta = |s|/\alpha$ gives

$$\lim_{|s|\to\infty} |s|^\lambda \sigma(s) = \int_0^\infty \beta^{\lambda-1}e^{-\pi\beta^2}d\beta \in (0, \infty),$$

so $\sigma(s)$ is comparable to $|s|^{-\lambda}$ for large $|s|$. Therefore, using

$$(1 + |s|)^{-\lambda} \lesssim_\lambda \sigma(s) \tag{3.8}$$

we can dominate the right hand-side of (3.5) up to a positive constant by $\sigma_{2^{k_j}}(x' - y')$. This in turn controls

$$\Gamma(F) \lesssim_\lambda \int_1^\infty \Big(\sum_{j=1}^{m} \sum_{l=k_{j-1}}^{k_j-1} \int_{\mathbb{R}^5} F(y, x')F(x, x')F(y, y')F(x, y')$$
$$g_{\alpha 2^{k_j}}(x' - y')\omega_{2^l}(x - q)\omega_{2^l}(y - q)\, dxdydx'dy'dq \Big)\alpha^{-\lambda}d\alpha. \tag{3.9}$$

Integrating in q, summing in l, and using $\widehat{\omega}(\xi)^2 = \sum_{i=-2}^{3} \big(\widehat{\chi}(2^i\xi) - \widehat{\chi}(2^{i+1}\xi)\big)$ we obtain

$$\sum_{l=k_{j-1}}^{k_j-1} \int_{\mathbb{R}} \omega_{2^l}(x-q)\omega_{2^l}(y-q)\,dq = \sum_{i=-2}^{3} (\chi_{2^{k_{j-1}+i}} - \chi_{2^{k_j+i}})(x-y).$$

Inserting this into (3.9), the integrand in α can be rewritten as

$$\sum_{i=-2}^{3}\sum_{j=1}^{m} \int_{\mathbb{R}^4} F(y,x')F(x,x')F(y,y')F(x,y')$$
$$g_{\alpha 2^{k_j}}(x'-y')(\chi_{2^{k_{j-1}+i}} - \chi_{2^{k_j+i}})(x-y)\,dxdydx'dy'.$$

It suffices to prove an estimate uniform in α for each summand corresponding to a fixed i and then integrate in α and sum over $-2 \le i \le 3$. For two functions $\tilde{\rho}, \rho \in \mathcal{S}(\mathbb{R})$ define

$$\Theta_{\tilde{\rho},\rho}(F) := \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(y,x')F(x,x')F(y,y')F(x,y')$$
$$\tilde{\rho}_{2^{k_j}}(x'-y')(\rho_{2^{k_{j-1}}} - \rho_{2^{k_j}})(x-y)\,dxdydx'dy'.$$

The needed estimate is a direct consequence of the following lemma applied with $\rho = \chi_{2^i}$.

Lemma 11. *For any real-valued $F \in \mathcal{S}(\mathbb{R}^2)$, real-valued $\rho \in \mathcal{S}(\mathbb{R})$ and $\alpha \in (0,\infty)$ we have*

$$\Theta_{g_\alpha,\rho}(F) \lesssim_\rho \|F\|_{\mathrm{L}^4(\mathbb{R}^2)}^4, \tag{3.10}$$

where $g(s) = e^{-\pi s^2}$.

Proof. Once again we normalize F as in (2.1). The first step is an application of the telescoping identity. If we denote

$$\widetilde{\Theta}_{\tilde{\rho},\rho}(F) := \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(y,x')F(x,x')F(y,y')F(x,y')$$
$$(\tilde{\rho}_{2^{k_{j-1}}} - \tilde{\rho}_{2^{k_j}})(x'-y')\rho_{2^{k_{j-1}}}(x-y)\,dxdydx'dy'$$

and for $t > 0$ define the single-scale quantity

$$\Xi_{\tilde{\rho},\rho,t}(F) := \int_{\mathbb{R}^4} F(y,x')F(x,x')F(y,y')F(x,y')\tilde{\rho}_t(x'-y')\rho_t(x-y)\,dxdydx'dy',$$

then we have

$$\Theta_{\tilde{\rho},\rho}(F) + \widetilde{\Theta}_{\tilde{\rho},\rho}(F) = \Xi_{\tilde{\rho},\rho,2^{k_0}}(F) - \Xi_{\tilde{\rho},\rho,2^{k_m}}(F), \tag{3.11}$$
$$\Xi_{\tilde{\rho},\rho,t}(F) \le \|\tilde{\rho}\|_{\mathrm{L}^1(\mathbb{R})}\|\rho\|_{\mathrm{L}^1(\mathbb{R})}. \tag{3.12}$$

The identity (3.11) follows from summation by parts: all intermediate terms cancel. To see (3.12) we substitute $u = x'-y'$, $v = x-y$, rewrite $\Xi_{\tilde{\rho},\rho,t}(F)$ as

$$\int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}^2} F(x-v,x')F(x,x')F(x-v,x'-u)F(x,x'-u)\,dxdx' \Big)\, \tilde{\rho}_t(u)\rho_t(v)\,dudv,$$

and apply Hölder's inequality in (x, x') for the exponents $(4, 4, 4, 4)$.

In order to show (3.10) we first use (3.11), which gives

$$\Theta_{g_\alpha,\rho}(F) = \Xi_{g_\alpha,\rho,2^{k_0}}(F) - \Xi_{g_\alpha,\rho,2^{k_m}}(F) - \widetilde{\Theta}_{g_\alpha,\rho}(F),$$

and hence applying (3.12) we get

$$|\Theta_{g_\alpha,\rho}(F)| \le |\Xi_{g_\alpha,\rho,2^{k_0}}(F)| + |\Xi_{g_\alpha,\rho,2^{k_m}}(F)| + |\widetilde{\Theta}_{g_\alpha,\rho}(F)| \lesssim_\rho 1 + \big|\widetilde{\Theta}_{g_\alpha,\rho}(F)\big|.$$

Therefore, it remains to estimate $\big|\widetilde{\Theta}_{g_\alpha,\rho}(F)\big|$.

By the fundamental theorem of calculus we rewrite $\widetilde{\Theta}_{g_\alpha,\rho}(F)$ as

$$\widetilde{\Theta}_{g_\alpha,\rho}(F) = \sum_{j=1}^{m} \int_{2^{k_{j-1}}}^{2^{k_j}} \int_{\mathbb{R}^4} F(y, x')F(x, x')F(y, y')F(x, y') \\ \big(- t\partial_t(g_{\alpha t}(x' - y'))\big)\rho_{2^{k_{j-1}}}(x - y)\, dxdydx'dy'\frac{dt}{t}.$$

For $h(s) := \sqrt{2/\pi}g'(\sqrt{2}s)$ we have $-t\partial_t(\widehat{g_{\alpha t}}(\xi)) = |\widehat{h_{\alpha t}}(\xi)|^2$ and hence

$$- t\partial_t(g_{\alpha t}(x' - y')) = \int_{\mathbb{R}} h_{\alpha t}(x' - p)h_{\alpha t}(y' - p)dp. \tag{3.13}$$

By this identity and the symmetry of $\widetilde{\Theta}_{g_\alpha,\rho}$, which results from four repetitions of the function F, we can express $\widetilde{\Theta}_{g_\alpha,\rho}(F)$ as

$$\sum_{j=1}^{m} \int_{2^{k_{j-1}}}^{2^{k_j}} \int_{\mathbb{R}^3} \Big(\int_{\mathbb{R}} F(y, x')F(x, x')h_{\alpha t}(x' - p)dx'\Big)^2 \rho_{2^{k_{j-1}}}(x - y)\, dxdydp\frac{dt}{t}. \tag{3.14}$$

Observe that the square in (3.14) is automatically non-negative, but the function ρ is not non-negative in general. To obtain positivity and an elementary tensor structure in x and y as in (3.13) we dominate $|\rho| \lesssim \sigma$ by applying (3.8) as before, where σ is the superposition of the Gaussians (3.7). This implies

$$\big|\widetilde{\Theta}_{g_\alpha,\rho}(F)\big| \lesssim_\rho \int_1^\infty \widetilde{\Theta}_{g_\alpha,g_\beta}(F)\beta^{-\lambda}d\beta.$$

We apply the telescoping identity (3.11) once more to get

$$\widetilde{\Theta}_{g_\alpha,g_\beta}(F) = \Xi_{g_\alpha,g_\beta,2^{k_0}}(F) - \Xi_{g_\alpha,g_\beta,2^{k_m}}(F) - \Theta_{g_\alpha,g_\beta}(F).$$

Now that we have reduced to Gaussian functions only, we have non-negativity of both $\Theta_{g_\alpha,g_\beta}(F)$ and $\widetilde{\Theta}_{g_\alpha,g_\beta}(F)$. This can be seen by the fundamental theorem of calculus and the equality (3.13), which allow us to write $\Theta_{g_\alpha,g_\beta}(F)$ and $\widetilde{\Theta}_{g_\alpha,g_\beta}(F)$ in the same way as we did with the form in (3.14). Therefore, by (3.12) once again,

$$\widetilde{\Theta}_{g_\alpha,g_\beta}(F) \le \Xi_{g_\alpha,g_\beta,2^{k_0}}(F) - \Xi_{g_\alpha,g_\beta,2^{k_m}}(F) \le 2\|g\|_{\mathrm{L}^1(\mathbb{R})}^2 \lesssim 1.$$

This finishes the proof of Lemma 11. □

4 Proof of Proposition 6

Our aim is to reduce the proposition to Lemma 11 from the previous section. Let χ and ω be the functions as in Section 3. Then $\widehat{\omega}$ equals 1 on $\{\xi+\eta : (\xi,\eta)\in \mathrm{supp}(\widehat{\Phi})\}$, so for each $(\xi,\eta)\in\mathbb{R}^2$ we can write

$$\widehat{\Phi}(\xi,\eta)=\widehat{\Phi}(\xi,\eta)\widehat{\omega}(\xi+\eta)^2,$$

similarly as in (3.2). Choosing the same substitution as in Section 3 and performing the analogous steps from (3.2) to (3.4) with k_j and l being replaced by j, it remains to estimate an analogous quantity to $\Gamma(F)$,

$$\sum_{j=-N}^{N}\int_{\mathbb{R}^6} F(y,x')F(x,x')F(y,y')F(x,y')$$
$$|\Phi|_{2^j}(x'-p,y'-p)\omega_{2^j}(x-q)\omega_{2^j}(y-q)\,dxdydx'dy'dpdq.$$

Using the decay assumption on Φ we obtain

$$\int_{\mathbb{R}}|\Phi|(x'-p,y'-p)\,dp \le \int_{\mathbb{R}} t(1+t|x'-y'|)^{-\lambda}(1+|x'+y'-2p|)^{-\lambda}dp$$
$$\lesssim_\lambda t(1+t|x'-y'|)^{-\lambda}.$$

Estimating the right hand-side as in (3.8) by the superposition σ defined in (3.7) and proceeding as we did with (3.9), it then suffices to bound

$$\sum_{j=-N}^{N}\int_{\mathbb{R}^4} F(y,x')F(x,x')F(y,y')F(x,y')g_{\alpha t 2^j}(x'-y')(\chi_{2^{j+i}}-\chi_{2^{j+i+1}})(x-y)\,dxdydx'dy'$$

uniformly in $\alpha,t\in(0,\infty)$ and for each fixed $-2\le i\le 3$. Such an estimate follows from the particular case of Lemma 11 when $\rho=\chi_{2^{i+1}}$ and $k_j=j$.

5 Proof of Proposition 7

We would like to decompose the kernel of the form appearing on the left hand side of (2.4) into elementary tensors analogous to those from Section 3. Then we could bound this form by the Cauchy-Schwarz inequality and iterations of the telescoping identity and positivity arguments. However, the multiplier support now intersects the axis $\eta=-\xi$, so a desired decomposition is not readily available.

To overcome this issue, the idea is to transfer to the multiplier with the symbol (5.8) below, which is homogeneous, i.e. constant on the rays through the origin, symmetric with respect to $\eta=-\xi$, and smooth away from that axis. Since the form with a constant multiplier is trivially bounded, we can then subtract the constant on $\eta=-\xi$ from that homogeneous multiplier. This leaves us with a function vanishing on $\eta=-\xi$ up to a certain positive order. By a bi-parameter lacunary decomposition with respect to the axes $\eta=\xi$ and $\eta=-\xi$ we reduce to the consideration of certain angular regions to which the arguments from Section 3 may be applied. Due to the vanishing along $\eta = -\xi$ we are able to sum over all such regions.

We can assume that $1 < \lambda < 2$, as the claim only becomes stronger as λ decreases to 1. Recall that the form from Proposition 7 is associated with the kernel

$$K(u,v) := \sum_{j=-N}^{N} \Phi_{2^j}(u,v).$$

Let θ be $\chi - \chi_2$, so that $\widehat{\theta}$ partitions the unity as in (2.21). Then $\int_0^\infty \widehat{\theta}(t\tau)\frac{dt}{t}$ is the same constant for all $0 \neq \tau \in \mathbb{R}$ and up to that constant $\widehat{K}(\xi,\eta)$ equals

$$\int_0^\infty \widehat{K}(\xi,\eta)\widehat{\theta}(t|(\xi,\eta)|)\frac{dt}{t} = \int_0^\infty \widehat{K^{(t)}}(t(\xi,\eta))\frac{dt}{t} \tag{5.1}$$

for all $(\xi,\eta) \neq (0,0)$, where $K^{(t)}$ is defined via its Fourier transform as

$$\widehat{K^{(t)}}(\xi,\eta) := \widehat{K}(t^{-1}(\xi,\eta))\widehat{\theta}(|(\xi,\eta)|).$$

Observe that the support of $\widehat{K^{(t)}}(\xi,\eta)$ lies in the intersection of the annulus $2^{-2} \leq |(\xi,\eta)| \leq 1$ with the quadruple cone $2^{-6} \leq |\eta/\xi| \leq 2^6$, which in turn is contained in the Cartesian product

$$([-1,-2^{-9}] \cup [2^{-9},1])^2. \tag{5.2}$$

Let ϑ be such that $\widehat{\vartheta}$ is a smooth nonnegative even function supported in $[-2,-2^{-10}] \cup [2^{-10},2]$ and such that $(\xi,\eta) \mapsto \widehat{\vartheta}(\xi)\widehat{\vartheta}(\eta)$ equals 1 on the set (5.2) and thus also on the support of each $\widehat{K^{(t)}}$. Then

$$\widehat{K^{(t)}}(\xi,\eta) = \widehat{K^{(t)}}(\xi,\eta)\widehat{\vartheta}(\xi)\widehat{\vartheta}(\eta),$$

which implies

$$K^{(t)}(u,v) = \int_{\mathbb{R}^2} K^{(t)}(a,b)\vartheta(u-a)\vartheta(v-b)\,dadb. \tag{5.3}$$

Using (5.1) and (5.3), the form from Proposition 7 can be rewritten as

$$\int_{\mathbb{R}^2}\int_0^\infty K^{(t)}(a,b)\int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v)$$
$$\vartheta_t(u-ta)\vartheta_t(v-tb)\,dxdydudv\,\frac{dt}{t}\,dadb. \tag{5.4}$$

Observe that for $\kappa \in \mathcal{S}(\mathbb{R}^2)$ defined by $\widehat{\kappa}(\xi,\eta) = \widehat{\theta}(|(\xi,\eta)|)$ we have

$$K^{(t)}(a,b) = \sum_{j=-N}^{N}\int_{\mathbb{R}^2} \Phi_{2^j/t}(a-x,b-y)\kappa(x,y)\,dxdy$$

and by the support conditions on $\widehat{\Phi}$ and $\widehat{\kappa}$ the sum is taken only over $-N \leq j \leq N$ that also satisfy $2^{-7} < 2^j/t < 2^7$. Thus, there are at most 14 non-zero summands for each fixed t and $2^j/t \sim 1$ holds for each of them. From the assumption (2.5) transformed into (2.6) and the rapid decay of κ it follows that

$$\begin{aligned}\big|K^{(t)}(a,b)\big| &\lesssim_\lambda \int_{\mathbb{R}^2} (1+|a-x|)^{-\lambda/2}(1+|b-y|)^{-\lambda/2}(1+|a-x-b+y|)^{-\lambda}|\kappa(x,y)|\,dxdy \\ &\lesssim_\lambda (1+|a|)^{-\lambda/2}(1+|b|)^{-\lambda/2}(1+|a-b|)^{-\lambda}.\end{aligned}$$

Taking absolute values in (5.4) and denoting

$$I(x,y,a,t) := \int_{\mathbb{R}} F(x+s,y)G(x,y+s)\vartheta_t(s-ta)ds,$$

we can now bound (5.4) by

$$\int_{\mathbb{R}^2} (1+|a|)^{-\lambda/2}(1+|b|)^{-\lambda/2}(1+|a-b|)^{-\lambda} \int_0^\infty \int_{\mathbb{R}^2} |I(x,y,a,t)I(x,y,b,t)|\,dxdy\,\frac{dt}{t}\,dadb.$$

Next, we apply the Cauchy-Schwarz inequality in x, y and t, which gives

$$\begin{aligned}\int_{\mathbb{R}^2} (1+|a-b|)^{-\lambda}(1+|a|)^{-\lambda/2}\Big(\int_0^\infty \int_{\mathbb{R}^2} I(x,y,a,t)^2\,dxdy\,\frac{dt}{t}\Big)^{1/2} \\ (1+|b|)^{-\lambda/2}\Big(\int_0^\infty \int_{\mathbb{R}^2} I(x,y,b,t)^2\,dxdy\,\frac{dt}{t}\Big)^{1/2} dadb. \qquad (5.5)\end{aligned}$$

If we denote

$$J(a) := (1+|a|)^{-\lambda/2}\Big(\int_0^\infty \int_{\mathbb{R}^2} I(x,y,a,t)^2\,dxdy\,\frac{dt}{t}\Big)^{1/2},$$

the expression (5.5) can be rewritten as

$$\int_{\mathbb{R}} \Big(\int_{\mathbb{R}} (1+|a-b|)^{-\lambda}J(a)da\Big)J(b)db.$$

Applying the Cauchy-Schwarz inequality in b we obtain

$$\Big(\int_{\mathbb{R}} \Big(\int_{\mathbb{R}} (1+|a-b|)^{-\lambda}J(a)da\Big)^2 db\Big)^{1/2}\Big(\int_{\mathbb{R}} J(b)^2 db\Big)^{1/2}. \qquad (5.6)$$

Note that the integral in a is the convolution of J with $s \mapsto (1+|s|)^{-\lambda}$. By Young's convolution inequality from $\mathrm{L}^1(\mathbb{R}) \times \mathrm{L}^2(\mathbb{R})$ to $\mathrm{L}^2(\mathbb{R})$, the expression (5.6) is bounded by a constant multiple of

$$\|J\|_{\mathrm{L}^2(\mathbb{R})}^2 \leq \int_{\mathbb{R}} (1+a^2)^{-\lambda/2} \int_0^\infty \int_{\mathbb{R}^2} I(x,y,a,t)^2\,dxdy\,\frac{dt}{t}da.$$

Expanding I, this equals

$$\begin{aligned}\int_{\mathbb{R}} (1+a^2)^{-\lambda/2} \int_0^\infty \int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v) \\ \vartheta_t(u-ta)\vartheta_t(v-ta)\,dxdydudv\frac{dt}{t}da. \qquad (5.7)\end{aligned}$$

Observe that this form is associated with the multiplier symbol

$$M(\xi,\eta) := \int_0^\infty \widehat{\vartheta}(t\xi)\widehat{\vartheta}(t\eta)\widehat{\rho}(t(\xi+\eta))\frac{dt}{t}, \tag{5.8}$$

where we have denoted

$$\rho(s) := (1+s^2)^{-\lambda/2}. \tag{5.9}$$

Note that the function $\widehat{\rho}$ is even and hence $M(\xi,\eta) = M(-\eta,-\xi)$. Moreover, M is constant on any line through the origin and in particular $M(\xi,-\xi) = M(1,-1)$ for any $0 \neq \xi \in \mathbb{R}$. Now we write

$$M(\xi,\eta) = M(1,-1) + \big(M(\xi,\eta) - M(1,-1)\big)$$

and split the form (5.7) into the two corresponding parts. The part associated with the constant multiplier yields $M(1,-1)$ times

$$\int_{\mathbb{R}^4} F(x+u,y)G(x,y+u)F(x+v,y)G(x,y+v)\delta_{(0,0)}(u,v)\,dxdydudv$$
$$= \int_{\mathbb{R}^4} F(x,y)^2G(x,y)^2\,dxdy \leq \|F\|^2_{\mathrm{L}^4(\mathbb{R}^2)}\|G\|^2_{\mathrm{L}^4(\mathbb{R}^2)} = 1,$$

where $\delta_{(0,0)}$ denotes the Dirac measure concentrated at the origin. Thus, our remaining task is to estimate the form associated with the symbol $M_0 := M - M(1,-1)$.

For each $(\xi,\eta) \in \mathbb{R}^2 \setminus \{(\xi,\eta) : \xi = \eta \text{ or } \xi = -\eta\}$ we decompose

$$M_0(\xi,\eta) = \sum_{k\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} M_0(\xi,\eta)\widehat{\theta}(2^{j+k}(\xi-\eta))\widehat{\theta}(2^j(\xi+\eta)). \tag{5.10}$$

If we denote

$$m^{(k)}(\xi,\eta) := M_0(\xi,\eta)\widehat{\theta}(2^k(\xi-\eta))\widehat{\theta}(\xi+\eta)$$

and

$$m(\xi,\eta) := \sum_{k\geq 0} m^{(k)}(\xi,\eta) = M_0(\xi,\eta)\Big(\sum_{k\geq 0}\widehat{\theta}(2^k(\xi-\eta))\Big)\widehat{\theta}(\xi+\eta),$$

and split the summation in (5.10) over the regions $k \geq 0$ and $k < 0$, we obtain

$$M_0(\xi,\eta) = \sum_{k\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} m^{(k)}(2^j(\xi,\eta)) = \sum_{j\in\mathbb{Z}} m(2^j(\xi,\eta)) + \sum_{k<0}\sum_{j\in\mathbb{Z}} m^{(k)}(2^j(\xi,\eta)).$$

Here we used that $M_0(\xi,\eta) = M_0(2^j(\xi,\eta))$ by homogeneity.

First we treat the form associated with the multiplier symbol

$$\sum_{j\in\mathbb{Z}} m(2^j(\xi,\eta)). \tag{5.11}$$

Observe that m is compactly supported in the strip $2^{-2} \leq |\xi+\eta| \leq 1$. Moreover, we have

$$|\check{m}(u,v)| \lesssim_\lambda (1+|u+v|)^{-\lambda}(1+|u-v|)^{-2}. \tag{5.12}$$

Indeed, this estimate can be seen by bounding the inverse Fourier transform of

$$(\xi,\eta) \mapsto m(\xi,\eta) + M(1,-1)\phi(\xi,\eta), \tag{5.13}$$

where we have set

$$\phi(\xi,\eta) := \Big(\sum_{k\geq 0}\widehat{\theta}(2^k(\xi-\eta))\Big)\widehat{\theta}(\xi+\eta).$$

Therefore, the inverse Fourier transform of (5.13) is nothing but

$$(u,v)\mapsto \int_{\mathbb{R}}\int_0^\infty \rho(a)\vartheta_t(u-ta)\vartheta_t(v-ta)\frac{dt}{t}da$$

convolved with the Schwartz function $\check{\phi}$ and by the support localization of ϕ we may assume that t ranges over a fixed bounded subinterval of $(0,\infty)$. It remains to observe that

$$\begin{aligned}\Big|\int_{\mathbb{R}}\rho(a)\vartheta_t(u-ta)\vartheta_t(v-ta)da\Big| &\lesssim_{\vartheta,\lambda} (1+|u-v|)^{-2}\int_{\mathbb{R}}\rho(a)(1+|u+v-2a|)^{-2\lambda}da\\ &\lesssim_\lambda (1+|u-v|)^{-2}(1+|u+v|)^{-\lambda}.\end{aligned}$$

which in turn implies (5.12). Boundedness of the form associated with (5.11) now follows from Proposition 6 applied with $\Phi=\check{m}$ and by letting $N\to\infty$.

It remains to consider the form associated with the symbol

$$\sum_{k<0}\sum_{j\in\mathbb{Z}} m^{(k)}(2^j(\xi,\eta)). \tag{5.14}$$

Note that $m^{(k)}$ is supported in the strip $2^{-2}\leq|\xi+\eta|\leq 1$ for each k. To estimate the form associated with (5.14) it now suffices to show that for each $k<0$ we have

$$|\widecheck{m^{(k)}}(u,v)| \lesssim_\lambda 2^{k(\lambda-1)}\,(1+|u+v|)^{-2}\,2^{-k}(1+2^{-k}|u-v|)^{-2}, \tag{5.15}$$

with the implicit constant independent of k. Once we have that, boundedness of the form associated with the symbol in (5.14) for a fixed k follows from Proposition 6 applied with $\Phi=\widecheck{m^{(k)}}$ and by letting $N\to\infty$. In the end it remains to sum the geometric series: $\sum_{k<0}2^{k(\lambda-1)}\lesssim_\lambda 1$.

The estimate (5.15) will be deduced by integration by parts in the Fourier expansion of $m^{(k)}$ once we verify the necessary symbol estimates. At this point we switch to the frequency coordinates $\xi-\eta$ and $\xi+\eta$, which are better suited for our problem. First, we claim that for any $0\leq n\leq 2$, $|\alpha|\sim 1$, and $0<|\beta|\leq 1$ we have

$$\big|\partial_\beta\partial_\alpha^n\big(M_0(\alpha+\beta,\beta-\alpha)\big)\big|\lesssim_\lambda|\beta|^{\lambda-2},\quad \big|\partial_\beta^2\partial_\alpha^n\big(M_0(\alpha+\beta,\beta-\alpha)\big)\big|\lesssim_\lambda|\beta|^{\lambda-3}. \tag{5.16}$$

For now let us assume that the estimates in (5.16) hold. For $0\leq n\leq 2$ define

$$\mu^{(n)}(\alpha,\beta) := \partial_\alpha^n\big(M_0(\alpha+\beta,\beta-\alpha)\big)$$

and note that $\mu^{(n)}(\alpha,0)=0$. Therefore, for any $|\alpha|\sim 1$, $0<|\beta|\leq 1$, and $0\leq n\leq 2$, the first estimate in (5.16) implies

$$\big|\partial_\alpha^n\big(M_0(\alpha+\beta,\beta-\alpha)\big)\big| = \Big|\int_0^\beta\partial_2\mu^{(n)}(\alpha,\gamma)d\gamma\Big|\lesssim_\lambda|\beta|^{\lambda-1}. \tag{5.17}$$

The estimates (5.16) and (5.17) together imply that for any $0 \leq l, n \leq 2$ one has

$$\big|\partial_\beta^l \partial_\alpha^n \big(M_0(\alpha + 2^k\beta, 2^k\beta - \alpha)\widehat{\theta}(2\alpha)\widehat{\theta}(2\beta)\big)\big| \lesssim_\lambda 2^{k(\lambda-1)},$$

which is by the homogeneity of M_0 equivalent to

$$\big|\partial_\beta^l \partial_\alpha^n \big(m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha)\big)\big| \lesssim_\lambda 2^{k(\lambda-1)}. \tag{5.18}$$

We proceed by verifying (5.15). Let us write

$$u\xi + v\eta = 2^{-k}(u-v)2^{k-1}(\xi - \eta) + (u+v)2^{-1}(\xi + \eta).$$

Changing variables $(\alpha, \beta) = (2^{k-1}(\xi - \eta), 2^{-1}(\xi + \eta))$ gives

$$\begin{aligned}\widetilde{m^{(k)}}(u,v) &= \int_{\mathbb{R}^2} m^{(k)}(\xi,\eta) e^{2\pi i(u\xi + v\eta)} d\xi d\eta \\ &= 2^{-k+1} \int_{\mathbb{R}^2} m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha) e^{2\pi i(2^{-k}(u-v)\alpha + (u+v)\beta)} d\alpha d\beta.\end{aligned}$$

If $|u - v| \leq 2^k$ and $|u + v| \leq 1$, then we bound

$$|\widetilde{m^{(k)}}(u,v)| \lesssim 2^{-k}\|m^{(k)}\|_{\mathrm{L}^\infty(\mathbb{R}^2)} \lesssim_\lambda 2^{k(\lambda-1)}\, 2^{-k},$$

which implies (5.15) in this case. Here we used (5.18) to control the L^∞ norm and observed

$$\big\{(\alpha,\beta) : m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha) \neq 0\big\} \subseteq ([-2^{-1}, -2^{-3}] \cup [2^{-3}, 2^{-1}])^2. \tag{5.19}$$

Now assume that $|u-v| \geq 2^k$ and $|u+v| \geq 1$. Integrating by parts we bound $|\widetilde{m^{(k)}}(u,v)|$ by a constant multiple of

$$2^{-k}(2^{-k}|u-v|)^{-2}|u+v|^{-2}\Big|\int_{\mathbb{R}^2} \partial_\beta^2\partial_\alpha^2(m^{(k)}(2^{-k}\alpha+\beta, \beta-2^{-k}\alpha))e^{2\pi i(2^{-k}(u-v)\alpha+(u+v)\beta)} d\alpha d\beta\Big|.$$

Together with (5.18) and (5.19) this shows (5.15) in the present case. If $|u - v| \geq 2^k$ and $|u + v| \leq 1$, or vice versa, we simply combine the arguments from both of the discussed cases.

It remains to show (5.16) and for that we need

$$|\widehat{\rho}'(\xi)| \lesssim_\lambda |\xi|^{\lambda-2}, \quad |\widehat{\rho}''(\xi)| \lesssim_\lambda |\xi|^{\lambda-3} \tag{5.20}$$

for $|\xi| \leq 1$, where ρ is our very particular choice of function (5.9). The following formulae that hold for $\xi > 0$ can be found using [40] or [1]:

$$\begin{aligned}\widehat{\rho}(\xi) &= 2\pi^{\lambda/2}\xi^{(\lambda-1)/2}\mathcal{K}_{(1-\lambda)/2}(2\pi\xi)/\Gamma(\lambda/2), \\ \widehat{\rho}'(\xi) &= -4\pi^{1+\lambda/2}\xi^{(\lambda-1)/2}\mathcal{K}_{(\lambda-3)/2}(2\pi\xi)/\Gamma(\lambda/2), \\ \widehat{\rho}''(\xi) &= 4\pi^{1+\lambda/2}\xi^{(\lambda-3)/2}\big(2\pi\xi\mathcal{K}_{(\lambda-5)/2}(2\pi\xi) - \mathcal{K}_{(\lambda-3)/2}(2\pi\xi)\big)/\Gamma(\lambda/2),\end{aligned}$$

where $\mathcal{K}_\alpha$ is the modified Bessel function of the second kind, given for $\alpha \notin \mathbb{Z}$ and $z > 0$ by the series

$$\mathcal{K}_\alpha(z) = \frac{\pi}{2\sin(\alpha\pi)}\Big(\sum_{n=0}^{\infty}\frac{1}{n!\Gamma(n-\alpha+1)}\Big(\frac{z}{2}\Big)^{2n-\alpha} - \sum_{n=0}^{\infty}\frac{1}{n!\Gamma(n+\alpha+1)}\Big(\frac{z}{2}\Big)^{2n+\alpha}\Big).$$

From this expansion we read off the asymptotic behaviors in a neighborhood of 0:

$$|\mathcal{K}_\alpha(z)| \sim_\alpha z^{\min\{\alpha,-\alpha\}}, \quad |\widehat{\rho}'(\xi)| \sim_\lambda |\xi|^{\lambda-2}, \quad |\widehat{\rho}''(\xi)| \sim_\lambda |\xi|^{\lambda-3},$$

which establish (5.20). Alternatively, to obtain these estimates one could decompose $\widehat{\rho}$ into the Littlewood-Paley pieces and argue by scaling. Finally, differentiation of $M(\alpha+\beta, \beta-\alpha)$ using (5.20) and the product rule gives (5.16).

6 Ergodic averages, deriving Theorem 1 from Theorem 2

Take $m \in \mathbb{N}$ and arbitrary positive integers $n_0 < n_1 < \cdots < n_m$. For $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ denote $A_t(F,G) := A_t^{\mathbb{1}_{[0,1)}}(F,G)$, so that

$$\begin{aligned} A_t(F,G)(x,y) &= \frac{1}{t}\int_{[0,t)} F(x+s,y)\,G(x,y+s)\,ds \\ &= \frac{1}{t}\int_{[x+y,x+y+t)} F(u-y,y)\,G(x,u-x)\,du. \end{aligned} \tag{6.1}$$

Applying Theorem 2 to the scales $t_j = n_j$ and arbitrary functions $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ normalized as in (2.1) gives

$$\sum_{j=1}^{m} \|A_{n_j}(F,G) - A_{n_{j-1}}(F,G)\|_{\mathrm{L}^2(\mathbb{R}^2)}^2 \lesssim 1. \tag{6.2}$$

Now we transfer the obtained estimate from $\mathbb{R}^2$ to $\mathbb{Z}^2$. Recall the definition (1.6) of the averages $\widetilde{A}_n$ and observe that they can be rewritten as

$$\widetilde{A}_n(\widetilde{F},\widetilde{G})(k,l) = \frac{1}{n}\sum_{\substack{i\in\mathbb{Z}\\ k+l\le i\le k+l+n-1}} \widetilde{F}(i-l,l)\,\widetilde{G}(k,i-k). \tag{6.3}$$

Pick arbitrary $\widetilde{F}, \widetilde{G} \in \ell^4(\mathbb{Z}^2)$ normalized by $\|\widetilde{F}\|_{\ell^4(\mathbb{Z}^2)} = \|\widetilde{G}\|_{\ell^4(\mathbb{Z}^2)} = 1$. Define the functions $F, G\colon \mathbb{R}^2 \to \mathbb{R}$ as

$$\begin{aligned} F(x,y) &:= \sum_{i,l\in\mathbb{Z}} \widetilde{F}(i-l,l)\,\mathbb{1}_{[i,i+1)}(x+y)\,\mathbb{1}_{[l,l+1)}(y), \\ G(x,y) &:= \sum_{i,k\in\mathbb{Z}} \widetilde{G}(k,i-k)\,\mathbb{1}_{[k,k+1)}(x)\,\mathbb{1}_{[i,i+1)}(x+y). \end{aligned}$$

Note that F and G are constant on certain skew parallelograms of area 1 and $\|F\|_{\mathrm{L}^4(\mathbb{R}^2)} = \|G\|_{\mathrm{L}^4(\mathbb{R}^2)} = 1$ as well. Splitting the integral (6.1) into the pieces over $i \le u < i+1$ we get

$$A_n(F,G)(k+\alpha, l+\beta) = \frac{1}{n}\sum_{i\in\mathbb{Z}} a_i\,\widetilde{F}(i-l,l)\,\widetilde{G}(k,i-k), \tag{6.4}$$

for any $k, l \in \mathbb{Z}$, $\alpha, \beta \in [0,1)$, where we have denoted

$$a_i = \big|[i, i+1) \cap [k+l+\alpha+\beta, k+l+\alpha+\beta+n)\big|.$$

Observe that

$$\begin{aligned} a_i &= 1 \quad \text{when } k+l+2 \leq i \leq k+l+n-1, \\ a_i &= 0 \quad \text{when } i \leq k+l-1 \text{ or } i \geq k+l+n+2, \\ a_i &\in [0,1] \quad \text{otherwise.} \end{aligned}$$

Comparing (6.4) with (6.3) it immediately follows that

$$\big|A_n(F,G)(k+\alpha, l+\beta) - \widetilde{A}_n(\widetilde{F},\widetilde{G})(k,l)\big| \leq \frac{1}{n} \sum_{i\in\{0,1,n,n+1\}} \big|\widetilde{F}(k+i,l)\,\widetilde{G}(k,l+i)\big|,$$

so for any $n \in \mathbb{N}$ we get

$$\big\|A_n(F,G)(k+\alpha, l+\beta) - \widetilde{A}_n(\widetilde{F},\widetilde{G})(k,l)\big\|_{\ell^2_{(k,l)}(\mathbb{Z}^2)} \leq \frac{4}{n}.$$

Observe that this estimate is uniform in $\alpha, \beta \in [0,1)$. Consequently,

$$\begin{aligned} \Big| \|A_{n_j}(F,G)(k+\alpha,l+\beta) - A_{n_{j-1}}(F,G)(k+\alpha,l+\beta)\|_{\ell^2_{(k,l)}(\mathbb{Z}^2)} \\ - \|\widetilde{A}_{n_j}(\widetilde{F},\widetilde{G}) - \widetilde{A}_{n_{j-1}}(\widetilde{F},\widetilde{G})\|_{\ell^2(\mathbb{Z}^2)}\Big| \leq \frac{8}{n_{j-1}}, \end{aligned}$$

so, taking the $\mathrm{L}^2([0,1)^2)$ norm in (α,β),

$$\Big| \|A_{n_j}(F,G) - A_{n_{j-1}}(F,G)\|_{\mathrm{L}^2(\mathbb{R}^2)} - \|\widetilde{A}_{n_j}(\widetilde{F},\widetilde{G}) - \widetilde{A}_{n_{j-1}}(\widetilde{F},\widetilde{G})\|_{\ell^2(\mathbb{Z}^2)}\Big| \leq \frac{8}{n_{j-1}}.$$

Combining this with (6.2) and using $\sum_{j=1}^{m} n_{j-1}^{-2} \leq \sum_{n=1}^{\infty} n^{-2} \lesssim 1$ we conclude

$$\sum_{j=1}^{m} \big\|\widetilde{A}_{n_j}(\widetilde{F},\widetilde{G}) - \widetilde{A}_{n_{j-1}}(\widetilde{F},\widetilde{G})\big\|_{\ell^2(\mathbb{Z}^2)}^2 \lesssim 1.$$

If we multiply the right hand side by $\|\widetilde{F}\|_{\ell^4(\mathbb{Z}^2)}^2 \|\widetilde{G}\|_{\ell^4(\mathbb{Z}^2)}^2$, then by homogeneity the inequality remains to hold for arbitrary $\widetilde{F}, \widetilde{G}$ and this establishes Corollary 3.

Finally, we transfer to the measure-preserving system $(X, \mathcal{F}, \mu, S, T)$. Let $f, g \in \mathrm{L}^4(X)$ be normalized by $\|f\|_{\mathrm{L}^4(X)} = \|g\|_{\mathrm{L}^4(X)} = 1$. Take a point $x \in X$ and fix a positive integer $N \geq n_m$. The function $\widetilde{F}_{x,N} \colon \mathbb{Z}^2 \to \mathbb{R}$ defined by

$$\widetilde{F}_{x,N}(k,l) := \begin{cases} f(S^k T^l x) & \text{if } 0 \leq k, l \leq 2N-1, \\ 0 & \text{otherwise} \end{cases}$$

and analogously defined $\widetilde{G}_{x,N}$ keep track of the values of f and g along the forward trajectory of x. Observe that for integers $0 \leq k, l < N$ and $0 < n \leq N$ we have

$$M_n(f,g)(S^k T^l x) = \frac{1}{n} \sum_{i=0}^{n-1} f(S^{k+i} T^l x) g(S^k T^{l+i} x) = \widetilde{A}_n\big(\widetilde{F}_{x,N}, \widetilde{G}_{x,N}\big)(k,l),$$

where we used $ST = TS$ and the definition (1.6). The fact that S and T are measure-preserving enables us to write

$$\|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|^2_{\mathrm{L}^2(X)} = \int_X \big|M_{n_j}(f,g)(x) - M_{n_{j-1}}(f,g)(x)\big|^2 d\mu(x)$$
$$= \frac{1}{N^2}\int_X \sum_{k,l=0}^{N-1} \big|M_{n_j}(f,g)(S^kT^lx) - M_{n_{j-1}}(f,g)(S^kT^lx)\big|^2 d\mu(x)$$
$$\leq \frac{1}{N^2}\int_X \big\|\widetilde{A}_{n_j}(\widetilde{F}_{x,N},\widetilde{G}_{x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{x,N},\widetilde{G}_{x,N})\big\|^2_{\ell^2(\mathbb{Z}^2)} d\mu(x)$$

for each $1 \leq j \leq m$. Similar computation as above gives

$$1 = \|f\|^4_{\mathrm{L}^4(X)} = \frac{1}{4N^2}\int_X \sum_{k,l=0}^{2N-1} |f(S^kT^lx)|^4 d\mu(x) = \frac{1}{4N^2}\int_X \|\widetilde{F}_{x,N}\|^4_{\ell^4(\mathbb{Z}^2)} d\mu(x).$$

Taking $\widetilde{F} = \widetilde{F}_{x,N}$, $\widetilde{G} = \widetilde{G}_{x,N}$ in Corollary 3 gives

$$\sum_{j=1}^{m} \big\|\widetilde{A}_{n_j}(\widetilde{F}_{x,N},\widetilde{G}_{x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{x,N},\widetilde{G}_{x,N})\big\|^2_{\ell^2(\mathbb{Z}^2)} \lesssim \|\widetilde{F}_{x,N}\|^4_{\ell^4(\mathbb{Z}^2)} + \|\widetilde{G}_{x,N}\|^4_{\ell^4(\mathbb{Z}^2)}.$$

Integrating this inequality in x over X and dividing by N^2 yields

$$\sum_{j=1}^{m} \|M_{n_j}(f,g) - M_{n_{j-1}}(f,g)\|^2_{\mathrm{L}^2(X)} \lesssim 1$$

for any $n_0 < n_1 < \cdots < n_m$. This completes the proof of Theorem 1.

7 Appendix

The following inequality (7.1) is taken from [21]; we reproduce a proof for the convenience of the reader. An alternative inequality serving the same purpose appears in [31].

Lemma 12. *If $a\colon [2^i, 2^{i+1}] \to \mathbb{R}$ is a continuously differentiable function, then*

$$\sup_{2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}} \sum_{j=1}^{m} |a(t_j) - a(t_{j-1})|^2 \lesssim \|a(t)\|_{\mathrm{L}^2_t((2^i,2^{i+1}),dt/t)} \|ta'(t)\|_{\mathrm{L}^2_t((2^i,2^{i+1}),dt/t)}, \tag{7.1}$$

$$\sup_{2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}} \sum_{j=1}^{m} |a(t_j) - a(t_{j-1})|^2 \leq \|ta'(t)\|^2_{\mathrm{L}^2_t((2^i,2^{i+1}),dt/t)}. \tag{7.2}$$

Proof. To obtain (7.1) we first show that for any $2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}$ and each index $1 \leq j \leq m$ one has

$$|a(t_j) - a(t_{j-1})|^2 \lesssim \|a(t)\|_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)} \|ta'(t)\|_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)}. \tag{7.3}$$

It suffices to prove this under the assumptions that a is non-negative and absolutely continuous. Indeed, in general we then split $a = a_+ - a_-$ where $a_+ = \max(a,0)$ and $a_- = -\min(a,0)$. Note that a_+, a_- and satisfy the required properties and that

$$\|a_+(t)\|_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)} \le \|a(t)\|_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)}, \ \|ta_+'(t)\|_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)} \le \|ta'(t)\|_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)}$$

and analogously for a_-, a_-'. Using the triangle inequality and applying (7.3) to a_+ and a_- we obtain the inequality for any real-valued absolutely continuous function a.

Let us assume that a is as claimed above. Then

$$|a(t_j) - a(t_{j-1})|^2 \le \big|a(t_j)^2 - a(t_{j-1})^2\big| \le \Big|\int_{t_{j-1}}^{t_j} t(a(t)^2)'\frac{dt}{t}\Big| = \Big|\int_{t_{j-1}}^{t_j} 2a(t)ta'(t)\frac{dt}{t}\Big|.$$

Applying the Cauchy-Schwarz inequality in t we bound this up to a constant by

$$\Big(\int_{t_{j-1}}^{t_j} a(t)^2\frac{dt}{t}\Big)^{1/2}\Big(\int_{t_{j-1}}^{t_j} (ta'(t))^2\frac{dt}{t}\Big)^{1/2},$$

which shows (7.3). Summing over j and applying the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}\sum_{j=1}^m |a(t_j) - a(t_{j-1})|^2 &\lesssim \Big(\sum_{j=1}^m \|a(t)\|^2_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)}\Big)^{1/2}\Big(\sum_{j=1}^m \|ta'(t)\|^2_{\mathrm{L}^2_t((t_{j-1},t_j),dt/t)}\Big)^{1/2}\\ &\le \|a(t)\|_{\mathrm{L}^2_t((2^i,2^{i+1}),dt/t)}\, \|ta'(t)\|_{\mathrm{L}^2_t((2^i,2^{i+1}),dt/t)}\end{aligned}$$

for any $2^i \le t_0 < \cdots < t_m \le 2^{i+1}$, which establishes (7.1).

To see (7.2) we estimate

$$|a(t_j) - a(t_{j-1})|^2 = \Big|\int_{t_{j-1}}^{t_j} ta'(t)\frac{dt}{t}\Big|^2 \le (t_j - t_{j-1})\int_{t_{j-1}}^{t_j} (ta'(t))^2\frac{dt}{t^2} \le 2^i\int_{t_{j-1}}^{t_j} (ta'(t))^2\frac{dt}{t^2}.$$

The first inequality follows from the Cauchy-Schwarz inequality in t, while for the second inequality we used the crude bound $t_j - t_{j-1} \le 2^i$. Thus,

$$\sum_{j=1}^m |a(t_j) - a(t_{j-1})|^2 \le 2^i\int_{2^i}^{2^{i+1}} (ta'(t))^2\frac{dt}{t^2} \le \int_{2^i}^{2^{i+1}} (ta'(t))^2\frac{dt}{t},$$

which gives (7.2). □

Acknowledgments

P. D. and C. T. are supported by the Hausdorff Center for Mathematics. V. K. and K. A. Š. are supported in part by the Croatian Science Foundation under the project 3526.

References

[1] M. Abramowitz, I. A. Stegun (Eds.), *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications, Inc., New York, 1992.

[2] T. Austin, *On the norm convergence of non-conventional ergodic averages*, Ergodic Theory Dynam. Systems **30** (2010), no. 2, 321–338.

[3] J. Avigad, J. Rute, *Oscillation and the mean ergodic theorem for uniformly convex Banach spaces*, Ergodic Theory Dynam. Systems **35** (2015), no. 4, 1009–1027.

[4] F. Bernicot, *L^p estimates for non smooth bilinear Littlewood-Paley square functions on $\mathbb{R}$*, Math. Ann. **351** (2011), no. 1, 1–49.

[5] F. Bernicot, S. Shrivastava, *Boundedness of smooth bilinear square functions and applications to some bilinear pseudo-differential operators*, Indiana Univ. Math. J. **60** (2011), no. 1, 233–268.

[6] G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U.S.A. **17** (1931), no. 12, 656–660.

[7] J. Bourgain, *Almost sure convergence and bounded entropy*, Israel J. Math. **63** (1988), no. 1, 79–97.

[8] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161.

[9] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, with an appendix by the author, H. Furstenberg, Y. Katznelson, and D. S. Ornstein, Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45.

[10] J.-P. Conze, E. Lesigne, *Théorèmes ergodiques pour des mesures diagonales*, Bull. Soc. Math. France **112** (1984), no. 2, 143–175.

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

[12] Y. Do, R. Oberlin, E. A. Palsson, *Variation-norm and fluctuation estimates for ergodic bilinear averages* (2015), to appear in Indiana Univ. Math. J., available at arXiv:1504.07134.

[13] S. Donoso, W. Sun, *Pointwise multiple averages for systems with two commuting transformations* (2015), preprint, available at arXiv:1509.09310.

[14] P. Durcik, *An L^4 estimate for a singular entangled quadrilinear form*, Math. Res. Lett. **22** (2015), no. 5, 1317–1332.

[15] P. Durcik, *L^p estimates for a singular entangled quadrilinear form* (2015), to appear in Trans. Amer. Math. Soc., available at arXiv:1506.08150.

[16] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Anal. Math. **31** (1977), 204–256.

[17] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Anal. Math. **38** (1978), no. 1, 275–291.

[18] H. Furstenberg, Y. Katznelson, D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 3, 527–552.

[19] R. L. Jones, R. Kaufman, J. M. Rosenblatt, M. Wierdl, *Oscillation in ergodic theory*, Ergodic Theory Dynam. Systems **18** (1998), no. 4, 889–935.

[20] R. L. Jones, I. V. Ostrovskii, J. M. Rosenblatt, *Square functions in ergodic theory*, Ergodic Theory Dynam. Systems **16** (1996), no. 2, 267–305.

[21] 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.

[22] V. Kovač, *Bellman function technique for multilinear estimates and an application to generalized paraproducts*, Indiana Univ. Math. J. **60** (2011), no. 3, 813–846.

[23] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. **28** (2012), no. 4, 1143–1164.

[24] V. Kovač, *Quantitative norm convergence of double ergodic averages associated with two commuting group actions*, Ergodic Theory Dynam. Systems **36** (2016), no. 3, 860–874.

[25] V. Kovač, K. A. Škreb, *One modification of the martingale transform and its applications to paraproducts and stochastic integrals*, J. Math. Anal. Appl. **426** (2015), no. 2, 1143–1163.

[26] V. Kovač, C. Thiele, *A T(1) theorem for entangled multilinear dyadic Calderón-Zygmund operators*, Illinois J. Math. **57** (2013), no. 3, 775–799.

[27] V. Kovač, C. Thiele, P. Zorin-Kranich, *Dyadic triangular Hilbert transform of two general and one not too general function*, Forum of Mathematics, Sigma **3** (2015), e25.

[28] M. Lacey, *On bilinear Littlewood-Paley square functions* Publ. Mat. **40** (1996), no. 2, 387–396.

[29] M. Lacey, C. Thiele, *L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$*, Ann. of Math. (2) **146** (1997), no. 3, 693–724.

[30] M. Lacey, C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2) **149** (1999), no. 2, 475–496.

[31] M. Mirek, E. Stein, B. Trojan, *$L^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: Variational estimates*, preprint, arXiv:1512.07523.

[32] P. Mohanty, S. Shrivastava, *A note on the bilinear Littlewood-Paley square function*, Proc. Amer. Math. Soc. **138** (2010), no. 6, 2095–2098.

[33] P. K. Ratnakumar, S. Shrivastava, *On bilinear Littlewood-Paley square functions*, Proc. Amer. Math. Soc. **140** (2012), no. 12, 4285–4293.

[34] J. L. Rubio de Francia, *A Littlewood-Paley inequality for arbitrary intervals* Rev. Mat. Iber. **1** (1985), no. 2, 1–14.

[35] T. Tao, *Cancellation for the multilinear Hilbert transform*, Collect. Math. **67** (2016), no. 2, 191–206.

[36] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems **28** (2008), no. 2, 657–688.

[37] C. Thiele, *Wave packet analysis*, CBMS Reg. Conf. Ser. Math. **105**, AMS, Providence, RI, 2006.

[38] J. von Neumann, *Proof of the Quasi-Ergodic Hypothesis*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), no. 1, 70–82.

[39] M. N. Walsh, *Norm convergence of nilpotent ergodic averages*, Ann. of Math. (2) **175** (2012), no. 3, 1667–1688.

[40] Wolfram Research, Inc., *Mathematica*, Ver. 9.0, Champaign, IL, 2012.

[41] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform*, preprint, available at arXiv:1507.02436.

Polona Durcik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: durcik@math.uni-bonn.de

Vjekoslav Kovač, Department of Mathematics, Faculty of Science, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia
E-mail address: vjekovac@math.hr

Kristina Ana Škreb, Faculty of Civil Engineering, University of Zagreb, Fra Andrije Kačića Miošića 26, 10000 Zagreb, Croatia
E-mail address: kskreb@grad.hr

Christoph Thiele, Mathematisches Institut, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: thiele@math.uni-bonn.de

Chapter 5

On side lengths of corners in positive density subsets of the Euclidean space

On side lengths of corners in positive density subsets of the Euclidean space

Polona Durcik, Vjekoslav Kovač, and Luka Rimanić

Abstract

We generalize a result by Cook, Magyar, and Pramanik [3] on three-term arithmetic progressions in subsets of $\mathbb{R}^d$ to corners in subsets of $\mathbb{R}^d \times \mathbb{R}^d$. More precisely, if $1 < p < \infty$, $p \neq 2$, and d is large enough, we show that an arbitrary measurable set $A \subseteq \mathbb{R}^d \times \mathbb{R}^d$ of positive upper Banach density contains corners (x, y), $(x + s, y)$, $(x, y + s)$ such that the ℓ^p-norm of the side s attains all sufficiently large real values. Even though we closely follow the basic steps from [3], the proof diverges at the part relying on harmonic analysis. We need to apply a higher-dimensional variant of a multilinear estimate from [5], which we establish using the techniques from [5] and [6].

1 Introduction

The *upper Banach density* of a set $A \subseteq \mathbb{R}^d$ is defined as

$$\overline{\delta}_d(A) := \limsup_{N\to\infty} \sup_{x\in\mathbb{R}^d} \frac{\big|A \cap (x + [0, N]^d)\big|}{|x + [0, N]^d|},$$

where $|\cdot|$ denotes the d-dimensional Lebesgue measure, so that $|x + [0, N]^d| = N^d$. If $d \geq 2$ and $\overline{\delta}_d(A) > 0$, then there exists a sufficiently large $\lambda_0(A) > 0$ such that for any real number $\lambda \geq \lambda_0(A)$ the set A contains points x and $x + s$ with $\|s\|_{\ell^2} = \lambda$. This fact was shown independently by Bourgain [2], Falconer and Marstrand [7], and Furstenberg, Katznelson, and Weiss [9]. Here $\|\cdot\|_{\ell^2}$ denotes the Euclidean norm. More generally, we denote the *ℓ^p-norm* on $\mathbb{R}^d$ by

$$\|s\|_{\ell^p} := \begin{cases} \Big(\sum_{i=1}^{d} |s_i|^p\Big)^{1/p} & \text{for } 1 \leq p < \infty, \\ \max_{1\leq i\leq d} |s_i| & \text{for } p = \infty, \end{cases}$$

if $s = (s_1, \ldots, s_d)$. It is another observation by Bourgain [2] that the same statement fails if we replace the trivial pattern x, $x + s$ by a 3-*term arithmetic progression*

$$x,\ x + s,\ x + 2s.$$

2010 *Mathematics Subject Classification.* Primary 05D10; Secondary 11B30, 42B20.

Indeed, the set A obtained as a union of the annuli $n - 1/10 \leq \|x\|_{\ell^2}^2 \leq n + 1/10$ as n runs over the positive integers clearly has density $\overline{\delta}_d(A) > 0$, but if $x, s \in \mathbb{R}^d$ are such that x, $x+s$, $x+2s \in A$, then the parallelogram law

$$2\|x+s\|_{\ell^2}^2 + 2\|s\|_{\ell^2}^2 = \|x+2s\|_{\ell^2}^2 + \|x\|_{\ell^2}^2$$

implies that $n - 2/5 \leq 2\|s\|_{\ell^2}^2 \leq n + 2/5$ for some integer n. Therefore, the ℓ^2-norms of the common differences s of the 3-term progressions in A cannot attain values in the set

$$\bigcup_{n=1}^{\infty} \Big(\sqrt{\frac{5n-3}{10}}, \sqrt{\frac{5n-2}{10}}\Big),$$

which contains arbitrarily large numbers.

An interesting phenomenon occurs in large dimensions if one replaces the ℓ^2-norm by other ℓ^p-norms. A recent result by Cook, Magyar, and Pramanik [3] sheds new light on the Euclidean density theorems by establishing that a set of positive upper Banach density still contains 3-term arithmetic progressions such that the ℓ^p-norms of their common differences attain all sufficiently large values when $1 < p < \infty$ and $p \neq 2$.

Theorem 1 (from [3])**.** *For any $p \in (1,2) \cup (2,\infty)$ there exists $d_p \geq 2$ such that for every integer $d \geq d_p$ the following holds. For any measurable set $A \subseteq \mathbb{R}^d$ satisfying $\overline{\delta}_d(A) > 0$ one can find $\lambda_0(A) > 0$ having the property that for any real number $\lambda \geq \lambda_0(A)$, there exist $x, s \in \mathbb{R}^d$ such that x, $x+s$, $x+2s \in A$ and $\|s\|_{\ell^p} = \lambda$.*

The authors of [3] place this result in the context of the Euclidean Ramsey theory and demonstrate that it is sharp with regard to the exponent p. Indeed, measuring the common differences in the ℓ^1 or the ℓ^∞-norm allows for quite straightforward counterexamples. They only leave the optimal value of the dimension threshold d_p as an open problem.

The aim of this paper is a generalization of Theorem 1 to so-called *corners*, which are patterns in $\mathbb{R}^d \times \mathbb{R}^d$ of the form

$$(x,y),\ (x+s,y),\ (x,y+s) \tag{1.1}$$

for some $x, y, s \in \mathbb{R}^d$, $s \neq 0$. The fact that any subset of $\mathbb{Z} \times \mathbb{Z}$ of positive upper density contains a corner was proved by Ajtai and Szemerédi [1], while the first "reasonable" quantitative upper bounds (of the form $n^2/(\log\log n)^c$ with $c > 0$) for subsets of $\{1,\ldots,n\} \times \{1,\ldots,n\}$ without corners are due to Shkredov [17], [18].

We are interested in finding corners exhibiting all sufficiently large side lengths in positive upper Banach density subsets of $\mathbb{R}^d \times \mathbb{R}^d$. Here is the main result of this paper.

Theorem 2. *For any $p \in (1,2) \cup (2,\infty)$ there exists $d_p \geq 2$ such that for every integer $d \geq d_p$ the following holds. For any measurable set $A \subseteq \mathbb{R}^d \times \mathbb{R}^d$ satisfying $\overline{\delta}_{2d}(A) > 0$ one can find $\lambda_0(A) > 0$ with the property that for any real number $\lambda \geq \lambda_0(A)$, there exist $x, y, s \in \mathbb{R}^d$ such that (x,y), $(x+s,y)$, $(x,y+s) \in A$ and $\|s\|_{\ell^p} = \lambda$.*

It is easy to see that Theorem 2 implies Theorem 1. One simply observes that if $A \subseteq \mathbb{R}^d$ has $\overline{\delta}_d(A) > 0$, then the set $\widetilde{A}$ defined by

$$\widetilde{A} := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : y - x \in A\}$$

satisfies $\overline{\delta}_{2d}(\widetilde{A}) > 0$. For this purpose it is convenient to change the coordinates on $\mathbb{R}^d \times \mathbb{R}^d$ to $(x', y') = (x + y, y - x)/\sqrt{2}$ and rotate the cubes $[0, N]^{2d}$ in the definition of $\overline{\delta}_{2d}(\widetilde{A})$, possibly at the cost of losing a multiplicative constant. Moreover, any corner in $\widetilde{A}$ with side s via the projection $(x, y) \mapsto y - x$ gives rise to a 3-term arithmetic progression in A with s as its common difference. The same argument also enables the use of the previously mentioned counterexamples, which rule out the possibility of Theorem 2 holding for $p = 1$, 2, or ∞.

We need to emphasize that our proof of Theorem 2 closely follows the outline of [3]. The most significant novelty appears in the harmonic analysis part of the proof, where we need to prove an estimate for certain "entangled" singular multilinear forms, stated as Theorem 3 below. For previous work on patterns in sufficiently dense subsets of the Euclidean space we refer for instance to [2], [12], and [15]. Bourgain [2] has shown that any set of positive upper Banach density in $\mathbb{R}^k$ contains isometric copies of all sufficiently large dilates of a fixed non-degenerate k-point (i.e. $(k-1)$-dimensional) simplex; non-degeneracy being essential there. Moreover, Lyall and Magyar [15] extended his result to Cartesian products of two non-degenerate simplices. In particular, they are able to detect patterns like (x, y), $(x + s, y)$, $(x, y + t)$ with $\|s\|_{\ell^2} = \|t\|_{\ell^2}$, which have more degrees of freedom than the corners in definition (1.1), and as such are easier to handle.

We now turn to the analytical ingredients that will be needed in the proof of Theorem 2. For $1 \leq p < \infty$ let $\|\cdot\|_{\mathrm{L}^p}$ denote the *Lebesgue* L^p*-norm* defined by

$$\|f\|_{\mathrm{L}^p} := \Big(\int_{\mathbb{R}^d} |f(x)|^p dx \Big)^{1/p}$$

and let $\mathrm{L}^p(\mathbb{R}^d)$ be the corresponding Banach space of a.e.-classes of measurable functions f such that $\|f\|_{\mathrm{L}^p} < \infty$. Denote by $\partial^\kappa f := \partial_1^{\kappa_1} \cdots \partial_d^{\kappa_d} f$ the partial derivative of a function $f \colon \mathbb{R}^d \to \mathbb{C}$ with respect to the multi-index $\kappa = (\kappa_1, \ldots, \kappa_d)$, the order of which will be written $|\kappa| := \kappa_1 + \cdots + \kappa_d$. Finally, we use the notation $\widehat{f}$ and $\check{f}$ for the Fourier transform and its inverse respectively, both initially defined for Schwartz functions f by (2.1) and (2.2) below, and then extended to tempered distributions.

Theorem 3. *Suppose that $m \in \mathrm{C}^\infty(\mathbb{R}^{2d})$ satisfies the standard symbol estimates, i.e. for any multi-index κ there exists a constant $C_\kappa \in [0, \infty)$ such that*

$$\big|(\partial^\kappa m)(\xi, \eta)\big| \leq C_\kappa \|(\xi, \eta)\|_{\ell^2}^{-|\kappa|} \tag{1.2}$$

for all $(\xi, \eta) \in \mathbb{R}^d \times \mathbb{R}^d$, $(\xi, \eta) \neq (0, 0)$. Suppose also that the tempered distribution $K = \check{m}$ is equal to a bounded compactly supported function (denoted by the same letter). Then for any real-valued $F, G \in \mathrm{L}^4(\mathbb{R}^{2d})$ we have the estimate

$$\Big| \int_{(\mathbb{R}^d)^4} F(x+u, y) G(x, y+u) F(x+v, y) G(x, y+v) K(u, v) du dv dx dy \Big| \leq C \|F\|_{\mathrm{L}^4}^2 \|G\|_{\mathrm{L}^4}^2,$$

with a constant $C \in [0, \infty)$ depending only on the dimension d and the constants C_κ.

We will only need a particular case of the theorem when $F = G$, but the given formulation is more natural since the proof will perform different changes of variables in F and G.

The singular integral form in Theorem 3 will appear by expanding out a certain square function quantity; see the proof of Proposition 8 below. It is more singular than the form used in [3] for the same purpose, so we cannot invoke any standard references on modulation-invariant operators. In fact, boundedness of a related singular integral operator, defined as

$$T(F,G)(x,y) := \text{p.v.} \int_{\mathbb{R}} F(x+u,y)G(x,y+u)\frac{du}{u}, \quad (x,y) \in \mathbb{R}^2, \tag{1.3}$$

and called the *triangular Hilbert transform*, is currently an open problem; see [14] for the partial results.

Only recently the techniques required for bounding the form in Theorem 3 were developed as byproducts of the papers [5] and [6], both of which are primarily concerned with unrelated problems. Indeed, Theorem 3 can be viewed as a higher-dimensional variant of an auxiliary estimate from [5], which established a norm-variation bound

$$\sup_{0<t_0<t_1<\cdots<t_m} \sum_{j=1}^{m} \|A_{t_j}(F,G) - A_{t_{j-1}}(F,G)\|_{\mathrm{L}^2}^2 \le C\|F\|_{\mathrm{L}^4}^2 \|G\|_{\mathrm{L}^4}^2 \tag{1.4}$$

for two-dimensional bilinear averages

$$A_t(F,G)(x,y) := \frac{1}{t}\int_0^t F(x+u,y)G(x,y+u)du, \quad (x,y) \in \mathbb{R}^2.$$

Inequality (1.4) in turn proved a quantitative result on the convergence of ergodic averages with respect to two commuting transformations. Moreover, the paper [6] studied multilinear analogs of (1.3), with a more modest goal of proving boundedness with constants growing like $(\log(R/r))^{1-\epsilon}$ as $R/r \to \infty$, where the integration variable u is now restricted to intervals $[-R,-r]$ and $[r,R]$ for $0 < r < R$. Interestingly, early instances of the method used for solving these problems were devised for bounding significantly less singular variants of the operator (1.3), such as

$$T(F,G)(x,y) := \text{p.v.} \int_{\mathbb{R}^2} F(x+u,y)G(x,y+v)K(u,v)dudv, \quad (x,y) \in \mathbb{R}^2;$$

see [4] and [13]. Roughly speaking, the mentioned technique can be described as follows. Instead of decomposing the given operator and bounding its pieces, one rather performs a structural induction and gradually symmetrizes it by repeated applications of the Cauchy-Schwarz inequality and an integration by parts identity. Eventually, the operator in question becomes so symmetric that a monotonicity argument applies, bounding it simply by single-scale objects.

Finally, let us say a few words about the organization of this paper. In Section 2 we give a detailed self-contained proof of Theorem 3. Unlike in [5], where the proof of a special case was given, we do not need any finer control of the constant C here, and are able to make use of further ideas from [6]. Section 3 contains the predominantly combinatorial part of the proof: we derive Theorem 2 from Theorem 3 by mimicking the steps from [3]. Consequently, we frequently refer to [3] and only comment on the ingredients that have to be altered. Finally, in Section 4 we discuss the current obstructions to extending Theorem 1 to longer progressions and Theorem 2 to generalized corners.

2 The analytical part: Proof of Theorem 3

If A and B are two nonnegative quantities, then $A \lesssim_P B$ will denote the inequality $A \leq CB$, with some finite constant C depending on a set of parameters P. We will write $A \sim_P B$ if both $A \lesssim_P B$ and $B \lesssim_P A$ hold. The standard inner product on $\mathbb{R}^d$ will be written $(x, y) \mapsto x \cdot y$, while the Euclidean norm $\| \cdot \|_{\ell^2}$ will simply be denoted by $\| \cdot \|$ in this section. Moreover, let $\mathcal{S}(\mathbb{R}^d)$ be the Schwartz space on $\mathbb{R}^d$ and let $\mathbb{i}$ denote the imaginary unit. We normalize the *Fourier transform* of a d-dimensional Schwartz function f as in

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{-2\pi \mathbb{i} x \cdot \xi} dx, \tag{2.1}$$

so that the *inverse Fourier transform* is given by the formula

$$\check{f}(x) = \int_{\mathbb{R}^d} f(\xi) e^{2\pi \mathbb{i} x \cdot \xi} d\xi. \tag{2.2}$$

Throughout this section we will use the following notation for the standard Gaussian function on $\mathbb{R}^d$ and its partial derivatives:

$$g(x) := e^{-\pi \|x\|^2},$$
$$h^i(x) := \partial_i g(x) \quad \text{for } i = 1, \ldots, d.$$

Moreover, for a function $f \colon \mathbb{R}^d \to \mathbb{C}$ we will denote by f_t its L^1-normalized dilate by $t > 0$, defined as

$$f_t(x) := t^{-d} f(t^{-1} x). \tag{2.3}$$

An important property of the Fourier transform is $\widehat{f_t}(\xi) = \widehat{f}(t\xi)$.

We begin by stating an "integration by parts" lemma, which will be used several times in the proof of Theorem 3. Its one-dimensional variant can be found in [4] or [5], but we prefer to give a self-contained proof. For real-valued functions $\psi, \varphi \in \mathcal{S}(\mathbb{R}^d)$ and $F \in \mathcal{S}(\mathbb{R}^{2d})$ we define the singular integral form

$$\Theta_{\psi,\varphi}(F) := \int_0^\infty \int_{(\mathbb{R}^d)^6} F(x, x') F(x, y') F(y, x') F(y, y')$$
$$\psi_t(x - q) \psi_t(y - q) \varphi_t(x' - p) \varphi_t(y' - p) dx dy dx' dy' dp dq \frac{dt}{t}. \tag{2.4}$$

Note that $\Theta_{\psi,\varphi}(F)$ can be rewritten as

$$\int_0^\infty \int_{(\mathbb{R}^d)^4} \Big(\int_{\mathbb{R}^d} F(x, x') F(x, y') \psi_t(x - q) dx \Big)^2 \varphi_t(x' - p) \varphi_t(y' - p) dx' dy' dp dq \frac{dt}{t}$$
$$= \int_0^\infty \int_{(\mathbb{R}^d)^4} \Big(\int_{\mathbb{R}^d} F(x, x') F(y, x') \varphi_t(x' - p) dx' \Big)^2 \psi_t(x - q) \psi_t(y - q) dx dy dp dq \frac{dt}{t},$$

so that $\Theta_{\psi,\varphi}(F) \geq 0$ when $\varphi \geq 0$ or $\psi \geq 0$.

Lemma 4. *For any real-valued function $F \in \mathcal{S}(\mathbb{R}^{2d})$ and any $\alpha, \beta > 0$ we have the estimate*

$$\sum_{i=1}^{d} \Theta_{h_\alpha^i, g_\beta}(F) \lesssim \|F\|_{\mathrm{L}^4}^4,$$

where the implicit constant is an absolute one, i.e. independent of α, β, d, and F.

Proof of Lemma 4. We claim that

$$\sum_{i=1}^{d} \big(\Theta_{h_\alpha^i, g_\beta}(F) + \Theta_{g_\alpha, h_\beta^i}(F)\big) = \pi \|F\|_{\mathrm{L}^4}^4. \tag{2.5}$$

By the remark preceding the lemma, all terms on the left-hand side of (2.5) are nonnegative. Therefore, (2.5) implies the inequalities

$$\sum_{i=1}^{d} \Theta_{h_\alpha^i, g_\beta}(F) \lesssim \|F\|_{\mathrm{L}^4}^4, \quad \sum_{i=1}^{d} \Theta_{g_\alpha, h_\beta^i}(F) \lesssim \|F\|_{\mathrm{L}^4}^4.$$

This establishes the claim of Lemma 4, up to the verification of (2.5).

To show the identity (2.5) we observe that by the fundamental theorem of calculus

$$\begin{aligned}
&\sum_{i=1}^{d} \Big(\int_0^\infty (2\pi\alpha t\xi_i)^2 e^{-2\pi\|\alpha t\xi\|^2} e^{-2\pi\|\beta t\eta\|^2} \frac{dt}{t} + \int_0^\infty e^{-2\pi\|\alpha t\xi\|^2} (2\pi\beta t\eta_i)^2 e^{-2\pi\|\beta t\eta\|^2} \frac{dt}{t} \Big) \\
&= \pi \int_0^\infty \Big(- t\partial_t \big(e^{-2\pi\|\alpha t\xi\|^2} e^{-2\pi\|\beta t\eta\|^2} \big) \Big) \frac{dt}{t} = \pi
\end{aligned}$$

for any $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$ and $\eta = (\eta_1, \ldots, \eta_d) \in \mathbb{R}^d$ such that $(\xi, \eta) \neq (0,0)$. Using $\widehat{g}(\xi) = e^{-\pi\|\xi\|^2}$ and $\widehat{h^i}(\xi) = 2\pi \mathbb{i} \xi_i \widehat{g}(\xi)$ this can be rewritten as

$$\sum_{i=1}^{d} \Big(\int_0^\infty |\widehat{h_{\alpha t}^i}(\xi)|^2 |\widehat{g_{\beta t}}(\eta)|^2 \frac{dt}{t} + \int_0^\infty |\widehat{g_{\alpha t}}(\xi)|^2 |\widehat{h_{\beta t}^i}(\eta)|^2 \frac{dt}{t} \Big) = \pi. \tag{2.6}$$

Note that for real-valued Schwartz functions φ and ψ one has

$$\begin{aligned}
&\int_{(\mathbb{R}^d)^2} |\widehat{\psi}_t(\xi)|^2 |\widehat{\varphi}_t(\eta)|^2 e^{2\pi \mathbb{i}((x-y)\cdot\xi + (x'-y')\cdot\eta)} d\xi d\eta \\
&= \int_{(\mathbb{R}^d)^2} \psi_t(x-q)\psi_t(y-q)\varphi_t(x'-p)\varphi_t(y'-p) dp dq.
\end{aligned} \tag{2.7}$$

Indeed, for a function ρ we denote $\widetilde{\rho}(s) := \overline{\rho(-s)}$, so that the Fourier transform of $\widetilde{\rho}$ is the complex conjugate of $\widehat{\rho}$. Equality (2.7) follows by noticing that its right-hand side equals

$$(\psi_t * \widetilde{\psi}_t)(x-y)(\varphi_t * \widetilde{\varphi}_t)(x'-y'),$$

which in turn transforms into the left-hand side using the Fourier inversion formula and

$$\widehat{\psi_t * \widetilde{\psi}_t} = |\widehat{\psi}_t|^2, \quad \widehat{\varphi_t * \widetilde{\varphi}_t} = |\widehat{\varphi}_t|^2.$$

Now we multiply (2.6) by

$$F(x,x')F(x,y')F(y,x')F(y,y')e^{2\pi\mathbb{i}((x-y)\cdot\xi+(x'-y')\cdot\eta)}$$

and integrate in x, y, x', y' and ξ, η. Then we apply the inversion formula (2.7) twice, once with $(\psi, \varphi) = (h^i_\alpha, g_\beta)$ and once with $(\psi, \varphi) = (g_\alpha, h^i_\beta)$, and recall the definition (2.4). This gives

$$\sum_{i=1}^{d} \big(\Theta_{h^i_\alpha, g_\beta}(F) + \Theta_{g_\alpha, h^i_\beta}(F)\big)$$
$$= \pi \int_{(\mathbb{R}^d)^4} F(x,x')F(x,y')F(y,x')F(y,y')\delta_{(0,0)}(x-y, x'-y')dxdydx'dy' = \pi\|F\|^4_{\mathrm{L}^4}.$$

Here $\delta_{(0,0)}$ denotes the Dirac measure concentrated at the origin and it is a well-known fact that its Fourier transform is the function constantly equal to 1 on the whole space $\mathbb{R}^d \times \mathbb{R}^d$. □

Observe that for $\nu > 0$ and $x \in \mathbb{R}^d$ we have

$$(1 + \|x\|)^{-\nu} \sim_\nu \int_1^\infty e^{-\pi\beta^{-2}\|x\|^2} \frac{d\beta}{\beta^{\nu+1}}. \tag{2.8}$$

This formula is easily verified by continuity and considering the limiting behavior as $\|x\| \to \infty$, when the ratio of the two sides converges to

$$\lim_{\|x\|\to\infty} \int_1^\infty e^{-\pi(\|x\|/\beta)^2}(\|x\|/\beta)^\nu \frac{d\beta}{\beta} = \int_0^\infty e^{-\pi\alpha^2}\alpha^{\nu-1}d\alpha = \frac{1}{2}\pi^{-\nu/2}\Gamma\Big(\frac{\nu}{2}\Big) \in (0,\infty).$$

It will be used in the proof of Theorem 3 to gradually reduce to forms in which all bump functions are Gaussians or their derivatives. Gaussians possess several convenient algebraic properties, such as positivity, elementary tensor structure, and the fact that they relate differentiation to multiplication.

Proof of Theorem 3. By a density argument we can assume that F and G are real-valued Schwartz functions. Substituting

$$x' = x + y + u, \quad y' = x + y + v$$

and introducing the functions

$$\widetilde{F}(a,b) := F(b-a, a), \quad \widetilde{G}(a,b) := G(a, b-a)$$

the form in question can be written as

$$\int_{(\mathbb{R}^d)^4} \widetilde{F}(y,x')\widetilde{G}(x,x')\widetilde{F}(y,y')\widetilde{G}(x,y')\check{m}(x'-x-y, y'-x-y)dxdydx'dy'. \tag{2.9}$$

We need to bound its absolute value by a constant times

$$\|F\|^2_{\mathrm{L}^4}\|G\|^2_{\mathrm{L}^4} = \|\widetilde{F}\|^2_{\mathrm{L}^4}\|\widetilde{G}\|^2_{\mathrm{L}^4}.$$

Let us henceforth omit writing tildes on the functions in (2.9). We will say that the form (2.9) is associated with the symbol m.

The first step is to decompose the kernel $\check{m}$ into elementary tensors in the variables x, y, x', y', which will allow for an application of the Cauchy-Schwarz inequality.

Let $\phi \in \mathcal{S}(\mathbb{R}^{2d})$ be a nonnegative radial function supported in the annulus $\{\tau \in \mathbb{R}^{2d} : 1 \leq \|\tau\| \leq 2\}$ and not identically equal to 0. The constants in any estimates that follow are allowed to depend on ϕ without explicit mention. Then

$$D := \int_0^\infty \phi(t\xi, t\eta)\|(t\xi, t\eta)\|^2 e^{-\pi\|(t\xi,t\eta)\|^2} \frac{dt}{t}$$

is the same constant for each $(\xi, \eta) \neq (0, 0)$. Therefore, for each such pair (ξ, η) we can write

$$m(\xi, \eta) = D^{-1} \int_0^\infty m(\xi, \eta)\phi(t\xi, t\eta)\|(t\xi, t\eta)\|^2 e^{-\pi\|(t\xi,t\eta)\|^2} \frac{dt}{t}.$$

Using the identity

$$\|(\xi, \eta)\|^2 = \|\xi + \eta\|^2 - 2\xi \cdot \eta \tag{2.10}$$

we can split further

$$m = m^{[1]} + m^{[2]},$$

where

$$m^{[1]}(\xi, \eta) := D^{-1} \int_0^\infty m^{(t)}(t\xi, t\eta)\|t\xi + t\eta\|^2 e^{-\pi\|(t\xi,t\eta)\|^2} \frac{dt}{t},$$
$$m^{[2]}(\xi, \eta) := -2D^{-1} \int_0^\infty m^{(t)}(t\xi, t\eta)(t\xi \cdot t\eta) e^{-\pi\|(t\xi,t\eta)\|^2} \frac{dt}{t},$$

and we have set

$$m^{(t)}(\xi, \eta) := m(t^{-1}\xi, t^{-1}\eta)\phi(\xi, \eta).$$

Now we separately study the forms associated with $m^{[1]}$ and $m^{[2]}$.

First we consider $m^{[1]}$. This is the easier term, as it vanishes on the plane $\xi + \eta = 0$, which brings useful cancellation to our form. The remaining part of the proof related to $m^{[1]}$ can be compared with Sections 3 and 4 in [5].

Define the functions $\varphi^{(t)}$ and $\vartheta^{(i,t)}$ via their Fourier transforms as

$$\widehat{\varphi^{(t)}}(\xi, \eta) := m^{(t)}(\xi, \eta) e^{2\pi\xi\cdot\eta},$$
$$\widehat{\vartheta^{(i,t)}}(\xi, \eta) := \widehat{\varphi^{(t)}}(\xi, \eta)\big((\xi_i + \eta_i)e^{-2^{-1}\pi\|\xi+\eta\|^2}\big)^2.$$

Observe that by $\|\xi + \eta\|^2 = \sum_{i=1}^d (\xi_i + \eta_i)^2$ and (2.10) used in the exponent we have

$$m^{[1]}(\xi, \eta) = D^{-1} \sum_{i=1}^d \int_0^\infty \widehat{\vartheta^{(i,t)}}(t\xi, t\eta) \frac{dt}{t}. \tag{2.11}$$

By the Fourier inversion formula we can write

$$\begin{aligned}
&\vartheta_t^{(i,t)}(x'-x-y,y'-x-y)\\
&=\int_{(\mathbb{R}^d)^2}\widehat{\varphi^{(t)}}(t\xi,t\eta)\big((t\xi_i+t\eta_i)e^{-2^{-1}\pi\|t\xi+t\eta\|^2}\big)^2e^{2\pi\mathbb{i}((x'-x-y)\cdot\xi+(y'-x-y)\cdot\eta)}d\xi d\eta\\
&=-\frac{1}{2\pi^2}\int_{(\mathbb{R}^d)^2}\widehat{\varphi_t^{(t)}}(\xi,\eta)e^{2\pi\mathbb{i}(x'\cdot\xi+y'\cdot\eta)}\big(\widehat{h^i_{2^{-1/2}t}}(-\xi-\eta)\big)^2e^{2\pi\mathbb{i}(x\cdot(-\xi-\eta)+y\cdot(-\xi-\eta))}d\xi d\eta.
\end{aligned}\tag{2.12}$$

To pass from the second to the third line we have used $\widehat{h^i}(\xi)=2\pi\mathbb{i}\xi_ie^{-\pi\|\xi\|^2}$. Using the definition of the Fourier transform, (2.12) can be, up to a constant, viewed as the integral of the Fourier transform of the $4d$-dimensional function

$$H(a,b,c,d):=\varphi_t^{(t)}(x'+a,y'+b)h^i_{2^{-1/2}t}(x+c)h^i_{2^{-1/2}t}(y+d)$$

over a $2d$-dimensional subspace of $\mathbb{R}^{4d}$ parametrized by

$$\{(\xi,\eta,-\xi-\eta,-\xi-\eta):\xi,\eta\in\mathbb{R}^d\}.\tag{2.13}$$

The integral of the Fourier transform of H over the above mentioned subspace equals the integral of the function H itself over the orthogonal complement of this subspace. This fact can be found for instance in [16], and it is easily verified by performing an orthogonal change of variables, which rotates the two subspaces onto $2d$-dimensional coordinate planes in $\mathbb{R}^{2d}\times\mathbb{R}^{2d}$. The orthogonal complement of (2.13) can be parametrized by

$$\{(-p-q,-p-q,-p,-q):p,q\in\mathbb{R}^d\}.$$

Therefore, (2.12) is a constant multiple of

$$\int_{(\mathbb{R}^d)^2}\varphi_t^{(t)}(x'-p-q,y'-p-q)h^i_{2^{-1/2}t}(x-p)h^i_{2^{-1/2}t}(y-q)dpdq.$$

Combining this with the decomposition of $m^{[1]}$ given in (2.11), we see that the form associated with $m^{[1]}$ can be, up to a constant, recognized as

$$\begin{aligned}\sum_{i=1}^d\int_0^\infty\int_{(\mathbb{R}^d)^6}&F(y,x')G(x,x')F(y,y')G(x,y')h^i_{2^{-1/2}t}(x-p)h^i_{2^{-1/2}t}(y-q)\\
&\varphi_t^{(t)}(x'-p-q,y'-p-q)dxdydx'dy'dpdq\frac{dt}{t}.\end{aligned}\tag{2.14}$$

Note that vanishing of the multiplier on $\xi+\eta=0$ is crucial for the cancellation in x and y on the spatial side.

Now we are ready to proceed with an application of the Cauchy-Schwarz inequality. We separate the functions in (2.14) with respect to the variables x,y and rewrite (2.14) as

$$\begin{aligned}\sum_{i=1}^d\int_0^\infty\int_{(\mathbb{R}^d)^4}&\Big(\int_{\mathbb{R}^d}F(y,x')F(y,y')h^i_{2^{-1/2}t}(y-q)dy\Big)\Big(\int_{\mathbb{R}^d}G(x,x')G(x,y')h^i_{2^{-1/2}t}(x-p)dx\Big)\\
&\varphi_t^{(t)}(x'-p-q,y'-p-q)dx'dy'dpdq\frac{dt}{t}.\end{aligned}$$

Then we apply the Cauchy-Schwarz inequality in x', y', p, q, t and in i, after which it remains to bound

$$\sum_{i=1}^{d} \int_0^\infty \int_{(\mathbb{R}^d)^4} \Big(\int_{\mathbb{R}^d} F(y,x')F(y,y')h^i_{2^{-1/2}t}(y-q)dy \Big)^2 \big|\varphi_t^{(t)}(x'-p,y'-p)\big| dx'dy'dpdq\frac{dt}{t} \tag{2.15}$$

and an analogous term involving the function G, which we omit. Note that we changed the variable p to $p-q$ while simplifying (2.15). The next step is to reduce to Gaussians using the formula (2.8). We have

$$\begin{aligned} |\varphi^{(t)}(u,v)| &\lesssim_{d,(C_\kappa)} (1+\|(u,v)\|)^{-2d-1} \\ &\sim_d \int_1^\infty e^{-\pi\beta^{-2}\|(u,v)\|^2}\frac{d\beta}{\beta^{2d+2}} = \int_1^\infty g_\beta(u)g_\beta(v)\frac{d\beta}{\beta^2}. \end{aligned} \tag{2.16}$$

The first estimate above can be verified integrating by parts in the Fourier expansion of $\varphi^{(t)}$. It holds uniformly in $t>0$, with the implicit constant depending only on d and the constants C_κ appearing in (1.2). The second estimate above is simply (2.8) for $x=(u,v)\in\mathbb{R}^{2d}$ and $\nu = 2d+1$. Substituting (2.16) into (2.15) and expanding out the square dominates (2.15) by a constant multiple of

$$\int_1^\infty \sum_{i=1}^{d} \Theta_{h^i_\alpha, g_\beta}(F)\frac{d\beta}{\beta^2},$$

where $\alpha = 2^{-1/2}$ and we recall the definition (2.4). By Lemma 4, the last display is bounded by a constant multiple of

$$\int_1^\infty \|F\|_{\mathrm{L}^4}^4 \frac{d\beta}{\beta^2} = \|F\|_{\mathrm{L}^4}^4,$$

which concludes the proof of boundedness of the form associated with $m^{[1]}$.

It remains to consider the form associated with the multiplier symbol $m^{[2]}$, which does not vanish on $\xi+\eta=0$. This part of the proof can be compared with Section 5 in [5]. In the one-dimensional case [5], the multiplier was symmetrized to become constant on the axis $\xi+\eta=0$. Then that constant was subtracted from the multiplier and a lacunary decomposition with respect to the critical axis was performed. In the present higher-dimensional setting we also reduce the problem to parts vanishing on the problematic plane $\xi+\eta=0$. However, working with Gaussians allows us to do that by using several related algebraic identities.

Applying the Fourier inversion formula to $m^{(t)}$ and using $\|(\xi,\eta)\|^2 = \|\xi\|^2+\|\eta\|^2$ in the exponent we can write

$$m^{[2]}(\xi,\eta) = -2D^{-1}\int_{(\mathbb{R}^d)^2}\int_0^\infty \widetilde{m^{(t)}}(u,v)(t\xi\cdot t\eta)e^{-\pi\|t\xi\|^2}e^{-2\pi \mathbb{i} u\cdot t\xi}e^{-\pi\|t\eta\|^2}e^{-2\pi \mathbb{i} v\cdot t\eta}\frac{dt}{t}dudv.$$

Using $\xi\cdot\eta = \sum_{i=1}^d \xi_i\eta_i$ and taking the inverse Fourier transform of

$$\xi_i e^{-\pi\|t\xi\|^2-2\pi\mathbb{i}u\cdot t\xi} \quad \text{and} \quad \eta_i e^{-\pi\|t\eta\|^2-2\pi\mathbb{i}v\cdot t\eta},$$

we see that the form associated with $m^{[2]}$ can be, up to a constant, recognized as

$$\sum_{i=1}^{d}\int_{(\mathbb{R}^d)^2}\int_0^\infty \widetilde{m^{(t)}}(u,v)\int_{(\mathbb{R}^d)^4} F(y,x')G(x,x')F(y,y')G(x,y')$$
$$h_t^i(x'-x-y-tu)h_t^i(y'-x-y-tv)dxdydx'dy'\frac{dt}{t}dudv. \qquad (2.17)$$

Now we would like to reduce the parameters u and v to only one parameter, which gives more symmetry. For this we first write (2.17) as

$$\sum_{i=1}^{d}\int_{(\mathbb{R}^d)^2}\int_0^\infty \widetilde{m^{(t)}}(u,v)\int_{(\mathbb{R}^d)^2}\Big(\int_{\mathbb{R}^d} F(y,x')G(x,x')h_t^i(x'-x-y-tu)dx'\Big)$$
$$\Big(\int_{\mathbb{R}^d} F(y,y')G(x,y')h_t^i(y'-x-y-tv)dy'\Big)dxdy\frac{dt}{t}dudv.$$

Then we use

$$\big|\widetilde{m^{(t)}}(u,v)\big| \lesssim_{d,(C_\kappa)} (1+\|u\|)^{-d-1}(1+\|v\|)^{-d-1},$$

which can be deduced analogously to the first estimate in (2.16), and apply the Cauchy-Schwarz inequality in x, y, and t. This yields

$$\sum_{i=1}^{d}\Big(\int_{\mathbb{R}^d}(1+\|u\|)^{-d-1}\Big(\int_0^\infty\int_{(\mathbb{R}^d)^2}\Big(\int_{\mathbb{R}^d} F(y,x')G(x,x')$$
$$h_t^i(x'-x-y-tu)dx'\Big)^2 dxdy\frac{dt}{t}\Big)^{1/2}du\Big)^2.$$

Indeed, note that after application of the Cauchy-Schwarz inequality the integrals in u and v have separated and they are equal. By another application of the Cauchy-Schwarz inequality, this time in u, we obtain

$$\Big(\int_{\mathbb{R}}(1+\|u\|)^{-d-1}du\Big)\sum_{i=1}^{d}\Big(\int_{\mathbb{R}^d}(1+\|u\|)^{-d-1}\int_0^\infty\int_{(\mathbb{R}^d)^2}\Big(\int_{\mathbb{R}^d} F(y,x')G(x,x')$$
$$h_t^i(x'-x-y-tu)dx'\Big)^2 dxdy\frac{dt}{t}du\Big). \qquad (2.18)$$

We evaluate the first integral in u and dominate $(1+\|u\|)^{-d-1}$ in the second integral using (2.8), analogously to the domination in (2.16). Expanding the square in (2.18), it then remains to bound

$$\int_1^\infty\sum_{i=1}^{d}\int_{\mathbb{R}^d} g_\alpha(u)\int_0^\infty\int_{(\mathbb{R}^d)^4} F(y,x')G(x,x')F(y,y')G(x,y')$$
$$h_t^i(x'-x-y-tu)h_t^i(y'-x-y-tu)dxdydx'dy'du\frac{dt}{t}\frac{d\alpha}{\alpha^2}. \qquad (2.19)$$

Note that it suffices to consider the expression in (2.19) for each fixed α and obtain estimates that are uniform in $\alpha \geq 1$. Taking the Fourier transform of

$$\sum_{i=1}^{d}\int_{\mathbb{R}^d} g_\alpha(u)h_t^i(a-tu)h_t^i(b-tu)du$$

in variable (a,b) gives a constant multiple of

$$(t\xi \cdot t\eta)e^{-\pi\|(t\xi,t\eta)\|^2}e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}.$$

Therefore, the form (2.19) for a fixed α is, up to a constant, associated with the symbol

$$\int_0^\infty (t\xi \cdot t\eta)e^{-\pi\|(t\xi,t\eta)\|^2}e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}\frac{dt}{t}. \tag{2.20}$$

Now that we have symmetrized the multiplier in u and v, we go backwards: we again use the identity (2.10) and write twice the expression in (2.20) as

$$\int_0^\infty \|t\xi+t\eta\|^2 e^{-\pi\|(t\xi,t\eta)\|^2}e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}\frac{dt}{t} \tag{2.21}$$

$$-\int_0^\infty \|(t\xi,t\eta)\|^2 e^{-\pi\|(t\xi,t\eta)\|^2}e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}\frac{dt}{t}. \tag{2.22}$$

The term (2.21) is easier to handle, and can be treated similarly as (2.11). Indeed, note that (2.21) can be further rewritten as

$$\alpha^{-2}\sum_{i=1}^d \int_0^\infty e^{-\pi\|(t\xi,t\eta)\|^2}\Big((\alpha t\xi_i+\alpha t\eta_i)e^{-2^{-1}\pi\|\alpha t\xi+\alpha t\eta\|^2}\Big)^2\frac{dt}{t}. \tag{2.23}$$

Performing the steps analogous to (2.11)–(2.14) and observing $\alpha^{-2}\leq 1$, since we only consider $\alpha\geq 1$, it suffices to bound

$$\sum_{i=1}^d \int_0^\infty \int_{(\mathbb{R}^d)^6} F(y,x')G(x,x')F(y,y')G(x,y')h^i_{\alpha t}(x-p)h^i_{\alpha t}(y-q)$$
$$g_{2^{1/2}t}(x'-p-q)g_{2^{1/2}t}(y'-p-q)dxdydx'dy'dpdq\frac{dt}{t} \tag{2.24}$$

uniformly in the parameter α. Separating the functions with respect to the variables x,y and applying the Cauchy-Schwarz inequality analogously to (2.15), we estimate the last display by

$$\Big(\sum_{i=1}^d \Theta_{h^i_\alpha,g_\beta}(F)\Big)^{1/2}\Big(\sum_{i=1}^d \Theta_{h^i_\alpha,g_\beta}(G)\Big)^{1/2} \lesssim \|F\|^2_{\mathrm{L}^4}\|G\|^2_{\mathrm{L}^4},$$

where $\beta=2^{1/2}$ and the last inequality follows from Lemma 4.

It remains to consider the second term (2.22). Here we first use an integration by parts identity to transfer to a multiplier vanishing on the critical plane $\xi+\eta=0$. By the fundamental theorem of calculus we have

$$2\pi\int_0^\infty \|(t\xi,t\eta)\|^2 e^{-\pi\|(t\xi,t\eta)\|^2}e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}\frac{dt}{t} \tag{2.25}$$

$$+\,2\pi\int_0^\infty e^{-\pi\|(t\xi,t\eta)\|^2}\|\alpha t\xi+\alpha t\eta\|^2 e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}\frac{dt}{t} \tag{2.26}$$

$$=\int_0^\infty \Big(-t\partial_t\big(e^{-\pi\|(t\xi,t\eta)\|^2}e^{-\pi\|\alpha t\xi+\alpha t\eta\|^2}\big)\Big)\frac{dt}{t}=1$$

for $(\xi,\eta)\neq(0,0)$. Since (2.22) is up to a constant equal to the term in (2.25), and the form associated with the constant symbol 1 is trivially bounded, it remains to consider the form associated with (2.26). Note that it is analogous to (2.21), up to scaling in α.

Expanding $\|\alpha t\xi+\alpha t\eta\|^2$ as in (2.23) and performing the steps analogous to (2.11)–(2.14) we again arrive at the form (2.24), which is bounded by the preceeding discussion. This finishes the proof. □

3 The combinatorial part: Proof of Theorem 2

As mentioned in the introduction, our strategy of proof closely follows that in [3]. In our presentation we try to find a compromise between elaborating the key steps and avoiding repetition.

For a fixed $1<p<\infty$ the authors of [3] start by defining a measure supported on $S_\lambda=\{s\in\mathbb{R}^d\colon \|s\|_{\ell^p}=\lambda\}$ that detects the correct size (of common differences or sides) in the ℓ^p-norm. More precisely, for each $\lambda>0$ we define σ_λ formally via the oscillatory integral

$$\sigma_\lambda(s):=\lambda^{-d+p}\int_{\mathbb{R}}e^{2\pi \mathrm{i}t(\|s\|_{\ell^p}^p-\lambda^p)}dt,$$

which turns out to be a measure that is mutually absolutely continuous with respect to the surface measure on S_λ. The form

$$\mathcal{N}_\lambda(f):=\int_{(\mathbb{R}^d)^2}\int_{S_\lambda}f(x,y)f(x+s,y)f(x,y+s)d\sigma_\lambda(s)dxdy$$

counts corners with respect to this measure. The main idea is to approximate $\mathcal{N}_\lambda(f)$ by a more convenient and smoother integral, defined using an appropriate Schwartz cutoff function, at which point we will be able to count the number of corners using a result from additive combinatorics.

Let $\psi\colon\mathbb{R}\to[0,1]$ be a Schwartz function such that $\widehat{\psi}$ is nonnegative and compactly supported, $\psi(0)=1$, and $\widehat{\psi}(1)>0$. All constants in any estimates that follow are allowed to depend on ψ and this dependence will be suppressed from the notation.

For $\varepsilon,\lambda>0$ define a function $\omega_\lambda^\varepsilon\colon\mathbb{R}^d\to\mathbb{C}$ that approximates the measure σ_λ by

$$\omega_\lambda^\varepsilon(s):=\lambda^{-d+p}\int_{\mathbb{R}}e^{2\pi \mathrm{i}t(\|s\|_{\ell^p}^p-\lambda^p)}\psi(\varepsilon\lambda^p t)dt=\lambda^{-d}\varepsilon^{-1}\widehat{\psi}\Big(\varepsilon^{-1}\big(1-\|\lambda^{-1}s\|_{\ell^p}^p\big)\Big).$$

It is a nonnegative, bounded, and compactly supported function (by our assumptions on $\widehat{\psi}$). Note that

$$\omega_\lambda^\varepsilon(s)=\lambda^{-d}\omega_1^\varepsilon(\lambda^{-1}s),$$

so the notation is still consistent with (2.3) from the previous section. Moreover, in [3] it is shown that

$$\int_{\mathbb{R}^d}\omega_\lambda^\varepsilon(s)ds=c_1(\varepsilon)\int_{\mathbb{R}^d}\omega_\lambda^1(s)ds, \tag{3.1}$$

where

$$c_1(\varepsilon)\sim_{p,d}1, \tag{3.2}$$

for $0 < \varepsilon < 1/100d$. Define

$$\mathcal{M}_\lambda^\varepsilon(f) := \int_{(\mathbb{R}^d)^3} f(x,y)f(x+s,y)f(x,y+s)\omega_\lambda^\varepsilon(s)dsdxdy.$$

The first goal is to prove that $\mathcal{M}_\lambda^1(f)$ is large provided that the function $0 \le f \le 1$ is dense.

Proposition 5. *For any $1 < p < \infty$, any positive integer d, any $0 < \delta \le 1$, and any λ and N satisfying $0 < \lambda \le N$ the following holds. If $f\colon \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ is a measurable function supported in $[0,N]^d \times [0,N]^d$ and such that $\int_{[0,N]^{2d}} f \ge \delta N^{2d}$, then*

$$\mathcal{M}_\lambda^1(f) \gtrsim_{p,d,\delta} N^{2d}.$$

When proving Proposition 5, we borrow the following idea from [3]. In that paper the authors cut $\mathbb{R}^d$ into boxes that can be thought of as scaled images of $[0,1]^d$. On each of these boxes one then uses Roth's theorem for compact abelian groups [2], the underlying group being the d-dimensional torus $\mathbb{T}^d$. We prove a similar result regarding corners in the unit box $[0,1]^d \times [0,1]^d$, which is equivalent to the same statement on $\mathbb{T}^d \times \mathbb{T}^d$.

Lemma 6. *Let $0 < \delta \le 1$ and let $f\colon \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ be a measurable function supported in $[0,1]^d \times [0,1]^d$ and such that $\int_{[0,1]^{2d}} f \ge \delta$. Then*

$$\int_{([0,1]^d)^3} f(x,y)f(x+s,y)f(x,y+s)dsdxdy \gtrsim_{d,\delta} 1.$$

Even though this lemma could be considered a quantitative variant of the well-known corners theorem [1], we could not find the exact reference to the corners theorem on compact abelian groups in the literature, so we deduce Lemma 6 from its more familiar finitary formulation using the averaging trick of Varnavides [20].

Proof of Lemma 6. Suppose that a positive integer n is large enough so that each subset $S \subseteq \{0,1,\ldots,n-1\}^2$ of cardinality at least $(\delta/8)n^2$ must contain a corner. Such n certainly exists by the result of Ajtai and Szemerédi [1], and by the theorem of Shkredov [18] we even know that it is sufficient to take any $n \ge \exp(\exp(8/\delta)^c)$ for some absolute constant c.

First, we note that the set

$$A := \Big\{(x,y) \in [0,1]^d \times [0,1]^d : f(x,y) \ge \frac{\delta}{2}\Big\}$$

has measure at least $\delta/2$ and that $f \ge (\delta/2)\mathbb{1}_A$, where $\mathbb{1}_A$ denotes the indicator function of A. Therefore, it is enough to show

$$\int_{([0,1]^d)^3} \mathbb{1}_A(x,y)\mathbb{1}_A(x+s,y)\mathbb{1}_A(x,y+s)dsdxdy \gtrsim_{d,\delta} 1. \tag{3.3}$$

Take $\epsilon = \delta/16d$ and observe

$$\fint_{(0,\epsilon/n]^d \times ([0,1-\epsilon]^d)^2} \Big(\frac{1}{n^2}\sum_{i,j=0}^{n-1} \mathbb{1}_A(u+it,v+jt)\Big)dtdudv \ge |A \cap [\epsilon,1-\epsilon]^{2d}| \ge \frac{\delta}{2} - 4d\epsilon = \frac{\delta}{4},$$

where $⨍$ denotes the average value of the function on the given set. Defining

$$T := \Big\{(t,u,v) \in (0,\epsilon/n]^d \times [0,1-\epsilon]^d \times [0,1-\epsilon]^d : \frac{1}{n^2}\sum_{i,j=0}^{n-1} \mathbb{1}_A(u+it, v+jt) \geq \frac{\delta}{8}\Big\},$$

from the previous estimate we get

$$|T| \geq \frac{\delta}{8}\Big(\frac{\epsilon}{n}\Big)^d (1-\epsilon)^{2d} \gtrsim_{d,\delta} 1. \tag{3.4}$$

For each triple $(t,u,v) \in T$ we consider the set

$$B_{t,u,v} := \big\{(i,j) \in \{0,1,\dots,n-1\}^2 : (u+it, v+jt) \in A\big\}.$$

Since $B_{t,u,v}$ contains at least $(\delta/8)n^2$ elements, by the choice of n we conclude that $B_{t,u,v}$ must contain a corner (i,j), $(i+k,j)$, $(i,j+k)$, which can be rewritten as

$$\sum_{\substack{i,j,k\in\{0,1,\dots,n-1\}\\ k\geq 1,\ i+k,j+k\leq n-1}} \mathbb{1}_A(u+it, v+jt)\mathbb{1}_A(u+it+kt, v+jt)\mathbb{1}_A(u+it, v+jt+kt) \geq 1.$$

Integrating this over $(t,u,v) \in T$, using (3.4), and changing variables to

$$x = u+it, \quad y = v+jt, \quad s = kt$$

we obtain

$$\sum_{\substack{i,j,k\in\{0,1,\dots,n-1\}\\ k\geq 1,\ i+k,j+k\leq n-1}} \frac{1}{k^d}\int_{[0,k\epsilon/n]^d\times([0,1]^d)^2} \mathbb{1}_A(x,y)\mathbb{1}_A(x+s,y)\mathbb{1}_A(x,y+s)dsdxdy \gtrsim_{d,\delta} 1. \tag{3.5}$$

It remains to observe that the left-hand side of (3.5) is at most n^3 times the left-hand side of (3.3), recalling that n can be taken to be a function depending only on δ. □

Proof of Proposition 5. It is straightforward to adapt the proof of the analogous proposition from [3] in the language of corners, replacing Roth's theorem on compact abelian groups [2] with Lemma 6. □

Our next aim is to prove that $\mathcal{N}_\lambda$ and $\mathcal{M}_\lambda^\varepsilon$ are in some sense close to each other.

Proposition 7. *For any $p \in (1,2)\cup(2,\infty)$ there exists $\gamma_p > 0$ such that for any positive integer d, any $0<\varepsilon<1$, and any λ and N satisfying $0<\lambda\leq N$ the following holds. If $f\colon \mathbb{R}^d\times\mathbb{R}^d \to [-1,1]$ is a measurable function supported in $[0,N]^d\times[0,N]^d$, then*

$$|\mathcal{N}_\lambda(f) - \mathcal{M}_\lambda^\varepsilon(f)| \lesssim_{p,d} \varepsilon^{d\gamma_p - 1} N^{2d}.$$

The proof of this proposition uses *uniformity norms* or the U^k*-norms*, which Gowers introduced in his work on Szemerédi's theorem on the integers [10],[11]. For a measurable function $f\colon \mathbb{R}^d \to \mathbb{C}$ we define the *Gowers uniformity norm on* $\mathbb{R}^d$ *of degree* k by

$$\|f\|_{\mathrm{U}^k}^{2^k} := \int_{(\mathbb{R}^d)^{k+1}} \Delta_{h_1}\cdots\Delta_{h_k} f(x)dxdh_1\cdots dh_k,$$

where $\Delta_h f(x) := f(x)\overline{f(x+h)}$. A linear change of variables immediately yields

$$\|f_t\|_{\mathrm{U}^k} = t^{-d(1-(k+1)/2^k)}\|f\|_{\mathrm{U}^k} \tag{3.6}$$

for all $t>0$.

Proof of Proposition 7. By a density argument we can assume that f is continuous.

From the discussion preceding the proof we know how $\|\omega_\lambda^\eta - \omega_\lambda^\varepsilon\|_{\mathrm{U}^3}$ is defined for $0 < \eta < \epsilon$. The authors of [3] also give a meaning to $\|\sigma_\lambda - \omega_\lambda^\varepsilon\|_{\mathrm{U}^3}$ by interpreting it as the limit $\lim_{\eta\to 0^+} \|\omega_\lambda^\eta - \omega_\lambda^\varepsilon\|_{\mathrm{U}^3}$, which is justified by the facts that $(\omega_\lambda^\eta)_{\eta>0}$ is a Cauchy net in the U^3-norm and that it converges vaguely to σ_λ as $\eta \to 0^+$. Moreover, in [3] it is shown that for any $1 < p < \infty$, $p \neq 2$ there exists a constant $\gamma_p > 0$ such that for each integer d, any $0 < \varepsilon < 1$, and any $\lambda > 0$ one has

$$\|\sigma_\lambda - \omega_\lambda^\varepsilon\|_{\mathrm{U}^3} \lesssim_{p,d} \lambda^{-d/2}\varepsilon^{d\gamma_p - 1}. \tag{3.7}$$

Indeed, it suffices to work out the case $\lambda = 1$ and the general result follows from the scaling identity (3.6).

On the other hand, by applying the Cauchy-Schwarz inequality three times, for an arbitrary measurable function $g\colon \mathbb{R}^d \to \mathbb{R}$ supported in a constant dilate of the cube $[-\lambda, \lambda]^d$ one obtains

$$\Big|\int_{(\mathbb{R}^d)^3} f(x,y)f(x+s,y)f(x,y+s)g(s)dsdxdy\Big| \lesssim N^{2d}\lambda^{d/2}\|g\|_{\mathrm{U}^3},$$

the so-called generalized von Neumann's theorem, this time for corners. Setting $g = \omega_\lambda^\eta - \omega_\lambda^\varepsilon$ and letting $\eta \to 0^+$ we get

$$|\mathcal{N}_\lambda(f) - \mathcal{M}_\lambda^\varepsilon(f)| \lesssim N^{2d}\lambda^{d/2}\|\sigma_\lambda - \omega_\lambda^\varepsilon\|_{\mathrm{U}^3}. \tag{3.8}$$

It remains to combine (3.7) and (3.8) and the claim follows. □

As the final step, we use Theorem 3 to connect $\mathcal{M}_\lambda^1(f)$ and $\mathcal{M}_\lambda^\varepsilon(f)$, where λ goes through a sequence of scalars. Motivated by [3], we define

$$k_\lambda^\varepsilon(s) := \omega_\lambda^\varepsilon(s) - c_1(\varepsilon)\omega_\lambda^1(s),$$

which is consistent with the notation (2.3), and also set

$$\mathcal{E}_\lambda^\varepsilon(f) := \mathcal{M}_\lambda^\varepsilon(f) - c_1(\varepsilon)\mathcal{M}_\lambda^1(f),$$

where $c_1(\varepsilon)$ is the constant from (3.1). We prove the following result.

Proposition 8. *Let $0 < \varepsilon < 1$, and let d and J be positive integers. Suppose that $\lambda_1 < \lambda_2 < \cdots < \lambda_J$ are positive numbers such that $\lambda_{j+1}/\lambda_j \geq 2$ for each $1 \leq j \leq J-1$. If $f\colon \mathbb{R}^d \times \mathbb{R}^d \to [-1,1]$ is a measurable function supported in $[0,N]^d \times [0,N]^d$, then*

$$\sum_{j=1}^{J} |\mathcal{E}_{\lambda_j}^\varepsilon(f)|^2 \lesssim_{d,\varepsilon} N^{4d}. \tag{3.9}$$

Proof of Proposition 8. Using the definition of $\mathcal{E}_\lambda^\varepsilon(f)$ and applying the Cauchy-Schwarz inequality we estimate:

$$\begin{aligned}
\sum_{j=1}^{J}|\mathcal{E}_{\lambda_j}^\varepsilon(f)|^2 &\leq \sum_{j=1}^{J}\Big(\int_{(\mathbb{R}^d)^2} f(x,y)\Big|\int_{\mathbb{R}^d} f(x+s,y)f(x,y+s)k_{\lambda_j}^\varepsilon(s)ds\Big|dxdy\Big)^2 \\
&\leq \|f\|_{\mathrm{L}^2}^2 \sum_{j=1}^{J}\int_{(\mathbb{R}^d)^2}\Big(\int_{\mathbb{R}^d} f(x+s,y)f(x,y+s)k_{\lambda_j}^\varepsilon(s)ds\Big)^2 dxdy \\
&= \|f\|_{\mathrm{L}^2}^2 \int_{(\mathbb{R}^d)^4} f(x+u,y)f(x,y+u)f(x+v,y)f(x,y+v)K(u,v)dudvdxdy,
\end{aligned}$$

where we have written $K(u,v) := \sum_{j=1}^{J} k_{\lambda_j}^\varepsilon(u)k_{\lambda_j}^\varepsilon(v)$. It was verified in [3] that $m = \widehat{K}$ satisfies the symbol estimates (1.2) with the constants C_κ depending only on κ, d, and ε. Therefore, Theorem 3 can be applied and yields

$$\sum_{j=1}^{J}|\mathcal{E}_{\lambda_j}^\varepsilon(f)|^2 \lesssim_{d,\varepsilon} \|f\|_{\mathrm{L}^2}^2\|f\|_{\mathrm{L}^4}^4 \leq N^{4d}. \qquad \square$$

We now deduce Theorem 2 from Propositions 5, 7, and 8.

Proof of Theorem 2. We argue by contradiction. Recall the constant γ_p from Proposition 7. If Theorem 2 does not hold, then for some $1 < p < \infty$, $p \neq 2$ and some $d > 1/\gamma_p$ there exists a measurable set $A \subseteq \mathbb{R}^{2d}$ with $\overline{\delta}_{2d}(A) > 0$ such that the side lengths of corners in A, measured in the ℓ^p-norm, avoid values from some positive sequence $(\lambda_j)_{j=1}^\infty$ converging to $+\infty$. We can sparsify this sequence if necessary, so that it satisfies $\lambda_{j+1}/\lambda_j \geq 2$ for each index j. Fix any positive integer J. It will be enough to consider finitely many scales $\lambda_1 < \cdots < \lambda_J$.

By the definition of upper Banach density, for any fixed $0 < \delta < \overline{\delta}_{2d}(A)$, there exists a number $N \geq \lambda_J$ for which there is $x_N \in \mathbb{R}^{2d}$ such that $|A \cap (x_N + [0,N]^{2d})| \geq \delta N^{2d}$. If we denote $A_N := (-x_N + A) \cap [0,N]^{2d}$, then A_N is a measurable subset of $[0,N]^{2d}$ with measure at least δN^{2d} such that the side length of any corner inside A_N avoids the values $\lambda_1, \ldots, \lambda_J$. The latter property immediately implies that $\mathcal{N}_{\lambda_j}(\mathbb{1}_{A_N}) = 0$ for each $1 \leq j \leq J$.

Let us apply the three auxiliary propositions with $f = \mathbb{1}_{A_N}$, recalling that this is the indicator function of A_N. Note that $\lim_{\varepsilon\to 0^+} \varepsilon^{d\gamma_p - 1} = 0$ by our choice of d. Therefore, if $\varepsilon > 0$ is taken small enough (depending on p, d, δ), then (3.2) and Propositions 5 and 7 give

$$|\mathcal{E}_{\lambda_j}^\varepsilon(f)| \geq c_1(\varepsilon)\mathcal{M}_{\lambda_j}^1(f) - |\mathcal{N}_{\lambda_j}(f) - \mathcal{M}_{\lambda_j}^\varepsilon(f)| \gtrsim_{p,d,\delta} N^{2d}.$$

Consequently,

$$\sum_{j=1}^{J}|\mathcal{E}_{\lambda_j}^\varepsilon(f)|^2 \gtrsim_{p,d,\delta} JN^{4d}. \tag{3.10}$$

Combining (3.9) and (3.10), and dividing by N^{4d}, we conclude that $J \lesssim_{p,d,\delta,\varepsilon} 1$. Recalling that J could have been taken arbitrarily large we arrive at the contradiction. $\square$

It is worth observing that a variant of the bound (3.9) with a constant $o(J)$ on the right-hand side would have been sufficient. It is plausible that such a bound could be easier to establish than the uniform estimates in Theorem 3 and Proposition 8. However, the scales $\lambda_1 < \cdots < \lambda_J$ in such a bound must comprise an arbitrary lacunary sequence. For instance, obtaining a $o(J)$ estimate for consecutive dyadic scales $\lambda_j = 2^j$ is considerably easier; compare with the closing remarks in the next section.

4 Remarks on possible generalizations

It is natural to ask if the generalization of Theorem 1 holds for k-term arithmetic progressions in $\mathbb{R}^d$,

$$x,\ x+s,\ x+2s,\ \ldots,\ x+(k-1)s,$$

and if Theorem 2 extends to the generalized k-element corners in $(\mathbb{R}^d)^{k-1}$,

$$(x_1,x_2,\ldots,x_{k-1}),\ (x_1+s,x_2,\ldots,x_{k-1}),\ (x_1,x_2+s,\ldots,x_{k-1}),\ \ldots,\ (x_1,x_2,\ldots,x_{k-1}+s).$$

The result that any positive upper density subset of $\mathbb{Z}^{k-1}$ has to contain a nontrivial k-element corner is popularly known as the *multidimensional Szemerédi theorem* and was first shown by Furstenberg and Katznelson [8].

The following proposition is a straightforward generalization of the aforementioned counterexample of Bourgain. It prohibits p from taking any integer value less than k.

Proposition 9. *Let d,k,p be positive integers such that $p \leq k-1$. There exists a measurable set $A \subseteq \mathbb{R}^d$ of positive upper Banach density such that no $\lambda_0 > 0$ satisfies the property that for each $\lambda \geq \lambda_0$ one can find a k-term arithmetic progression x, $x+s$, $\ldots$, $x+(k-1)s$ in A with $\|s\|_{\ell^p} = \lambda$.*

Proof of Proposition 9. Take $x = (x_1,\ldots,x_d)$ and $s = (s_1,\ldots,s_d)$ in $\mathbb{R}^d$ such that $x+js$ has nonnegative coordinates for $j = 0,\ldots,p$ and observe the identity

$$\sum_{j=0}^{p} (-1)^{p-j} \binom{p}{j} \|x+js\|_{\ell^p}^p = p!\|s\|_{\ell^p}^p. \tag{4.1}$$

It is a direct consequence of the scalar identity

$$\sum_{j=0}^{p} (-1)^{p-j} \binom{p}{j} (\alpha + j\beta)^l = \begin{cases} 0 & \text{for } l = 0,1,\ldots,p-1, \\ p!\beta^p & \text{for } l = p, \end{cases}$$

applied with $l = p$, $\alpha = x_i$, $\beta = s_i$, $i = 1,\ldots,d$, which in turn can be easily established by induction on p.

Led by the example for three-term progressions, we define

$$A := \bigcup_{n=1}^{\infty} \big\{ x \in [0,\infty)^d : n - 2^{-p-2} \leq \|x\|_{\ell^p}^p \leq n + 2^{-p-2} \big\}. \tag{4.2}$$

As before, the set A is made up of parts of spherical shells, but this time with respect to the ℓ^p-norm. It is easy to see that it still satisfies $\overline{\delta}_d(A) > 0$.

Suppose that $x, s \in \mathbb{R}^d$ are such that $x + js \in A$ for $j = 0, 1, \ldots, k-1$. We only need to consider the first $p+1$ terms of this progression. By construction, $\|x + js\|_{\ell^p}^p$ differs from some positive integer n_j by at most 2^{-p-2}. From (4.1) we see that $p!\|s\|_{\ell^p}^p$ differs from the integer $\sum_{j=0}^{p}(-1)^{p-j}\binom{p}{j}n_j$ by at most

$$\sum_{j=0}^{p}\binom{p}{j}\big|\|x + js\|_{\ell^p}^p - n_j\big| \leq \sum_{j=0}^{p}\binom{p}{j}2^{-p-2} = \frac{1}{4}.$$

Consequently, $\|s\|_{\ell^p}$ cannot attain values in the set

$$\bigcup_{n=1}^{\infty}\Big(\Big(\frac{4n-3}{4p!}\Big)^{1/p}, \Big(\frac{4n-1}{4p!}\Big)^{1/p}\Big), \tag{4.3}$$

which is unbounded from above. □

Example (4.2) from the previous proof also leads to a counterexample for generalized corners, by considering

$$\widetilde{A} := \big\{(x_1, x_2, \ldots, x_{k-1}) \in (\mathbb{R}^d)^{k-1} : x_1 + 2x_2 + \cdots + (k-1)x_{k-1} \in A\big\}.$$

Once again, this set has $\overline{\delta}_{(k-1)d}(\widetilde{A}) > 0$, but the ℓ^p-norm of the side s of each k-element corner in $\widetilde{A}$ cannot belong to the set (4.3).

There is still a chance that Theorems 1 and 2 generalize to $k \geq 4$ and any $1 \leq p \leq \infty$ other than $1, 2, \ldots, k-1$, and ∞. However, the corresponding analogs of Theorem 3 would involve operators of complexity similar as to the so-called *multilinear and simplex Hilbert transforms* (see [14],[19],[21]), for which no L^p-boundedness results are known at the time of writing. An encouraging sign is that the papers [19] and [21] establish estimates for the truncations of these operators with constants $o(J)$ in the number of consecutive dyadic scales J, while [6] improves this bound to $J^{1-\epsilon}$ for some $\epsilon > 0$. As one needs to consider arbitrary (and not only consecutive dyadic) scales for the intended application, we believe that generalizations to large values of k are still out of reach of the currently available techniques.

Acknowledgments

P. D. is supported by the Hausdorff Center for Mathematics. V. K. is supported in part by the Croatian Science Foundation under the project 3526. L. R. is supported by the School of Mathematics at the University of Bristol. P. D. and V. K. are partially supported by the bilateral DAAD-MZO grant *Multilinear singular integrals and applications.* The authors would like to thank Christoph Thiele, Julia Wolf, and the members of the HARICOT seminar for useful discussions and helpful comments.

References

[1] M. Ajtai, E. Szemerédi, *Sets of lattice points that form no squares*, Stud. Sci. Math. Hungar. **9** (1974), 9–11.

[2] J. Bourgain, *A Szemerédi type theorem for sets of positive density in* $\mathbf{R}^k$, Israel J. Math. **54** (1986), no. 3, 307–316.

[3] B. Cook, Á. Magyar, M. Pramanik, *A Roth type theorem for dense subsets of* $\mathbb{R}^d$ (2015), to appear in Bull. London Math. Soc., available at arXiv:1511.06010.

[4] P. Durcik, *An* L^4 *estimate for a singular entangled quadrilinear form*, Math. Res. Lett. **22** (2015), no. 5, 1317–1332.

[5] P. Durcik, V. Kovač, K. A. Škreb, C. Thiele, *Norm-variation of ergodic averages with respect to two commuting transformations* (2016), preprint, available at arXiv:1603.00631.

[6] P. Durcik, V. Kovač, C. Thiele, *Power-type cancellation for the simplex Hilbert transform* (2016), to appear in J. Anal. Math., available at arXiv:1608.00156.

[7] K. J. Falconer, J. M. Marstrand, *Plane sets with positive density at infinity contain all large distances*, Bull. London Math. Soc. **18** (1986), no. 5, 471–474.

[8] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Anal. Math. **38** (1978), no. 1, 275–291.

[9] H. Furstenberg, Y. Katznelson, B. Weiss, *Ergodic theory and configurations in sets of positive density. Mathematics of Ramsey theory*, pp. 184–198, Algorithms Combin. **5**, Springer, Berlin, 1990.

[10] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), no. 3, 529–551.

[11] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[12] L. Huckaba, N. Lyall, Á. Magyar, *Simplices and sets of positive upper density in* $\mathbb{R}^d$ (2016), to appear in Proc. Amer. Math. Soc., available at arXiv:1509.09283.

[13] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. **28** (2012), no. 4, 1143–1164.

[14] V. Kovač, C. Thiele, P. Zorin-Kranich, *Dyadic triangular Hilbert transform of two general functions and one not too general function*, Forum of Mathematics, Sigma **3** (2015), e25.

[15] N. Lyall, Á. Magyar, *Product of simplices and sets of positive upper density in* $\mathbb{R}^d$ (2016), to appear in Math. Proc. Cambridge Philos. Soc., available at arXiv:1605.04890.

[16] C. Muscalu, W. Schlag, *Classical and multilinear harmonic analysis. Vol. II*, Cambridge Studies in Advanced Mathematics **138**, Cambridge University Press, 2013.

[17] I. D. Shkredov, *On a problem of Gowers*, Izv. Ross. Akad. Nauk Ser. Mat. **70** (2006), no. 2, 179–221; English translation in Izv. Math. **70** (2006), no. 2, 385–425.

[18] I. D. Shkredov, *On a generalization of Szemerédi's theorem*, Proc. London Math. Soc. (3) **93** (2006), no. 3, 723–760.

[19] T. Tao, *Cancellation for the multilinear Hilbert transform*, Collect. Math. **67** (2016), no. 2, 191–206.

[20] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360.

[21] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform* (2015), to appear

in Math. Res. Lett., available at arXiv:1507.02436.

Polona Durcik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: durcik@math.uni-bonn.de

Vjekoslav Kovač, Department of Mathematics, Faculty of Science, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia
E-mail address: vjekovac@math.hr

Luka Rimanić, School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom
E-mail address: luka.rimanic@bristol.ac.uk

Chapter 6

Power-type cancellation for the simplex Hilbert transform

Power-type cancellation for the simplex Hilbert transform

Polona Durcik, Vjekoslav Kovač, and Christoph Thiele

Abstract

We prove L^p bounds for the truncated simplex Hilbert transform which grow with a power less than one of the truncation range in the logarithmic scale.

1 Introduction

The simplex Hilbert transform of degree $n \geq 1$ is given by

$$\Lambda_n := \text{p.v.} \int_{\mathbb{R}^{n+1}} \prod_{i=0}^{n} F_i(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \frac{1}{x_0 + \cdots + x_n} dx_0 \ldots dx_n.$$

It is a multilinear form in the $n+1$ functions $F_0, \ldots, F_n$, which for simplicity we assume to be in the Schwartz class. If $n = 1$, then the simplex Hilbert transform is the form obtained by dualization of the classical Hilbert transform. The case $n = 2$ was called the triangular Hilbert transform in [4]. A major open problem is whether for $n \geq 2$ the simplex Hilbert transform satisfies any L^p bounds of the type

$$|\Lambda_n| \leq C \prod_{i=0}^{n} \|F_i\|_{p_i}.$$

Partial progress in the case $n = 2$ was made in [4] for a dyadic model and under the additional assumption that one of the functions F_i takes certain special forms.

The papers [5] and [6] initiated the study of growth of the bounds for the truncated simplex Hilbert transform

$$\Lambda_{n,r,R} := \int_{r \leq |x_0 + \cdots + x_n| \leq R} \prod_{i=0}^{n} F_i(x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \frac{1}{x_0 + \cdots + x_n} dx_0 \ldots dx_n$$

for some truncation parameters $0 < r < R$. The trivial estimate

$$|\Lambda_{n,r,R}| \leq 2\Big(\log \frac{R}{r}\Big) \prod_{i=0}^{n} \|F_i\|_{p_i} \tag{1.1}$$

with Banach space exponents $1 \leq p_i \leq \infty$ satisfying the Hölder scaling $\sum_{i=0}^{n} 1/p_i = 1$ follows by substituting $x_0 = x - x_1 - \cdots - x_n$, applying Hölder's inequality in $x_1, \ldots, x_n$, and integrating in x. Alternatively, if one is careless about the actual constant 2, one can simply break the kernel into about $\log(R/r)$ many scales and estimate each scale separately.

Using techniques from additive combinatorics, Zorin-Kranich [6] improved this bound to $o(\log(R/r))$ when $R/r \to \infty$ in the open range $1 < p_i < \infty$ with the Hölder scaling. A special case of this result was shown before by Tao [5].

The main result of this paper is the following bound.

Theorem 1. *There exists a finite constant C depending only on n such that for any Schwartz functions $F_0, \ldots, F_n$ on $\mathbb{R}^n$ and any $0 < r < R$ we have*

$$|\Lambda_{n,r,R}| \leq C\Big(\log\frac{R}{r}\Big)^{1-2^{-n+1}} \|F_0\|_{2^n} \prod_{i=1}^{n} \|F_i\|_{2^{n-i+1}}. \tag{1.2}$$

By interpolation of (1.2) with (1.1) we obtain the following corollary.

Corollary 2. *Let $1 < p_0, \ldots, p_n < \infty$ and $1/p_0 + \cdots + 1/p_n = 1$. There exist a finite constant C and a number $\epsilon > 0$, both depending only on n and $p_0, \ldots, p_n$, such that for any Schwartz functions $F_0, \ldots, F_n$ on $\mathbb{R}^n$ and any $0 < r < R$ we have*

$$|\Lambda_{n,r,R}| \leq C\Big(\log\frac{R}{r}\Big)^{1-\epsilon} \prod_{i=0}^{n} \|F_i\|_{p_i}.$$

In particular, this strengthens the results from [5] and [6]. The special case $n = 2$ was commented on in [2], where it followed from boundedness of a certain square function. A modification of our arguments could yield bounds for a simplex transform associated with more general Calderón-Zygmund kernels on $\mathbb{R}$ replacing $K(t) = 1/t$, but we do not aim for that kind of generality here. The reader can also consult [4] and [6] for the ways of encoding various lower-dimensional or less singular operators into Λ_n, so that Corollary 2 gives nontrivial estimates for the truncations of these operators too, even though some of them are already known to be (uniformly) bounded.

The proof of Theorem 1 is a special case of a more general estimate in Lemma 3 on auxiliary forms involving an additional parameter $1 \leq k \leq n$, which is in turn proved by induction on that parameter. The induction uses higher-dimensional analogues of the arguments in [1], [2], and [3], i.e. intertwined applications of the Cauchy-Schwarz inequality (2.8) and an integration by parts identity (2.15). The base case is closely related to the quadrilinear forms studied in [1] and [3].

In Section 4 we discuss a dyadic version of Theorem 1.

2 Proof of Theorem 1

We fix an integer $n \geq 2$ and numbers $0 < r < R$. One can suppose that $\log(R/r) > 1$, since otherwise (1.2) is even weaker than (1.1). We also fix Schwartz functions $F_0, \ldots, F_n$ as in Theorem 1. It is enough to work with real-valued functions, since complex-valued functions may be split into their real and imaginary parts. By homogeneity we may assume that the functions are normalized as

$$\|F_0\|_{2^n} = \|F_1\|_{2^n} = \|F_2\|_{2^{n-1}} = \cdots = \|F_n\|_{2^1} = 1. \tag{2.1}$$

Next, we pass from rough to smooth truncations of the simplex Hilbert transform. Let us write

$$\varphi(x) := \frac{\mathbb{1}_{[-R,R]\setminus[-r,r]}(x) - (g(x/R) - g(x/r))}{x} = \frac{\mathbb{1}_{[-1,1]}(x/R) - g(x/R)}{x} - \frac{\mathbb{1}_{[-1,1]}(x/r) - g(x/r)}{x},$$

where g is the Gaussian function $g(x) := e^{-\pi x^2}$. Note that φ is integrable uniformly in the truncation parameters $0 < r < R$ and that the bound

$$\Big| \int_{\mathbb{R}^{n+1}} \prod_{i=0}^{n} F_i(x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \varphi(x_0 + \dots + x_n) dx_0 \dots dx_n \Big| \le \|\varphi\|_1 \|F_0\|_{2^n} \prod_{i=1}^{n} \|F_i\|_{2^{n-i+1}}$$

follows from the change of variables $x_0 = x - x_1 - \dots - x_n$ and Hölder's inequality in $x_1, \dots, x_n$. Therefore, in order to prove Theorem 1 it suffices to prove the estimate for the kernel

$$\frac{g(x/R) - g(x/r)}{x} = -\int_r^R t^{-2} g'(t^{-1}x) dt.$$

That is, it suffices to obtain, in lieu of (1.2),

$$\Big| \int_r^R \int_{\mathbb{R}^{n+1}} \prod_{i=0}^{n} F_i(x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_n) h_t(x_0 + \dots + x_n) dx_0 \dots dx_n \frac{dt}{t} \Big| \le C \Big(\log \frac{R}{r} \Big)^{1-2^{-n+1}}, \tag{2.2}$$

where h is the derivative of g, and we use subscripts to denote L^1-normalized dilates of functions:

$$h_t(x) := t^{-1} h(t^{-1}x).$$

For the inductive statement we need to define further expressions. For $0 \le k \le n$ we define $\mathcal{F}^k$ as a function of variables $x_0, \dots, x_n, x_0^0, x_0^1, \dots, x_n^0, x_n^1 \in \mathbb{R}$ by

$$\mathcal{F}^k := \prod_{i=0}^{k} \prod_{(r_{k+1}, \dots, r_n) \in \{0,1\}^{n-k}} F_i(x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_k, x_{k+1}^{r_{k+1}}, \dots, x_n^{r_n}). \tag{2.3}$$

Note that $\mathcal{F}^k$ does not depend on $x_{k+1}, \dots, x_n$ and $x_0^0, x_0^1, \dots, x_k^0, x_k^1$. Each factor F_i in the product has the property that for each $k+1 \le j \le n$ it is independent of precisely one of the variables x_j^0 or x_j^1. If $n = 3$, the structure of $\mathcal{F}^k$ for $k = 3$, 2, and 1 is illustrated in Figures 1–3 in the next section. The set $\{0, \dots, k\} \times \{0,1\}^{n-k}$ is viewed as set of vertices of a polytope in $\mathbb{R}^n$. To each hyper-face of the polytope we associate a variable and to each vertex a function F_j of the adjacent n variables. In the cases $k = 0$ and $k = 1$, the polytope is an n-dimensional cube, while for $k = n$ the polytope is an n-dimensional simplex.

For $2 \le k \le n$ and $\alpha, \alpha_k, \dots, \alpha_n \in (0, \infty)$ we define

$$\Lambda^k_{\alpha, \alpha_k, \dots, \alpha_n} := \int_r^R \int_{\mathbb{R}^{n-k+1}} \int_{\mathbb{R}^{2n-2k}} \int_{\mathbb{R}^k} \Big| \int_{\mathbb{R}} \mathcal{F}^k h_{t\alpha_k}(x_k - p_k) dx_k \Big| g_{t\alpha}(x_0 + \dots + x_{k-1} + p_k + \dots + p_n) dx_0 \dots dx_{k-1} \Big(\prod_{i=k+1}^{n} g_{t\alpha_i}(x_i^0 - p_i) g_{t\alpha_i}(x_i^1 - p_i) dx_i^0 dx_i^1 \Big) dp_k \dots dp_n \frac{dt}{t}. \tag{2.4}$$

For $1 \le k \le n$ and $\alpha, \alpha_k, \dots, \alpha_n \in (0,\infty)$ we define

$$\widetilde{\Lambda}^k_{\alpha,\alpha_k,\dots,\alpha_n} := \int_r^R \int_{\mathbb{R}^{n-k+1}} \int_{\mathbb{R}^{2n-2k}} \Big| \int_{\mathbb{R}^k} \int_{\mathbb{R}} \mathcal{F}^k\, h_{t\alpha_k}(x_k - p_k) dx_k$$
$$h_{t\alpha}(x_0 + \dots + x_{k-1} + p_k + \dots + p_n) dx_0 \dots dx_{k-1} \Big|$$
$$\Big(\prod_{i=k+1}^n g_{t\alpha_i}(x_i^0 - p_i) g_{t\alpha_i}(x_i^1 - p_i) dx_i^0 dx_i^1 \Big) dp_k \dots dp_n \frac{dt}{t}. \tag{2.5}$$

The differences between (2.4) and (2.5) are the occurrence of $g_{t\alpha}$ versus $h_{t\alpha}$ and the position of the absolute value signs. Also, we have no need to define (2.4) for $k = 1$. Observe the trivial identity

$$h = 2^{1/2}\, h_{2^{-1/2}} * g_{2^{-1/2}}.$$

Therefore the left hand-side of (2.2) is bounded by

$$2^{1/2} \Lambda^n_{2^{-1/2},\, 2^{-1/2}}.$$

The estimate (2.2) is then a consequence of the following lemma.

All constants in what follows will depend on n and k and we write $A \lesssim B$ if there exists a finite constant C depending on n and k such that $A \le CB$.

Lemma 3. *For any $2 \le k \le n$ and any $\alpha, \alpha_k, \dots, \alpha_n \in [2^{-(n-k+1)/2}, \infty)$ we have the estimates*

$$\Lambda^k_{\alpha,\alpha_k,\dots,\alpha_n},\, \widetilde{\Lambda}^k_{\alpha,\alpha_k,\dots,\alpha_n} \lesssim (\alpha\alpha_k \dots \alpha_n)^2 \Big(\log \frac{R}{r} \Big)^{1-2^{-k+1}}. \tag{2.6}$$

For $k = 1$ and any $\alpha, \alpha_1, \dots, \alpha_n \in (0,\infty)$ we have the estimate

$$\widetilde{\Lambda}^1_{\alpha,\alpha_1,\dots,\alpha_n} \lesssim 1.$$

Proof of Lemma 3. We induct on $1 \le k \le n$ and let us begin by establishing the inductive step. Take $2 \le k \le n$ and $\alpha, \alpha_k, \dots, \alpha_n \in [2^{-(n-k+1)/2}, \infty)$. We first reduce the desired bound on $\widetilde{\Lambda}^k_{\alpha,\alpha_k,\dots,\alpha_n}$ to that on $\Lambda^k_{\alpha,\alpha_k,\dots,\alpha_n}$. We can dominate pointwise

$$|h(x)| \lesssim \int_1^\infty g_\beta(x) \beta^{-4} d\beta \tag{2.7}$$

for each $x \in \mathbb{R}$. Indeed, the right hand-side of (2.7) is comparable to x^{-4} for large $|x|$. By the triangle inequality and (2.7) we can then bound

$$\widetilde{\Lambda}^k_{\alpha,\alpha_k,\dots,\alpha_n} \lesssim \int_1^\infty \Lambda^k_{\alpha\beta,\alpha_k,\dots,\alpha_n} \beta^{-4} d\beta.$$

Assuming the estimate (2.6) for $\Lambda^k_{\alpha,\alpha_k,\dots,\alpha_n}$, the right hand side of the last display is integrable in β. Since $\alpha \in [2^{-(n-k+1)/2}, \infty)$ is arbitrary, it suffices to prove upper bounds for $\Lambda^k_{\alpha,\alpha_k,\dots,\alpha_n}$.

Now we apply the Cauchy-Schwarz inequality in the variable t, which yields

$$\big(\Lambda^k_{\alpha,\alpha_k,\ldots,\alpha_n}\big)^2 \leq \Big(\log\frac{R}{r}\Big)\int_r^R \Big(\int_{\mathbb{R}^{n-k+1}}\int_{\mathbb{R}^{2n-2k}}\int_{\mathbb{R}^k}\Big|\int_{\mathbb{R}}\mathcal{F}^k\, h_{t\alpha_k}(x_k-p_k)dx_k\Big| \\ g_{t\alpha}(x_0+\cdots+x_{k-1}+p_k+\cdots+p_n)dx_0\ldots dx_{k-1} \\ \Big(\prod_{i=k+1}^{n} g_{t\alpha_i}(x_i^0-p_i)g_{t\alpha_i}(x_i^1-p_i)dx_i^0dx_i^1\Big)dp_k\ldots dp_n\Big)^2\frac{dt}{t}.$$

We expand the definition of $\mathcal{F}^k$ and for each fixed t we apply the Cauchy-Schwarz inequality in all remaining integration variables but x_k. This way we obtain

$$\big(\Lambda^k_{\alpha,\alpha_k,\ldots,\alpha_n}\big)^2 \leq \Big(\log\frac{R}{r}\Big)\int_r^R \mathcal{M}_t\,\mathcal{N}_t\,\frac{dt}{t}, \tag{2.8}$$

where

$$\mathcal{M}_t := \int_{\mathbb{R}^{n-k+1}}\int_{\mathbb{R}^{2n-2k}}\int_{\mathbb{R}^k}\Big(\int_{\mathbb{R}}\prod_{i=0}^{k-1}\prod_{(r_{k+1},\ldots,r_n)\in\{0,1\}^{n-k}} F_i(x_0,\ldots,x_{i-1},x_{i+1},\ldots,x_k,x_{k+1}^{r_{k+1}},\ldots,x_n^{r_n}) \\ h_{t\alpha_k}(x_k-p_k)dx_k\Big)^2 g_{t\alpha}(x_0+\cdots+x_{k-1}+p_k+\cdots+p_n)dx_0\ldots dx_{k-1} \\ \Big(\prod_{i=k+1}^{n} g_{t\alpha_i}(x_i^0-p_i)g_{t\alpha_i}(x_i^1-p_i)dx_i^0dx_i^1\Big)dp_k\ldots dp_n$$

and

$$\mathcal{N}_t := \int_{\mathbb{R}^{n-k+1}}\int_{\mathbb{R}^{2n-2k}}\int_{\mathbb{R}^k}\prod_{(r_{k+1},\ldots,r_n)\in\{0,1\}^{n-k}} F_k(x_0,\ldots,x_{k-1},x_{k+1}^{r_{k+1}},\ldots,x_n^{r_n})^2 \\ g_{t\alpha}(x_0+\cdots+x_{k-1}+p_k+\cdots+p_n)dx_0\ldots dx_{k-1} \\ \Big(\prod_{i=k+1}^{n} g_{t\alpha_i}(x_i^0-p_i)g_{t\alpha_i}(x_i^1-p_i)dx_i^0dx_i^1\Big)dp_k\ldots dp_n.$$

To estimate $\mathcal{N}_t$ pointwise for each fixed t, we first integrate in p_k getting rid of $g_{t\alpha}$, then introduce the variables y_i and q_i via $x_i^0 = x_i^1 - y_i$ and $p_i = x_i^1 - q_i$, respectively. Next, we apply Hölder's inequality in variables x_0 through x_{k-1} and x_{k+1}^1 through x_n^1. Finally, we integrate the remaining Gaussian factors in y_i and q_i for $k+1 \leq i \leq n$. This yields

$$\mathcal{N}_t \leq \|F_k^2\|_{2^{n-k}}^{2^{n-k}} = \|F_k\|_{2^{n-k+1}}^{2^{n-k+1}} = 1, \tag{2.9}$$

so we have obtained an estimate which is uniform in $t > 0$.

It remains to control

$$\int_r^R \mathcal{M}_t\frac{dt}{t}. \tag{2.10}$$

Expanding the square in the definition of $\mathcal{M}_t$, the expression (2.10) becomes the special case $k = j$ of the following more general expressions defined for $j \geq k$:

$$\Theta^{(j)} := \int_r^R \int_{\mathbb{R}^{n-k+1}} \int_{\mathbb{R}^{2n-2k+2}} \int_{\mathbb{R}^k} \mathcal{F}^{k-1} \tag{2.11}$$

$$g_{t\alpha}(x_0 + \cdots + x_{k-1} + p_k + \cdots + p_n) dx_0 \dots dx_{k-1}$$

$$h_{t\alpha_j}(x_j^0 - p_j) h_{t\alpha_j}(x_j^1 - p_j) dx_j^0 dx_j^1 \Big(\prod_{\substack{i=k \\ i \neq j}}^{n} g_{t\alpha_i}(x_i^0 - p_i) g_{t\alpha_i}(x_i^1 - p_i) dx_i^0 dx_i^1 \Big) dp_k \dots dp_n \frac{dt}{t}.$$

Also define

$$\Theta := -\Big(1 + \alpha^{-2} \sum_{j=k}^{n} \alpha_j^2\Big) \int_r^R \int_{\mathbb{R}^{n-k+2}} \int_{\mathbb{R}^{2n-2k+2}} \int_{\mathbb{R}^{k-1}} \int_{\mathbb{R}} \mathcal{F}^{k-1}\, h_{t\alpha 2^{-1/2}}(x_{k-1} - p_{k-1}) dx_{k-1}$$

$$h_{t\alpha 2^{-1/2}}(x_0 + \cdots + x_{k-2} + p_{k-1} + \cdots + p_n) dx_0 \dots dx_{k-2}$$

$$\Big(\prod_{i=k}^{n} g_{t\alpha_i}(x_i^0 - p_i) g_{t\alpha_i}(x_i^1 - p_i) dx_i^0 dx_i^1 \Big) dp_{k-1} \dots dp_n \frac{dt}{t}.$$

We claim that

$$\Theta + \sum_{j=k}^{n} \Theta^{(j)} \lesssim 1. \tag{2.12}$$

Before proving the claim, we show how it can be used to control $\Theta^{(k)}$. Note that $\Theta^{(k)}$ is non-negative because the real-valued terms in the expression assemble into an integral of a square that came from previous application of the Cauchy-Schwarz inequality. The terms $\Theta^{(j)}$ are also non-negative for each $j \geq k$; the argument is the same after renaming the variables. Therefore, comparing the definitions of Θ and $\widetilde{\Lambda}^k_{\alpha,\alpha_k,\dots,\alpha_n}$,

$$\Theta^{(k)} \leq \sum_{j=k}^{n} \Theta^{(j)} \lesssim 1 + |\Theta| \leq 1 + \Big(1 + \alpha^{-2} \sum_{j=k}^{n} \alpha_j^2\Big) \widetilde{\Lambda}^{k-1}_{\alpha 2^{-1/2}, \alpha 2^{-1/2}, \alpha_k, \dots, \alpha_n}.$$

By the induction hypothesis (i.e. the statement for $k-1$), we may estimate this display further by

$$\lesssim \alpha^{-2} \Big(\alpha^2 + \sum_{j=k}^{n} \alpha_j^2\Big) (\alpha^2 \alpha_k \dots \alpha_n)^2 \Big(\log \frac{R}{r}\Big)^{1-2^{-k+2}} \lesssim (\alpha \alpha_k \dots \alpha_n)^4 \Big(\log \frac{R}{r}\Big)^{1-2^{-k+2}},$$

where we have estimated the sum of the squared alphas by their product. We combine this estimate with (2.8) and (2.9). Multiplying with $\log(R/r)$ and taking the square root shows (2.6) for the given k, completing the induction step up to the verification of the claim (2.12).

To see this claim, we employ the Fourier transform which we normalize as

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

For fixed $x_0, \ldots, x_{k-1}, x_k^0, x_k^1, \ldots, x_n^0, x_n^1$ the integral in $p_k, \ldots, p_n$ in $\Theta^{(j)}$ is the integral of the function

$$
\begin{aligned}
H(q, q_k^0, q_k^1, \ldots, q_n^0, q_n^1) :=& g_{t\alpha}(q + x_0 + \cdots + x_{k-1}) \\
& h_{t\alpha_j}(q_j^0 - x_j^0) h_{t\alpha_j}(q_j^1 - x_j^1) \Big(\prod_{\substack{i=k \\ i \neq j}}^{n} g_{t\alpha_i}(q_i^0 - x_i^0) g_{t\alpha_i}(q_i^1 - x_i^1) \Big)
\end{aligned}
$$

over the $(n-k+1)$-dimensional subspace

$$\{(p_k + \cdots + p_n,\, p_k,\, p_k,\, p_{k+1},\, p_{k+1}, \ldots, p_n,\, p_n) : p_k, \ldots, p_n \in \mathbb{R}\}$$

of $\mathbb{R}^{2n-2k+3}$. The orthogonal complement of this subspace is

$$\{(\eta,\, \xi_k,\, -\xi_k - \eta,\, \xi_{k+1},\, -\xi_{k+1} - \eta, \ldots, \xi_n,\, -\xi_n - \eta) : \eta, \xi_k, \ldots, \xi_n \in \mathbb{R}\}.$$

The previously mentioned integral is equal to the integral of the Fourier transform of H over this orthogonal complement, which in turn becomes

$$
\begin{aligned}
& \int_{\mathbb{R}^{n-k+2}} \widehat{g_{t\alpha}}(\eta) \widehat{h_{t\alpha_j}}(\xi_j) \widehat{h_{t\alpha_j}}(-\xi_j - \eta) \prod_{\substack{i=k \\ i \neq j}}^{n} \widehat{g_{t\alpha_i}}(\xi_i) \widehat{g_{t\alpha_i}}(\xi_i + \eta) \\
& e^{2\pi \mathbb{i} \left((x_0 + \cdots + x_{k-1})\eta - \sum_{i=k}^{n} (x_i^0 \xi_i + x_i^1(-\xi_i - \eta)) \right)} \, d\eta d\xi_k \ldots d\xi_n. \qquad (2.13)
\end{aligned}
$$

Quite similarly, the integral in $p_{k-1}, \ldots, p_n$ in Θ can be expressed as

$$
\begin{aligned}
& \int_{\mathbb{R}^{n-k+2}} \widehat{h_{t\alpha 2^{-1/2}}}(\eta) \widehat{h_{t\alpha 2^{-1/2}}}(-\eta) \prod_{i=k}^{n} \widehat{g_{t\alpha_i}}(\xi_i) \widehat{g_{t\alpha_i}}(\xi_i + \eta) \\
& e^{2\pi \mathbb{i} \left((x_0 + \cdots + x_{k-1})\eta - \sum_{i=k}^{n} (x_i^0 \xi_i + x_i^1(-\xi_i - \eta)) \right)} \, d\eta d\xi_k \ldots d\xi_n. \qquad (2.14)
\end{aligned}
$$

Now we state the crucial "telescoping" or "integration by parts" identity

$$
\begin{aligned}
& \Big(1 + \alpha^{-2} \sum_{j=k}^{n} \alpha_j^2 \Big) \int_r^R \widehat{h_{t\alpha 2^{-1/2}}}(\eta) \widehat{h_{t\alpha 2^{-1/2}}}(-\eta) \prod_{i=k}^{n} \widehat{g_{t\alpha_i}}(\xi_i) \widehat{g_{t\alpha_i}}(\xi_i + \eta) \frac{dt}{t} \\
& + \sum_{j=k}^{n} \int_r^R \widehat{g_{t\alpha}}(\eta) \widehat{h_{t\alpha_j}}(\xi_j) \widehat{h_{t\alpha_j}}(-\xi_j - \eta) \prod_{\substack{i=k \\ i \neq j}}^{n} \widehat{g_{t\alpha_i}}(\xi_i) \widehat{g_{t\alpha_i}}(\xi_i + \eta) \frac{dt}{t} \\
& = \pi \big(G_r(\eta, \xi_k, \ldots, \xi_n) - G_R(\eta, \xi_k, \ldots, \xi_n) \big), \qquad (2.15)
\end{aligned}
$$

where for $t > 0$ we have denoted

$$G_t(\eta, \xi_k, \ldots, \xi_n) := \widehat{g_{t\alpha}}(\eta) \prod_{j=k}^{n} \widehat{g_{t\alpha_j}}(\xi_j) \widehat{g_{t\alpha_j}}(\xi_j + \eta).$$

To see this identity, we use the fundamental theorem of calculus, together with $\widehat{g}(\xi) = e^{-\pi\xi^2}$, which yields that the right hand side of the identity (2.15) equals

$$
\begin{aligned}
& - \int_r^R \pi t \partial_t (G_t(\eta, \xi_k, \ldots, \xi_n)) \frac{dt}{t} \\
& = \int_r^R 2\pi^2 t^2 \Big(\alpha^2 \eta^2 + \sum_{j=k}^{n} \alpha_j^2 (\xi_j^2 + (\xi_j + \eta)^2) \Big) G_t(\eta, \xi_k, \ldots, \xi_n) \frac{dt}{t}.
\end{aligned}
$$

Using $\widehat{h}(\xi) = 2\pi\mathbb{i}\xi\widehat{g}(\xi)$ gives

$$\widehat{h}(t\alpha 2^{-1/2}\eta)\widehat{h}(t\alpha 2^{-1/2}(-\eta)) = (2\pi\mathbb{i}t\alpha)^2 2^{-1}\eta(-\eta)\widehat{g}(t\alpha 2^{-1/2}\eta)\widehat{g}(t\alpha 2^{-1/2}\eta) = 2\pi^2 t^2\alpha^2\eta^2\widehat{g}(t\alpha\eta)$$

and

$$\widehat{h}(t\alpha_j\xi_j)\widehat{h}(t\alpha_j(-\xi_j-\eta)) = 4\pi^2 t^2\alpha_j^2\xi_j(\xi_j+\eta)\widehat{g}(t\alpha_j\xi_j)\widehat{g}(t\alpha_j(\xi_j+\eta)),$$

so the left hand side of (2.15) becomes

$$\int_r^R \Big(1+\alpha^{-2}\sum_{j=k}^n \alpha_j^2\Big) 2\pi^2 t^2\alpha^2\eta^2\, G_t(\eta,\xi_k,\ldots,\xi_n)\frac{dt}{t}$$
$$+\int_r^R \Big(\sum_{j=k}^n 4\pi^2 t^2\alpha_j^2\xi_j(\xi_j+\eta)\Big)\, G_t(\eta,\xi_k,\ldots,\xi_n)\frac{dt}{t}.$$

A straightforward polynomial identity finally establishes (2.15).

The terms on the left hand side of (2.15) correspond to the terms on the left hand side of (2.12): one only needs to multiply (2.15) with $\mathcal{F}^{k-1}$ and the complex exponential from (2.13), (2.14), and perform the remaining integrations. We thus need to show that the corresponding terms for the right hand side of (2.15) can be bounded by a constant. However, for $t=r$ or $t=R$ we have

$$\Big|\int_{\mathbb{R}^{n-k+1}}\int_{\mathbb{R}^{2n-2k+2}}\int_{\mathbb{R}^k}\mathcal{F}^{k-1}$$
$$g_{t\alpha}(x_0+\cdots+x_{k-1}+p_k+\cdots+p_n)dx_0\ldots dx_{k-1}$$
$$\Big(\prod_{i=k}^n g_{t\alpha_i}(x_i^0-p_i)g_{t\alpha_i}(x_i^1-p_i)dx_i^0 dx_i^1\Big)dp_k\ldots dp_n\Big| \le \|F_0\|_{2^n}^{2^{n-k+1}}\prod_{i=1}^{k-1}\|F_i\|_{2^{n-i+1}}^{2^{n-k+1}} = 1, \tag{2.16}$$

i.e. these single-scale estimates are uniform in $t>0$ and $\alpha_i>1$. This follows by first introducing new variables y, y_i, and q_i via $x_0 = y-x_1-x_2-\cdots-x_{k-1}$, $x_i^0 = x_i^1 - y_i$, and $p_i = x_i^1 - q_i$. With these new variables, we first apply Hölder's inequality in $x_1,\ldots,x_{k-1}$, then integrate in y, then apply Hölder's inequality in $x_k^1,\ldots,x_n^1$, and finally integrate in y_i and q_i for $k\le i\le n$.

Inserting (2.15) into (2.13) and (2.14), passing to the spatial side and using the estimate (2.16) we obtain the desired claim (2.12). This completes the proof of the inductive step.

It remains to establish the base case $k=1$ of the induction, i.e. to estimate $\widetilde{\Lambda}^1_{\alpha,\alpha_k,\ldots,\alpha_n}$. Unlike in the inductive step we do not dominate one of the functions h. Instead we apply the Cauchy-Schwarz inequality to (2.5) immediately in such a way that each of the terms on the right hand side invokes cancellative functions h. This is possible only in the case $k=1$ because here the integration in the variables x_0 and x_1 separates. More precisely, we apply the Cauchy-Schwarz inequality in the integrals over the variables $x_2^0, x_2^1, \ldots x_n^0, x_n^1$, $p_1,\ldots,p_n$, and t to obtain

$$\big(\widetilde{\Lambda}^1_{\alpha,\alpha_1,\ldots,\alpha_n}\big)^2 \le \widetilde{\Theta}^{(1)}(F_0)\widetilde{\Theta}^{(1)}(F_1), \tag{2.17}$$

where for $1 \leq j \leq n$ and a Schwartz function F on $\mathbb{R}^n$ we have set

$$\widetilde{\Theta}^{(j)}(F) := \int_r^R \int_{\mathbb{R}^n} \int_{\mathbb{R}^{2n}} \prod_{(r_1,\ldots,r_n)\in\{0,1\}^n} F(x_1^{r_1},\ldots,x_n^{r_n})$$
$$h_{t\alpha_j}(x_j^0 - p_j) h_{t\alpha_j}(x_j^1 - p_j) dx_j^0 dx_j^1 \Big(\prod_{\substack{i=1 \\ i\neq j}}^{n} g_{t\alpha_i}(x_i^0 - p_i) g_{t\alpha_i}(x_i^1 - p_i) dx_i^0 dx_i^1 \Big) dp_1 \ldots dp_n \frac{dt}{t}.$$

Similarly as in the inductive step, we now have

$$\sum_{j=1}^{n} \widetilde{\Theta}^{(j)}(F) \lesssim 1 \tag{2.18}$$

for any F with $\|F\|_{2^n} = 1$. Namely, $\widetilde{\Theta}^{(j)}(F)$ coincides with $\Theta^{(j)}$ for $k = 1$, except for the choice of functions F making up $\mathcal{F}^{k-1}$. Moreover, $\mathcal{F}^0$ does not depend on x_0, so the integral in x_0 is merely the integral of a Gaussian. Likewise, the integral in x_0 in the definition of Θ for $k = 1$ is an integral over the derivative of a Gaussian and hence vanishes. Thus claim (2.18) follows analogously to claim (2.12).

It remains to observe that $\widetilde{\Theta}^{(j)} \geq 0$ for each $1 \leq j \leq n$, which is again analogous to the proof of the inductive step: simply observe that we are integrating squares of real-valued expressions. Together with (2.18) this implies

$$\widetilde{\Theta}^{(1)}(F_0),\, \widetilde{\Theta}^{(1)}(F_1) \lesssim 1,$$

which by (2.17) concludes the proof of the base case $k = 1$ of the induction. □

3 An illustration of the induction steps

Figures 1–3 represent the induction scheme for $n = 3$. The polyhedra in Figures 1–3 represent the structure of $\mathcal{F}^k$ for $k = 3$, 2, and 1 in this order. The vertices represent the various factors F_j in the definition of $\mathcal{F}^k$, while the faces represent the arguments in these factors, such that adjacency of a face to a vertex means that the argument appears in the corresponding factor of $\mathcal{F}^k$.

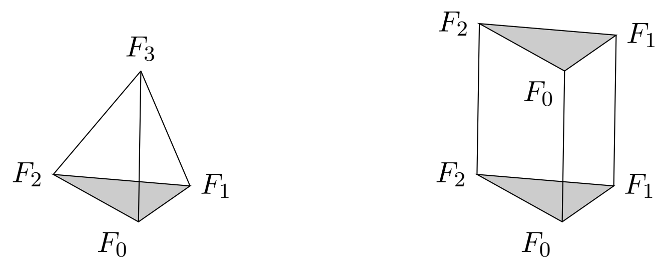

Figure 1: Case $k = 3$.

The passage from left to right polyhedron in each figure represents the effect of the Cauchy-Schwarz inequality (2.8), passing from a form $\Lambda^k_{\alpha,\alpha_k,\ldots,\alpha_n}$ involving $\mathcal{F}^k$ on the left to a form $\mathcal{M}_t$ or $\Theta^{(k)}$ involving $\mathcal{F}^{k-1}$ on the right.

The shaded faces of the left polyhedra correspond to the variable x_k in $\Lambda^k_{\alpha,\alpha_k,\dots,\alpha_n}$ appearing in the cancellative function h. On the right hand side this variable has bifurcated into two variables x_k^0 and x_k^1 in $\Theta^{(k)}$, both of which still carry cancellation.

Comparing the right polyhedron in one figure to the left polyhedron in the next figure, the shaded faces move to a different location indicating the effect of the telescoping estimate (2.12). Note that the picture depicts only the most important of, in general many, terms in the telescoping identity. In all but the last figure we have only one shaded face on the left polyhedron, since after domination of one function h by Gaussians only one function h survives.

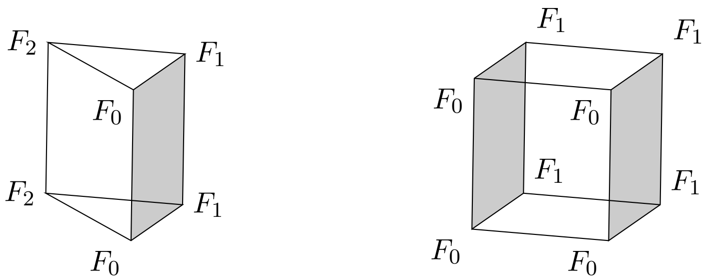

Figure 2: Case $k = 2$.

The last figure corresponds to the base case, which is treated differently. On the one hand we have two shaded faces of the left polyhedron, and on the other hand the Cauchy-Schwarz inequality does not change the geometry of the polyhedron, but merely the labeling of the corners. This stabilization of the process is ultimately the reason that the recursion stops.

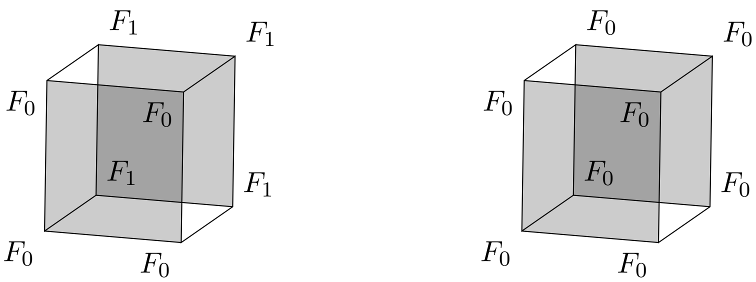

Figure 3: Case $k = 1$.

4 Dyadic model of the simplex Hilbert transform

In this section we discuss the analogue of Theorem 1 for the dyadic model of the truncated simplex Hilbert transform. Define

$$\Lambda^{\mathrm{d}}_{n,m} := \sum_{l=0}^{m-1} \sum_{(I_0,\dots,I_n)\in\mathcal{I}_l} \epsilon_{l,I_0,\dots,I_n} \int_{(\mathbb{R}_+)^{n+1}} \prod_{i=0}^{n} F_i(x_0,\dots,x_{i-1},x_{i+1},\dots,x_n) 2^{-l} \Big(\prod_{i=0}^{n} \mathrm{h}_{I_i}(x_i) dx_i \Big),$$

where $n, m \geq 1$, $\mathbb{R}_+ = [0, \infty)$, and for $l \in \mathbb{Z}$ we denote

$$\mathcal{I}_l := \{(I_0, \ldots, I_n) : 0 \in I_0 \oplus \cdots \oplus I_n,\ I_i \text{ dyadic interval},\ I_i \subset \mathbb{R}_+,\ |I_i| = 2^l,\ 1 \leq i \leq n\}.$$

Here a dyadic interval is any interval of the form $[2^l m, 2^l(m+1))$ with $m, l \in \mathbb{Z}$ and $\oplus$ is the addition of the Walsh group; see [4] for further details. The otherwise arbitrary coefficients $\epsilon_{l,I_0,\ldots,I_n}$ are assumed to be bounded in the absolute value by 1 and we have denoted by $\mathbb{h}_I$ the L^∞-normalized Haar function on I. A convenient property of the Haar functions is that

$$\mathbb{h}_{I_1 \oplus I_2}(x_1 \oplus x_2) = \mathbb{h}_{I_1}(x_1)\mathbb{h}_{I_2}(x_2)$$

whenever I_1, I_2 are dyadic intervals of the same length, $x_1 \in I_1$, $x_2 \in I_2$, and $I_1 \oplus I_2$ is defined to be yet another dyadic interval of that same length whose left endpoint is the $\oplus$-sum of the left endpoints of I_1 and I_2. Indeed, this is simply the character property of the more general Walsh functions. In dyadic models it is common to replace $1/(x_0 + \cdots + x_n)$ with kernels such as

$$K(x_0, \ldots, x_n) = \sum_l \epsilon_l 2^{-l} \mathbb{h}_{[0,2^l)}(x_1 \oplus \cdots \oplus x_n) = \sum_l \sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l} \epsilon_l 2^{-l} \prod_{i=0}^{n} \mathbb{h}_{I_i}(x_i).$$

This time the trivial estimate grows linearly in the number of scales m and we want to improve on this trivial bound with a power less than one.

Theorem 4. *There exists a finite constant C depending only on $n \geq 1$ such that for any tuple $F_0, \ldots, F_n$ of finite linear combinations of Haar functions and any $m \geq 1$ we have*

$$|\Lambda^{\mathrm{d}}_{n,m}| \leq C m^{1-2^{-n+1}} \|F_0\|_{2^n} \prod_{i=1}^{n} \|F_i\|_{2^{n-i+1}}.$$

Sketch of proof. Fix positive integers n, m and functions $F_0, \ldots, F_n$ normalized as in (2.1). In order to perform the structural induction we introduce expressions indexed by $1 \leq k \leq n$

$$\Lambda^{\mathrm{d},k} := \sum_{l=0}^{m-1} \sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l} \int_{(\mathbb{R}_+)^{2n-2k}} \Big| \int_{(\mathbb{R}_+)^{k+1}} \mathcal{F}^k (2^{-l})^{n-k+1} \Big(\prod_{i=0}^{k} \mathbb{h}_{I_i}(x_i) dx_i \Big) \Big|$$
$$\Big(\prod_{i=k+1}^{n} \mathbb{1}_{I_i}(x_i^0) \mathbb{1}_{I_i}(x_i^1) dx_i^0 dx_i^1 \Big),$$

where $\mathcal{F}^k$ is defined as in (2.3). We claim that

$$\Lambda^{\mathrm{d},k} \lesssim m^{1-2^{-k+1}} \tag{4.1}$$

for each $1 \leq k \leq n$. Since $|\Lambda^{\mathrm{d}}_{n,m}| \leq \Lambda^{\mathrm{d},n}$, this then implies the theorem.

We prove (4.1) by induction on k and begin with the inductive step. Let $2 \leq k \leq n$. Performing the analogous steps from (2.8) to (2.9) we obtain

$$\big(\Lambda^{\mathrm{d},k}\big)^2 \lesssim m \sum_{l=0}^{m-1} \mathcal{M}^{\mathrm{d}}_l, \tag{4.2}$$

where

$$\mathcal{M}_l^{\mathrm{d}} := \sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l} \int_{(\mathbb{R}_+)^{2n-k}} \Big| \int_{\mathbb{R}_+} \prod_{i=0}^{k-1} \prod_{(r_{k+1},\ldots,r_n)\in\{0,1\}^{n-k}} F_i(x_0,\ldots,x_{i-1},x_{i+1},\ldots,x_k,x_{k+1}^{r_{k+1}},\ldots,x_n^{r_n})$$
$$2^{-l}\mathbb{h}_{I_k}(x_k)dx_k\Big|^2 (2^{-l})^{n-k}\Big(\prod_{i=0}^{k-1}\mathbb{1}_{I_i}(x_i)dx_i\Big)\Big(\prod_{i=k+1}^{n}\mathbb{1}_{I_i}(x_i^0)\mathbb{1}_{I_i}(x_i^1)dx_i^0dx_i^1\Big).$$

Therefore it remains to control $\sum_{l=0}^{m-1}\mathcal{M}_l^{\mathrm{d}}$, which can be rewritten, in analogy with display (2.11), as

$$\sum_{l=0}^{m-1} \sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l} \int_{(\mathbb{R}_+)^{2n-k+2}} \mathcal{F}^{k-1} \tag{4.3}$$
$$(2^{-l})^{n-k+2}\Big(\prod_{i=0}^{k-1}\mathbb{1}_{I_i}(x_i)dx_i\Big)\Big(\mathbb{h}_{I_k}(x_k^{(0)})\mathbb{h}_{I_k}(x_k^{(1)})dx_k^{(0)}dx_k^{(1)}\Big)\Big(\prod_{i=k+1}^{n}\mathbb{1}_{I_i}(x_i^{(0)})\mathbb{1}_{I_i}(x_i^{(1)})dx_i^{(0)}dx_i^{(1)}\Big).$$

The identity (2.15) is now replaced by the dyadic "telescoping" identity

$$\sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l} \bigg(\Big(\prod_{i=0}^{k-1}\mathbb{h}_{I_i}(x_i)\Big)\Big(\prod_{i=k}^{n}\big(\mathbb{1}_{I_i}(x_i^{(0)})\mathbb{h}_{I_i}(x_i^{(1)})+\mathbb{h}_{I_i}(x_i^{(0)})\mathbb{1}_{I_i}(x_i^{(1)})\big)\Big)$$
$$+\Big(\prod_{i=0}^{k-1}\mathbb{1}_{I_i}(x_i)\Big)\Big(\prod_{i=k}^{n}\big(\mathbb{1}_{I_i}(x_i^{(0)})\mathbb{1}_{I_i}(x_i^{(1)})+\mathbb{h}_{I_i}(x_i^{(0)})\mathbb{h}_{I_i}(x_i^{(1)})\big)\Big)\bigg)$$
$$= 2^{n-k+2}\sum_{(I_0,\ldots,I_n)\in\mathcal{I}_{l-1}}\Big(\prod_{i=0}^{k-1}\mathbb{1}_{I_i}(x_i)\Big)\Big(\prod_{i=k}^{n}\mathbb{1}_{I_i}(x_i^{(0)})\mathbb{1}_{I_i}(x_i^{(1)})\Big). \tag{4.4}$$

In order to verify it, we split each interval I_i on the left hand side into its left "child" I_i^0 and its right "child" I_i^1, so that (4.4) turns into

$$\frac{1}{2}\sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l}\bigg(\Big(\prod_{i=0}^{k-1}\big(\mathbb{1}_{I_i^0}(x_i)-\mathbb{1}_{I_i^1}(x_i)\big)\Big)\Big(\prod_{i=k}^{n}\big(\mathbb{1}_{I_i^0}(x_i^{(0)})\mathbb{1}_{I_i^0}(x_i^{(1)})-\mathbb{1}_{I_i^1}(x_i^{(0)})\mathbb{1}_{I_i^1}(x_i^{(1)})\big)\Big)$$
$$+\Big(\prod_{i=0}^{k-1}\big(\mathbb{1}_{I_i^0}(x_i)+\mathbb{1}_{I_i^1}(x_i)\big)\Big)\Big(\prod_{i=k}^{n}\big(\mathbb{1}_{I_i^0}(x_i^{(0)})\mathbb{1}_{I_i^0}(x_i^{(1)})+\mathbb{1}_{I_i^1}(x_i^{(0)})\mathbb{1}_{I_i^1}(x_i^{(1)})\big)\Big)\bigg)$$
$$=\sum_{(I_0,\ldots,I_n)\in\mathcal{I}_{l-1}}\Big(\prod_{i=0}^{k-1}\mathbb{1}_{I_i}(x_i)\Big)\Big(\prod_{i=k}^{n}\mathbb{1}_{I_i}(x_i^{(0)})\mathbb{1}_{I_i}(x_i^{(1)})\Big).$$

This identity becomes apparent once we observe that the tuple $(I_0^{s_0},\ldots,I_n^{s_n})$ for some $(s_0,\ldots,s_n)\in\{0,1\}^{n+1}$ belongs to $\mathcal{I}_{l-1}$ if and only if the number of s_i that are equal to 1 is even.

What we have in (4.3) can be recognized as one of the terms beginning with $\mathbb{1}$'s in (4.4), after multiplying (4.4) by $\mathcal{F}^{k-1}$, integrating and finally summing over the intervals and l. All terms in the second line of (4.4) lead to non-negative expressions analogous to (2.11), so it suffices to control their sum. What remains after summing the above identity in l, up to single-scale quantities analogous to (2.16), are the terms beginning with $\mathbb{h}$'s. By the triangle inequality, these terms lead to at most 2^n times

$$\sum_{l=0}^{m-1} \sum_{(I_0,\ldots,I_n)\in\mathcal{I}_l} \int_{(\mathbb{R}_+)^{2n-2k+2}} \bigg| \int_{(\mathbb{R}_+)^k} \mathcal{F}^{k-1} \, (2^{-l})^{n-k+2} \Big(\prod_{i=0}^{k-1} \mathbb{h}_{I_i}(x_i) dx_i \Big) \bigg| \Big(\prod_{i=k}^{n} \mathbb{1}_{I_i}(x_i^{(0)}) \mathbb{1}_{I_i}(x_i^{(1)}) dx_i^{(0)} dx_i^{(1)} \Big),$$

which can be recognized as $\Lambda^{\mathrm{d},k-1}$. Applying the induction hypothesis combined with (4.2) finishes the inductive step.

The base case $k = 1$ can be deduced similarly as in the previous section. □

Acknowledgments

P. D. and C. T. are supported by the Hausdorff Center for Mathematics. V. K. is supported in part by the Croatian Science Foundation under the project 3526.

References

[1] P. Durcik, *An L^4 estimate for a singular entangled quadrilinear form*, Math. Res. Lett. **22** (2015), no. 5, 1317–1332.

[2] P. Durcik, V. Kovač, K. A. Škreb, C. Thiele, *Norm-variation of ergodic averages with respect to two commuting transformations*, (2016) preprint, available at arXiv:1603.00631.

[3] V. Kovač, *Boundedness of the twisted paraproduct*, Rev. Mat. Iberoam. **28** (2012), no. 4, 1143–1164.

[4] V. Kovač, C. Thiele, P. Zorin-Kranich, *Dyadic triangular Hilbert transform of two general and one not too general function*, Forum of Mathematics, Sigma **3** (2015), e25.

[5] T. Tao, *Cancellation for the multilinear Hilbert transform*, Collect. Math. **67** (2016), no. 2, 191–206.

[6] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform* (2015), To appear in Math. Res. Lett., available at arXiv:1507.02436.

Polona Durcik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: durcik@math.uni-bonn.de

Vjekoslav Kovač, Department of Mathematics, Faculty of Science, University of Zagreb, Bijenička cesta 30, 10000 Zagreb, Croatia
E-mail address: vjekovac@math.hr

Christoph Thiele, Mathematisches Institut, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: thiele@math.uni-bonn.de

Vita

2011/09	B.Sc. in Mathematics University of Ljubljana, Slovenia
2014/02	M.Sc. in Mathematics University of Bonn, Germany
2014/04–present	Ph.D. studies in Mathematics University of Bonn, Germany

2008–2014	Zois scholarship Slovenian national merit scholarship
2014–2017	Hausdorff scholarship doctoral scholarship by Bonn Intl. Graduate School of Mathematics