Norm variation of ergodic averages with respect to two commuting transformations

POLONA DURCIK†, VJEKOSLAV KOVAČ‡, KRISTINA ANA ŠKREB§ and
CHRISTOPH THIELE†

† *Mathematisches Institut, Universität Bonn, Endenicher Allee 60,*
53115 Bonn, Germany
(e-mail: durcik@math.uni-bonn.de, thiele@math.uni-bonn.de)
‡ *Department of Mathematics, Faculty of Science, University of Zagreb,*
Bijenička cesta 30, 10000 Zagreb, Croatia
(e-mail: vjekovac@math.hr)
§ *Faculty of Civil Engineering, University of Zagreb, Fra Andrije Kačića Miošića 26,*
10000 Zagreb, Croatia
(e-mail: kskreb@grad.hr)

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 $S : X \to X$ 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* [**33**], which guarantees convergence of the *single ergodic averages*

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

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

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

for any positive integers $n_0 < n_1 < \cdots < n_m$ and with an absolute finite constant C. The work of Bourgain [8] 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 §5, 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 [16–18] connecting ergodic theory with arithmetic combinatorics. In this paper we are concerned with the bilinear case. Let $S, T : X \to X$ 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 $f, g \in L^\infty(X)$ 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 $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$, 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 $f, g \in L^4(X)$, we have*

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

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, 8, 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 (a.e.) 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 [8] and Jones *et al* [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 $T = S^{-1}$. It was successfully studied by the analytic approach and an almost everywhere convergence result was established by Bourgain [9]. 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 $a : \mathcal{U} \to B$ with $\mathcal{U} \subseteq \mathbb{R}$ is defined as

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

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 in Theorem 1 is the best possible one. 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 multilinear and multiscale singular integral operators gradually developed by the authors in [14, 15, 22–26]. 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 us to iterate these steps, with the Cauchy–Schwarz inequality used in between to reduce the complexity of the expressions.

Let us 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 φ. 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 non-zero function φ.

Analytic reformulation of the aforementioned open problem on the a.e. convergence of the averages (1.3) would require a strengthening of Theorem 2 involving pointwise variation of the bilinear averages on $\mathbb{R}^2$. Even though our techniques are not sufficient for controlling the latter quantity in its full generality, we can still establish an estimate for the so-called 'short pointwise variation'. The following corollary is not really a consequence of Theorem 2, but rather a byproduct of Lemma 9 below and the discussion in §2.3. We formulate it here to emphasize that our short variation argument does not distinguish between pointwise and norm variations.

COROLLARY 3. *For any Schwartz function φ there exists a finite constant C_{φ} such that for any $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ we have*

$$\left(\sum_{i=-\infty}^{\infty} \big\| \|A_t^{\varphi}(F, G)(x, y)\|_{\mathrm{V}_t^2([2^i, 2^{i+1}], \mathbb{C})} \big\|_{\mathrm{L}_{(x,y)}^2(\mathbb{R}^2)}^2 \right)^{1/2} \leq C_{\varphi} \|F\|_{\mathrm{L}^4(\mathbb{R}^2)} \|G\|_{\mathrm{L}^4(\mathbb{R}^2)}.$$

While deriving Theorem 1 from Theorem 2, the following discrete estimate will appear along the way. It is also worth stating as a separate corollary due to its elegant formulation. For any two double sequences $\widetilde{F}, \widetilde{G} : \mathbb{Z}^2 \to \mathbb{R}$, 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 4. *There exists a finite constant C such that for any $\widetilde{F}, \widetilde{G} \in \ell^4(\mathbb{Z}^2)$ we have*

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

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

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

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

COROLLARY 5. *For any $F, G \in \mathrm{L}^4(\mathbb{R}^2)$ we have*

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

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 §2.

In contrast with Corollary 5, no bounds are known for the corresponding bilinear singular integral

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

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 [37], 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 5 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 non-negative function on $\mathbb{R}$ that is constantly 1 on the interval $[-1, 1]$. By substituting $z = x - y$, observing that

$$\int_{-R}^R \int_{-R}^R \left| \int_{\mathbb{R}} F(x + s, y)G(x, y + s)2^{-j}\psi(2^{-j}s) \, ds \right|^2 dx \, dy$$

$$\ge \int_{-R}^R \left| \int_{\mathbb{R}} f(z + s)g(z - s)2^{-j}\psi(2^{-j}s) \, ds \right|^2 dz,$$

applying Corollary 5, 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

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

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

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

for an arbitrary collection of disjoint intervals Ω, which would be a bilinear variant of the well-known result by Rubio de Francia [**35**]. Here $\widecheck{\mathbb{1}}_\omega$ 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**, **34**]. 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 §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 Lemmas 9 and 10. Propositions 6 and 7 are the key results here. Their proofs are postponed to §§3 and 4 and these two sections contain the main novelties of our approach. Finally, the somewhat standard transition from Theorem 2 to Corollary 4 and then to Theorem 1 is presented in detail in §5.

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 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 the form

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

whenever $\varphi, \widehat{\varphi} \in L^1(\mathbb{R}^d)$. 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 two propositions that will be the key ingredients in the proof of Theorem 2. Their own proofs will be postponed to the subsequent sections.

PROPOSITION 6. *Let $\lambda > 1$ and let ϑ, $\varphi \in \mathcal{S}(\mathbb{R})$ be such that*

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

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

$$\left| \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(x + u, y)G(x, y + u)F(x + v, y)G(x, y + v) \right.$$

$$\left. \times \vartheta_{2^{k_j}}(u)(\varphi_{2^{k_j}} - \varphi_{2^{k_{j-1}}})(v) \, dx \, dy \, du \, dv \right| \lesssim_\lambda 1. \tag{2.2}$$

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

$$|\Phi(u, v)| \le (1 + |u + v|)^{-\lambda}(1 + |u - v|)^{-2\lambda} \tag{2.3}$$

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

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

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

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

implies (2.3) within an absolute constant. Moreover, (2.3) implies (2.5) 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) within an absolute constant, since there is no control on $k_j - k_{j-1}$. However, the form in Proposition 6 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.

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, non-negative, 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 non-negative 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 6 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*

$$|\phi * \chi_{2^4}(s)| \leq C_0(1 + |s|)^{-\lambda}, \quad |\phi(s)| \leq C_1(1 + |s|)^{-\lambda} \qquad (2.6)$$

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

$$|\phi(u)\overline{\phi(v)}| \leq C_2(1 + |u + v|)^{-\lambda}(1 + |u - v|)^{-2\lambda} \qquad (2.7)$$

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 $F, G \in L^4(\mathbb{R}^2)$ are normalized by (2.1), then

$$\|A_{2^k}^{\phi}(F, G)\|_{V_k^2(\mathbb{Z}, L^2(\mathbb{R}^2))} \lesssim_\lambda C_0^{1/2} C_1^{1/2} + C_2^{1/2}. \qquad (2.8)$$

Observe that if $\widehat{\phi}$ vanishes on $[-2^{-2}, 2^{-2}]$, then the first estimate in (2.6) holds with $C_0 = 0$. In this case Lemma 8 yields

$$\|A_{2^k}^{\phi}(F, G)\|_{V_k^2(\mathbb{Z}, L^2(\mathbb{R}^2))} \lesssim_\lambda C_2^{1/2}. \qquad (2.9)$$

Proof of Lemma 8. Standard limiting arguments reduce the estimate (2.8) 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

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

Expanding the L^2 norm gives

$$V(F, G) = \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(x + u, y)G(x, y + u)F(x + v, y)G(x, y + v)$$
$$\times (\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v) \, dx \, dy \, du \, dv.$$

We have the identity

$$(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v) = \left(\phi_{2^{k_{j-1}}}(u)\phi_{2^{k_{j-1}}}(v) - \phi_{2^{k_j}}(u)\phi_{2^{k_j}}(v) \right)$$
$$+ \phi_{2^{k_j}}(u)(\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(v)$$
$$+ (\phi_{2^{k_j}} - \phi_{2^{k_{j-1}}})(u)\phi_{2^{k_j}}(v). \qquad (2.10)$$

Summing (2.10) 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 (2.7) implies $\int_{\mathbb{R}^2} |\phi(u)\phi(v)| \, du \, dv \lesssim_\lambda C_2$, we obtain

$$\left| \int_{\mathbb{R}^4} F(x + u, y)G(x, y + u)F(x + v, y)G(x, y + v) \right.$$
$$\left. \times (\phi_{2^{k_0}}(u)\phi_{2^{k_0}}(v) - \phi_{2^{k_m}}(u)\phi_{2^{k_m}}(v)) \, dx \, dy \, du \, dv \right| \lesssim_\lambda C_2 \|F\|_{L^4(\mathbb{R}^2)}^2 \|G\|_{L^4(\mathbb{R}^2)}^2 = C_2.$$
$$(2.11)$$

By symmetry of the second and the third terms on the right-hand side of (2.10), it then suffices to bound

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

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

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

By the assumptions (2.6) on ϕ, Proposition 6 gives

$$|\Lambda_{\chi_{2^4}}(F, G)| \lesssim_{\lambda} C_0 C_1. \tag{2.12}$$

Rewrite Λ_{ω} by separating the functions in u and v as

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

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

$$|\Lambda_{\omega}(F, G)| \le \widetilde{\Lambda}_{\omega}(F, G)^{1/2} V(F, G)^{1/2}, \tag{2.13}$$

where for a function ρ we have set

$$\widetilde{\Lambda}_{\rho}(F, G) := \sum_{j=1}^{m} \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} F(x + u, y)G(x, y + u)(\phi * \rho)_{2^{k_j}}(u) \, du \right)^2 dx \, dy.$$

Note that, up to increasing the quantity $\widetilde{\Lambda}_{\omega}(F, G)$ by adding non-negative 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 $\widetilde{\Lambda}_{\omega}(F, G)$, we can write this form as

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

By the assumption (2.7), Proposition 7 implies that

$$\widetilde{\Lambda}_{\omega}(F, G) \lesssim_{\lambda} C_2. \tag{2.14}$$

Inequalities (2.11), (2.12), (2.13), and (2.14) together give a bootstrapping estimate

$$V(F, G) \lesssim_{\lambda} C_2 + C_0 C_1 + C_2^{1/2} V(F, G)^{1/2}.$$

This shows that $V(F, G) \lesssim_{\lambda} C_0 C_1 + C_2$ and hence proves (2.8). $\square$

2.2. *Short variation.* The following two closely related lemmas 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*

$$\left| \int_1^2 t \partial_t(\phi_t(u)) t \partial_t(\overline{\phi_t(v)}) \frac{dt}{t} \right| \leq C_3 (1 + |u + v|)^{-\lambda} (1 + |u - v|)^{-2\lambda} \qquad (2.15)$$

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

$$\left(\sum_{i=-N}^{N} \left\| \|A_t^\phi(F, G)(x, y)\|_{\mathrm{V}_t^2([2^i, 2^{i+1}], \mathbb{C})} \right\|_{\mathrm{L}^2_{(x,y)}(\mathbb{R}^2)}^2 \right)^{1/2} \lesssim_\lambda C_3^{1/2}, \qquad (2.16)$$

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 (2.7) 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*

$$\left(\sum_{i=-N}^{N} \left\| \|A_t^\phi(F, G)(x, y)\|_{\mathrm{V}_t^2([2^i, 2^{i+1}], \mathbb{C})} \right\|_{\mathrm{L}^2_{(x,y)}(\mathbb{R}^2)}^2 \right)^{1/2} \lesssim_\lambda C_2^{1/4} C_3^{1/4}, \qquad (2.17)$$

with the implicit constant independent of N.

Observe that Lemmas 9 and 10 actually establish pointwise short variation estimates. Since we clearly have

$$\|A_t^\phi(F, G)\|_{\mathrm{V}_t^2([2^i, 2^{i+1}], \mathrm{L}^2(\mathbb{R}^2))} \leq \left\| \|A_t^\phi(F, G)(x, y)\|_{\mathrm{V}_t^2([2^i, 2^{i+1}], \mathbb{C})} \right\|_{\mathrm{L}^2_{(x,y)}(\mathbb{R}^2)},$$

these in turn also imply the corresponding norm-variation estimates.

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 13 (in the Appendix A) applied with $a(t) = A_t^\phi(F, G)(x, y)$ for each fixed (x, y), we have

$$\sup_{2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}} \sum_{j=1}^{m} |A_{t_j}^\phi(F, G)(x, y) - A_{t_{j-1}}^\phi(F, G)(x, y)|^2$$

$$\leq \int_1^2 \left(A_{2^i t}^\psi(F, G)(x, y) \right)^2 \frac{dt}{t}.$$

Indeed, this follows from $A_t^\psi(F, G) = -t \partial_t(A_t^\phi(F, G))$ and by rescaling in t. Integrating in x, y and summing over $-N \leq i \leq N$ yields

$$\sum_{i=-N}^{N} \left\| \|A_t^\phi(F, G)(x, y)\|_{\mathrm{V}_t^2([2^i, 2^{i+1}], \mathbb{C})} \right\|_{\mathrm{L}^2_{(x,y)}(\mathbb{R}^2)}^2$$

$$\leq \sum_{i=-N}^{N} \int_{\mathbb{R}^2} \int_1^2 \left(A_{2^i t}^\psi(F, G)(x, y) \right)^2 \frac{dt}{t} \, dx \, dy.$$

Expanding the square on the right-hand side, in order to finish the proof of Lemma 9 we need to bound

$$\sum_{i=-N}^{N} \int_{\mathbb{R}^4} F(x+u, y)G(x, y+u)F(x+v, y)G(x, y+v)$$
$$\times \left(\int_1^2 \psi_{2^i t}(u)\psi_{2^i t}(v)\frac{dt}{t} \right) dx\, dy\, du\, dv. \tag{2.18}$$

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 (2.15). Proposition 7 implies boundedness of (2.18) within an absolute constant times C_3, which yields (2.16). $\qquad\square$

Proof of Lemma 10. Let all the notation and the assumptions be as in the proof of the previous lemma. By Lemma 13 and the Cauchy–Schwarz inequality in x, y, and i, we deduce that

$$\sum_{i=-N}^{N} \big\| \|A_t^\phi(F, G)(x, y)\|_{V_t^2([2^i, 2^{i+1}], \mathbb{C})} \big\|_{L_{(x,y)}^2(\mathbb{R}^2)}^2$$
$$\lesssim \prod_{\rho \in \{\phi, \psi\}} \left(\sum_{i=-N}^{N} \int_{\mathbb{R}^2} \int_1^2 \left(A_{2^i t}^\rho(F, G)(x, y) \right)^2 \frac{dt}{t}\, dx\, dy \right)^{1/2}.$$

By the support assumptions on ϕ, (2.7), and (2.15), Proposition 7 applied twice gives 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 (2.17). $\qquad\square$

Finally, we are ready to deduce Theorem 2 from these lemmas. 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 Lemmas 8 and 9 combined with the standard separation into long and short jumps imply that

$$\big\| A_t^\varphi(F, G) \big\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim_\lambda C_0^{1/2}C_1^{1/2} + C_2^{1/2} + C_3^{1/2} \lesssim_\varphi 1. \tag{2.19}$$

The details can be found for instance in [12] or [21]. 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 (2.19) 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

$$\sum_{k \in \mathbb{Z}} \widehat{\theta}(2^k \xi) = 1 \tag{2.20}$$

for all $0 \neq \xi \in \mathbb{R}$. Then we can write

$$\varphi = c\chi + (\varphi - c\chi) = c\chi + \sum_{k \in \mathbb{Z}} (\varphi - c\chi) * \theta_{2^k}, \tag{2.21}$$

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 (2.21) 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 (2.21). For the part associated with $c\chi$, boundedness follows from (2.19) since χ is Schwartz and $\widehat{\chi}$ is constant near the origin:

$$\|A_t^{c\chi}(F, G)\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 1. \tag{2.22}$$

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

$$\|A_t^{\vartheta}(F, G)\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 2^{-|k|} \tag{2.23}$$

for any $k \in \mathbb{Z}$. By scaling invariance of the left-hand side of (2.23) 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 (2.21). Then, from (2.21), (2.22), (2.23), Minkowski's inequality, and Fatou's lemma, we obtain

$$\|A_t^{\varphi}(F, G)\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 1 + \sum_{k \in \mathbb{Z}} 2^{-|k|} \lesssim 1,$$

which finishes the proof.

In order to verify (2.23), 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,

$$\big\| |\xi|^{\alpha} D^{\beta} \widehat{\vartheta}(\xi) \big\|_{L_{\xi}^{\infty}(\mathbb{R})} \lesssim_{\alpha, \beta} 2^{-|k|}$$

for any $\alpha, \beta \geq 0$. Therefore, $2^{|k|} \vartheta$ satisfies (2.6), (2.7), and (2.15) with the constants independent of k. The estimate (2.23) then follows from (2.19) applied with $\varphi = 2^{|k|} \vartheta$ and by homogeneity.

Let us remark that the same arguments also establish Corollary 3. We simply use Lemma 9, this time to get a short pointwise variation estimate in the same particular case of φ, and then perform decomposition (2.21) of a general Schwartz function.

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

$$\mathbb{1}_{[0,1)} = \mathbb{1}_{[0,1)} * \chi + \sum_{k=-\infty}^{-1} \mathbb{1}_{[0,1)} * \theta_{2^k}. \tag{2.24}$$

By the Plancherel identity, the series in (2.24) 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 §2.3, we obtain

$$\left\| A_t^{\mathbb{1}_{[0,1)} * \chi}(F, G) \right\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 1. \tag{2.25}$$

Now we concentrate on the individual terms in (2.24) 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

$$\left\| A_t^{\widetilde{\theta}}(F, G) \right\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 1. \tag{2.26}$$

By scaling invariance of the left-hand side, (2.26) 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

$$\left\| A_t^{\vartheta}(F, G) \right\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 2^{k/8}. \tag{2.27}$$

Once this is shown, by scaling invariance of the left-hand side, the estimate (2.27) 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 (2.24), (2.25), (2.26), (2.27), Minkowski's inequality, and Fatou's lemma, we finally obtain

$$\left\| A_t^{\mathbb{1}_{[0,1)}}(F, G) \right\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim 1 + \sum_{k \leq -1} (2^k + 2^{k/8}) \lesssim 1.$$

In order to see (2.27), 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

$$|\vartheta(u)\vartheta(v)| \lesssim_{\lambda,\nu} 2^{k(2-\lambda)}(1 + |u|)^{-\lambda/2}(1 + |v|)^{-\lambda/2}(1 + |u - v|)^{-\nu}, \tag{2.28}$$

$$\left| \int_1^2 t\partial_t(\vartheta_t(u))t\partial_t(\vartheta_t(v))\frac{dt}{t} \right| \lesssim_{\lambda,\nu} 2^{k(1-\lambda)}(1 + |u|)^{-\lambda/2}(1 + |v|)^{-\lambda/2}(1 + |u - v|)^{-\nu}. \tag{2.29}$$

We have already commented on how bounds of this form with $\nu = 3\lambda$ transform into bounds (2.7) and (2.15). Once these two estimates are verified, the separation into short and long jumps together with (2.9) and Lemma 10, which require (2.28) and (2.29) to hold with $\lambda > 1$, give

$$\left\| A_t^{\vartheta}(F, G) \right\|_{V_t^2((0,\infty), L^2(\mathbb{R}^2))} \lesssim_\lambda C_2^{1/2} + C_2^{1/4} C_3^{1/4}$$

with $C_2 \sim 2^{k(2-\lambda)}$ and $C_3 \sim 2^{k(1-\lambda)}$. Choosing $\lambda = 5/4$, we obtain (2.27).

Proof of (2.28). By the rapid decay of $\widetilde{\theta}$, we have

$$|\widetilde{\theta}(u - 2^{-k})\widetilde{\theta}(v - 2^{-k})| \lesssim_{\lambda,\nu} (1 + |u - 2^{-k}|)^{-\lambda/2-\nu}(1 + |v - 2^{-k}|)^{-\lambda/2-\nu}$$
$$\leq (1 + |u - 2^{-k}|)^{-\lambda/2}(1 + |v - 2^{-k}|)^{-\lambda/2}(1 + |u - v|)^{-\nu},$$

where we used $|u - v| \leq |u - 2^{-k}| + |v - 2^{-k}|$. From

$$(1 + |u - 2^{-k}|)^{-\lambda/2} \leq (1 + |u|)^{-\lambda/2}(1 + 2^{-k})^{\lambda/2} \lesssim_{\lambda} (1 + |u|)^{-\lambda/2}2^{-k\lambda/2},$$

we then conclude (2.28). $\qquad\square$

Proof of (2.29). 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 (2.29) corresponding to $(s\vartheta'(s))_t$, that is,

$$\left|\int_1^2 (u\vartheta'(u))_t(v\vartheta'(v))_t \frac{dt}{t}\right| \lesssim_{\lambda,\nu} 2^{k(1-\lambda)}(1 + |u|)^{-\lambda/2}(1 + |v|)^{-\lambda/2}(1 + |u - v|)^{-\nu}. \tag{2.30}$$

The analogous inequalities corresponding to the other terms are treated in the same manner. To see (2.30), we first observe that

$$(s\vartheta'(s))_t = (s2^k\theta(s - 2^{-k}))_t = st^{-1}2^k\theta_t(s - t2^{-k})$$

and bound $|\theta_t(s)| \lesssim_{\lambda,\nu} (1 + |s|)^{-\lambda/2-\nu-1}$ using $t \in [1, 2]$. Then we estimate

$$\left|\int_1^2 u\theta_t(u - t2^{-k})v\theta_t(v - t2^{-k})\frac{dt}{t^3}\right|$$
$$\lesssim_{\lambda,\nu} |uv|\int_1^2 \left((1 + |u - t2^{-k}|)(1 + |v - t2^{-k}|)\right)^{-\lambda/2-\nu-1} dt.$$

By the triangle inequality $|u - v| \leq |u - t2^{-k}| + |v - t2^{-k}|$ and the Cauchy–Schwarz inequality in t, this is bounded by

$$(1 + |u - v|)^{-\nu}|u|\left(\int_1^2 (1 + |u - t2^{-k}|)^{-\lambda-2} dt\right)^{1/2}|v|\left(\int_1^2 (1 + |v - t2^{-k}|)^{-\lambda-2} dt\right)^{1/2}.$$

Now, if $|u| \leq 2^{-k+2}$, then we estimate

$$(1 + |u|)^{\lambda/2}|u|\left(\int_1^2 (1 + |u - t2^{-k}|)^{-\lambda-2} dt\right)^{1/2}$$
$$\leq (1 + |u|)^{\lambda/2+1}\left(\int_{-\infty}^{\infty} (1 + |u - t2^{-k}|)^{-\lambda-2} dt\right)^{1/2}$$
$$\lesssim_{\lambda} 2^{k/2}(1 + |u|)^{\lambda/2+1} \lesssim_{\lambda} 2^{k(-1-\lambda)/2},$$

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

$$(1 + |u|)^{\lambda/2}|u|\left(\int_1^2 (1 + |u - t2^{-k}|)^{-\lambda-2} dt\right)^{1/2}$$
$$\lesssim_{\lambda} (1 + |u|)^{\lambda/2+1}(1 + |u|)^{-\lambda/2-1} = 1 \leq 2^{k(-1-\lambda)/2}.$$

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 (2.30). $\qquad\square$

3. *Proof of Proposition 6*

Let us rewrite the form (2.2) from Proposition 6 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

$$x' = x + y + u, \quad y' = x + y + v,$$
$$\widetilde{F}(y, x') := F(x' - y, y), \quad \widetilde{G}(x, x') := G(x, x' - x).$$

Note that we still have $\|\widetilde{F}\|_{L^4(\mathbb{R}^2)} = \|\widetilde{G}\|_{L^4(\mathbb{R}^2)} = 1$. Omitting the tildes for notational simplicity, it then suffices to show the inequality

$$\left| \sum_{j=1}^m \sum_{l=k_{j-1}}^{k_j-1} \int_{\mathbb{R}^4} F(y, x')G(x, x')F(y, y')G(x, y') \right.$$

$$\left. \times \vartheta_{2^{k_j}}(x' - x - y)\psi_{2^l}(y' - x - y) \, dx \, dy \, dx' \, dy' \right| \lesssim 1.$$

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 non-negative 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 the set $[-2, -2^{-4}] \cup [2^{-4}, 2]$. 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

$$\vartheta_{2^{k_j}}(x' - x - y)\psi_{2^l}(y' - x - y)$$
$$= \int_{\mathbb{R}^2} \widehat{\vartheta}(2^{k_j}\xi)e^{2\pi ix'\xi}\widehat{\psi}(2^l\eta)e^{2\pi iy'\eta}\widehat{\omega}(2^l(-\xi - \eta))$$
$$\times e^{2\pi ix(-\xi-\eta)}\widehat{\omega}(2^l(-\xi - \eta))e^{2\pi iy(-\xi-\eta)} \, d\xi \, d\eta.$$

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

$$\int_{\mathbb{R}^2} \vartheta_{2^{k_j}}(x' - p - q)\psi_{2^l}(y' - p - q)\omega_{2^l}(x - p)\omega_{2^l}(y - q)\, dp\, dq$$

and the object we need to bound is

$$\sum_{j=1}^{m} \sum_{l=k_{j-1}}^{k_j-1} \int_{\mathbb{R}^6} F(y, x')G(x, x')F(y, y')G(x, y')\vartheta_{2^{k_j}}(x' - p - q)$$
$$\times \psi_{2^l}(y' - p - q)\omega_{2^l}(x - p)\omega_{2^l}(y - q)\, dx\, dy\, dx'\, dy'\, dp\, dq. \qquad (3.3)$$

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

$$\sum_{j=1}^{m} \sum_{l=k_{j-1}}^{k_j-1} \int_{\mathbb{R}^4} \left(\int_{\mathbb{R}} F(y, x')F(y, y')\omega_{2^l}(y - q)\, dy \right)$$
$$\times \left(\int_{\mathbb{R}} G(x, x')G(x, y')\omega_{2^l}(x - p)\, dx \right)$$
$$\times \vartheta_{2^{k_j}}(x' - p - q)\psi_{2^l}(y' - p - q)\, dx'\, dy'\, dp\, dq.$$

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

$$\Gamma(F)^{1/2}\Gamma(G)^{1/2}, \qquad (3.4)$$

where we have denoted

$$\Gamma(F) := \sum_{j=1}^{m} \sum_{l=k_{j-1}}^{k_j-1} \int_{\mathbb{R}^4} \left(\int_{\mathbb{R}} F(y, x')F(y, y')\omega_{2^l}(y - q)\, dy \right)^2$$
$$\times |\vartheta|_{2^{k_j}}(x' - p)|\psi|_{2^l}(y' - p)\, dx'\, dy'\, dp\, dq.$$

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

$$\int_{\mathbb{R}} |\vartheta|_{2^{k_j}}(x' - p)|\psi|_{2^l}(y' - p)\, dp \lesssim_\lambda 2^{-k_j}(1 + 2^{-k_j}|x' - y'|)^{-\lambda}. \qquad (3.5)$$

This fact can be shown along the lines of [**38**, 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

$$\|\vartheta_{2^{k_j}}\|_{L^\infty(\mathbb{R})}\|\psi_{2^l}\|_{L^1(\mathbb{R})} \lesssim_\lambda \|\vartheta\|_{L^\infty(\mathbb{R})}\|\psi\|_{L^1(\mathbb{R})}2^{-k_j}(1 + 2^{-k_j}|x' - y'|)^{-\lambda}.$$

If $|x' - y'| \geq 2^{k_j+1}/(\lambda - 1)$, then let us denote by c the mid point of x' and y'. Without loss of generality, we may assume that $x' < c < y'$. We split the integral as $\int_{\mathbb{R}} = \int_{-\infty}^{c} + \int_{c}^{\infty}$ and estimate it by

$$\|\vartheta\|_{L^1(\mathbb{R})} 2^{-l}(1 + 2^{-l}|y' - c|)^{-\lambda} + 2^{-k_j}(1 + 2^{-k_j}|x' - c|)^{-\lambda}\|\psi\|_{L^1(\mathbb{R})}. \tag{3.6}$$

Since $|x' - c| = |y' - c| = |x' - y'|/2, l \leq k_j$,

$$2^{-l-1}|x' - y'| \geq 2^{-k_j-1}|x' - y'| \geq (\lambda - 1)^{-1},$$

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

$$(\|\vartheta\|_{L^1(\mathbb{R})} + \|\psi\|_{L^1(\mathbb{R})}) 2^{-k_j}(1 + 2^{-k_j-1}|x' - y'|)^{-\lambda}.$$

It remains to note that $\|\vartheta\|_{L^\infty(\mathbb{R})} \leq 1$, $\|\vartheta\|_{L^1(\mathbb{R})} \lesssim_\lambda 1$, and $\|\psi\|_{L^1(\mathbb{R})} \lesssim_\lambda 1$, 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 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 $\sigma(0) = \lambda^{-1}$ 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

$$\Gamma(F) \lesssim_\lambda \int_1^\infty \left(\sum_{j=1}^m \sum_{l=k_j-1}^{k_j-1} \int_{\mathbb{R}^5} F(y, x')F(x, x')F(y, y')F(x, y') \right.$$

$$\left. \times g_{\alpha 2^{k_j}}(x' - y')\omega_{2^l}(x - q)\omega_{2^l}(y - q) \, dx \, dy \, dx' \, dy' \, dq \right)\alpha^{-\lambda} \, d\alpha. \tag{3.9}$$

Integrating in q, summing in l, and using $\widehat{\omega}(\xi)^2 = \sum_{i=-2}^{3}(\widehat{\chi}(2^i\xi) - \widehat{\chi}(2^{i+1}\xi))$, 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

$$\sum_{i=-2}^{3} \sum_{j=1}^m \int_{\mathbb{R}^4} F(y, x')F(x, x')F(y, y')F(x, y')$$

$$\times g_{\alpha 2^{k_j}}(x' - y')(\chi_{2^{k_j-1+i}} - \chi_{2^{k_j+i}})(x - y) \, dx \, dy \, dx' \, dy'.$$

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 $-2 \leq i \leq 3$. For two functions $\tilde{\rho}, \rho \in \mathcal{S}(\mathbb{R})$, define

$$\Theta_{\tilde{\rho},\rho}(F) := \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(y, x') F(x, x') F(y, y') F(x, y')$$
$$\times \tilde{\rho}_{2^{k_j}}(x' - y')(\rho_{2^{k_{j-1}}} - \rho_{2^{k_j}})(x - y)\, dx\, dy\, dx'\, dy'.$$

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*

$$\Theta_{g_\alpha,\rho}(F) \lesssim_\rho \|F\|_{L^4(\mathbb{R}^2)}^4, \tag{3.10}$$

where $g(s) = e^{-\pi s^2}$.

Proof of Lemma 11. Once again we normalize F as in (2.1). The first step is an application of the telescoping identity. If we denote

$$\widetilde{\Theta}_{\tilde{\rho},\rho}(F) := \sum_{j=1}^{m} \int_{\mathbb{R}^4} F(y, x') F(x, x') F(y, y') F(x, y')$$
$$\times (\tilde{\rho}_{2^{k_{j-1}}} - \tilde{\rho}_{2^{k_j}})(x' - y')\rho_{2^{k_{j-1}}}(x - y)\, dx\, dy\, dx'\, dy'$$

and for $t > 0$ define the single-scale quantity

$$\Xi_{\tilde{\rho},\rho,t}(F) := \int_{\mathbb{R}^4} F(y, x') F(x, x') F(y, y') F(x, y') \tilde{\rho}_t(x' - y')\rho_t(x - y)\, dx\, dy\, dx'\, dy',$$

then we have

$$\Theta_{\tilde{\rho},\rho}(F) + \widetilde{\Theta}_{\tilde{\rho},\rho}(F) = \Xi_{\tilde{\rho},\rho,2^{k_0}}(F) - \Xi_{\tilde{\rho},\rho,2^{k_m}}(F), \tag{3.11}$$

$$\Xi_{\tilde{\rho},\rho,t}(F) \leq \|\tilde{\rho}\|_{L^1(\mathbb{R})} \|\rho\|_{L^1(\mathbb{R})}. \tag{3.12}$$

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

$$\int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} F(x - v, x') F(x, x') F(x - v, x' - u) F(x, x' - u)\, dx\, dx' \right) \tilde{\rho}_t(u)\rho_t(v)\, du\, dv,$$

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

$$|\Theta_{g_\alpha,\rho}(F)| \leq |\Xi_{g_\alpha,\rho,2^{k_0}}(F)| + |\Xi_{g_\alpha,\rho,2^{k_m}}(F)| + |\widetilde{\Theta}_{g_\alpha,\rho}(F)| \lesssim_\rho 1 + |\widetilde{\Theta}_{g_\alpha,\rho}(F)|.$$

Therefore, it remains to estimate $|\widetilde{\Theta}_{g_\alpha,\rho}(F)|$.

By the fundamental theorem of calculus, we rewrite $\widetilde{\Theta}_{g_\alpha,\rho}(F)$ as

$$\widetilde{\Theta}_{g_\alpha,\rho}(F) = \sum_{j=1}^{m} \int_{2^{k_{j-1}}}^{2^{k_j}} \int_{\mathbb{R}^4} F(y, x') F(x, x') F(y, y') F(x, y')$$
$$\times (-t\partial_t(g_{\alpha t}(x' - y'))) \rho_{2^{k_{j-1}}}(x - y)\, dx\, dy\, dx'\, dy'\, \frac{dt}{t}.$$

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

$$\sum_{j=1}^{m} \int_{2^{k_{j-1}}}^{2^{k_j}} \int_{\mathbb{R}^3} \left(\int_{\mathbb{R}} F(y, x')F(x, x')h_{\alpha t}(x' - p)\, dx' \right)^2 \rho_{2^{k_{j-1}}}(x - y)\, dx\, dy\, dp\, \frac{dt}{t}. \tag{3.14}$$

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 that

$$\left| \widetilde{\Theta}_{g_\alpha,\rho}(F) \right| \lesssim_\rho \int_1^\infty \widetilde{\Theta}_{g_\alpha,g_\beta}(F)\beta^{-\lambda}\, d\beta.$$

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,

$$\widetilde{\Theta}_{g_\alpha,g_\beta}(F) \le \Xi_{g_\alpha,g_\beta,2^{k_0}}(F) - \Xi_{g_\alpha,g_\beta,2^{k_m}}(F) \le 2\|g\|_{L^1(\mathbb{R})}^2 \lesssim 1.$$

This finishes the proof of Lemma 11. $\qquad\square$

4. *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 §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 (4.8) 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 analogous to those from §3 may be applied. Due to the vanishing along $\eta = -\xi$, we are able to sum over all such regions.

We start with a lemma which considers multiplier symbols supported away from the axis $\eta = -\xi$. It will be applied several times in the proof of Proposition 7.

LEMMA 12. *Let $\lambda > 1$, $t > 0$ and let $\Phi \in \mathcal{S}(\mathbb{R}^2)$ be such that*

$$|\Phi(u, v)| \leq (1 + |u + v|)^{-\lambda} t (1 + t|u - v|)^{-\lambda}$$

for all $u, v \in \mathbb{R}$. Moreover, assume that $2^{-2} \leq |\xi + \eta| \leq 1$ 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 (2.4).

Proof of Lemma 12. Our aim is to reduce Lemma 12 to Lemma 11 from the previous section. Let χ and ω be the functions as in §3. Then $\widehat{\omega}(\xi + \eta)$ equals 1 on $\{\xi + \eta : (\xi, \eta) \in \text{supp}(\widehat{\Phi})\}$, 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 §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 $\Gamma(F)$,

$$\sum_{j=-N}^{N} \int_{\mathbb{R}^6} F(y, x') F(x, x') F(y, y') F(x, y')$$

$$\times |\Phi|_{2^j}(x' - p, y' - p)\omega_{2^j}(x - q)\omega_{2^j}(y - q) \, dx \, dy \, dx' \, dy' \, dp \, dq.$$

Using the decay assumption on Φ, we obtain

$$\int_{\mathbb{R}} |\Phi|(x' - p, y' - p) \, dp \leq \int_{\mathbb{R}} t(1 + t|x' - y'|)^{-\lambda}(1 + |x' + y' - 2p|)^{-\lambda} \, dp$$

$$\lesssim_{\lambda} t(1 + t|x' - y'|)^{-\lambda}.$$

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

$$\sum_{j=-N}^{N} \int_{\mathbb{R}^4} F(y, x') F(x, x') F(y, y') F(x, y') g_{\alpha t 2^j}(x' - y')$$

$$\times (\chi_{2^{j+i}} - \chi_{2^{j+i+1}})(x - y) \, dx \, dy \, dx' \, dy'$$

uniformly in $\alpha, t \in (0, \infty)$ and for each fixed $-2 \leq i \leq 3$. Such an estimate follows from the particular case of Lemma 11 when $\rho = \chi_{2^{i+1}}$ and $k_j = j$. $\square$

Now we are ready to proceed with the proof of Proposition 7. 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 (2.20). Then $\int_0^\infty \widehat{\theta}(t\tau)(dt/t)$ is the same constant for all $0 \neq \tau \in \mathbb{R}$ and up to that constant $\widehat{K}(\xi, \eta)$ equals

$$\int_0^\infty \widehat{K}(\xi, \eta)\widehat{\theta}(t|(\xi, \eta)|)\frac{dt}{t} = \int_0^\infty \widehat{K^{(t)}}(t(\xi, \eta))\frac{dt}{t} \tag{4.1}$$

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

$$([-1, -2^{-9}] \cup [2^{-9}, 1])^2. \tag{4.2}$$

Let ϑ be such that $\widehat{\vartheta}$ is a smooth non-negative 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 (4.2) 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 that

$$K^{(t)}(u, v) = \int_{\mathbb{R}^2} K^{(t)}(a, b)\vartheta(u - a)\vartheta(v - b) \, da \, db. \tag{4.3}$$

Using (4.1) and (4.3), the form from Proposition 7 can be rewritten as

$$\int_{\mathbb{R}^2} \int_0^\infty K^{(t)}(a, b) \int_{\mathbb{R}^4} F(x + u, y)G(x, y + u)F(x + v, y)G(x, y + v)$$
$$\times \vartheta_t(u - ta)\vartheta_t(v - tb) \, dx \, dy \, du \, dv \frac{dt}{t} \, da \, db. \tag{4.4}$$

Observe that for $\kappa \in \mathcal{S}(\mathbb{R}^2)$ defined by $\widehat{\kappa}(\xi, \eta) = \widehat{\theta}(|(\xi, \eta)|)$ we have

$$K^{(t)}(a, b) = \sum_{j=-N}^N \int_{\mathbb{R}^2} \Phi_{2^j/t}(a - x, b - y)\kappa(x, y) \, dx \, dy$$

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 (2.3) transformed into (2.5) and the rapid decay of κ, it follows that

$$\left| K^{(t)}(a, b) \right| \lesssim_\lambda \int_{\mathbb{R}^2} (1 + |a - x|)^{-\lambda/2}(1 + |b - y|)^{-\lambda/2}$$
$$\times (1 + |a - x - b + y|)^{-\lambda}|\kappa(x, y)| \, dx \, dy$$
$$\lesssim_\lambda (1 + |a|)^{-\lambda/2}(1 + |b|)^{-\lambda/2}(1 + |a - b|)^{-\lambda}.$$

Taking absolute values in (4.4) 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 (4.4) by

$$\int_{\mathbb{R}^2} (1 + |a|)^{-\lambda/2}(1 + |b|)^{-\lambda/2}(1 + |a - b|)^{-\lambda}$$
$$\times \int_0^\infty \int_{\mathbb{R}^2} |I(x, y, a, t)I(x, y, b, t)| \, dx \, dy \frac{dt}{t} \, da \, db.$$

Next, we apply the Cauchy–Schwarz inequality in x, y and t, which gives

$$\int_{\mathbb{R}^2} (1 + |a - b|)^{-\lambda}(1 + |a|)^{-\lambda/2}\left(\int_0^\infty \int_{\mathbb{R}^2} I(x, y, a, t)^2 \, dx \, dy \frac{dt}{t}\right)^{1/2}$$
$$\times (1 + |b|)^{-\lambda/2}\left(\int_0^\infty \int_{\mathbb{R}^2} I(x, y, b, t)^2 \, dx \, dy \frac{dt}{t}\right)^{1/2} da \, db. \qquad (4.5)$$

If we denote

$$J(a) := (1 + |a|)^{-\lambda/2}\left(\int_0^\infty \int_{\mathbb{R}^2} I(x, y, a, t)^2 \, dx \, dy \frac{dt}{t}\right)^{1/2},$$

the expression (4.5) can be rewritten as

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} (1 + |a - b|)^{-\lambda} J(a) \, da\right) J(b) \, db.$$

Applying the Cauchy–Schwarz inequality in b, we obtain

$$\left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} (1 + |a - b|)^{-\lambda} J(a) \, da\right)^2 db\right)^{1/2}\left(\int_{\mathbb{R}} J(b)^2 \, db\right)^{1/2}. \qquad (4.6)$$

Note that the integral in a is the convolution of J with $s \mapsto (1 + |s|)^{-\lambda}$. By Young's convolution inequality from $L^1(\mathbb{R}) \times L^2(\mathbb{R})$ to $L^2(\mathbb{R})$, the expression (4.6) is bounded by a constant multiple of

$$\|J\|_{L^2(\mathbb{R})}^2 \le \int_{\mathbb{R}} (1 + a^2)^{-\lambda/2} \int_0^\infty \int_{\mathbb{R}^2} I(x, y, a, t)^2 \, dx \, dy \frac{dt}{t} \, da.$$

Expanding I, this equals

$$\int_{\mathbb{R}} (1 + a^2)^{-\lambda/2} \int_0^\infty \int_{\mathbb{R}^4} F(x + u, y)G(x, y + u)F(x + v, y)G(x, y + v)$$
$$\times \vartheta_t(u - ta)\vartheta_t(v - ta) \, dx \, dy \, du \, dv \frac{dt}{t} \, da. \qquad (4.7)$$

Observe that this form is associated with the multiplier symbol

$$M(\xi, \eta) := \int_0^\infty \widehat{\vartheta}(t\xi)\widehat{\vartheta}(t\eta)\widehat{\rho}(t(\xi + \eta)) \frac{dt}{t}, \qquad (4.8)$$

where we have denoted

$$\rho(s) := (1 + s^2)^{-\lambda/2}. \qquad (4.9)$$

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 $0 \ne \xi \in \mathbb{R}$. Now we write

$$M(\xi, \eta) = M(1, -1) + \big(M(\xi, \eta) - M(1, -1)\big)$$

and split the form (4.7) into the two corresponding parts. The part associated with the constant multiplier yields $M(1, -1)$ times

$$\int_{\mathbb{R}^4} F(x + u, y)G(x, y + u)F(x + v, y)G(x, y + v)\delta_{(0,0)}(u, v) \, dx \, dy \, du \, dv$$
$$= \int_{\mathbb{R}^4} F(x, y)^2 G(x, y)^2 \, dx \, dy \le \|F\|_{L^4(\mathbb{R}^2)}^2 \|G\|_{L^4(\mathbb{R}^2)}^2 = 1,$$

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

$$M_0(\xi, \eta) = \sum_{k \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} M_0(\xi, \eta)\widehat{\theta}(2^{j+k}(\xi - \eta))\widehat{\theta}(2^j(\xi + \eta)). \qquad (4.10)$$

If we denote

$$m^{(k)}(\xi, \eta) := M_0(\xi, \eta)\widehat{\theta}(2^k(\xi - \eta))\widehat{\theta}(\xi + \eta)$$

and

$$m(\xi, \eta) := \sum_{k \geq 0} m^{(k)}(\xi, \eta) = M_0(\xi, \eta)\left(\sum_{k \geq 0} \widehat{\theta}(2^k(\xi - \eta))\right)\widehat{\theta}(\xi + \eta),$$

and split the summation in (4.10) 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

$$\sum_{j \in \mathbb{Z}} m(2^j(\xi, \eta)). \qquad (4.11)$$

Observe that m is compactly supported in the strip $2^{-2} \leq |\xi + \eta| \leq 1$. Moreover, we have

$$|\widecheck{m}(u, v)| \lesssim_\lambda (1 + |u + v|)^{-\lambda}(1 + |u - v|)^{-2}. \qquad (4.12)$$

Indeed, this estimate can be seen by bounding the inverse Fourier transform of

$$(\xi, \eta) \mapsto m(\xi, \eta) + M(1, -1)\phi(\xi, \eta), \qquad (4.13)$$

where we have set

$$\phi(\xi, \eta) := \left(\sum_{k \geq 0} \widehat{\theta}(2^k(\xi - \eta))\right)\widehat{\theta}(\xi + \eta).$$

Therefore, the inverse Fourier transform of (4.13) 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 $\widecheck{\phi}$ 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

$$\left| \int_{\mathbb{R}} \rho(a)\vartheta_t(u - ta)\vartheta_t(v - ta)\, da \right|$$
$$\lesssim_{\vartheta,\lambda} (1 + |u - v|)^{-2} \int_{\mathbb{R}} \rho(a)(1 + |u + v - 2a|)^{-2\lambda}\, da$$
$$\lesssim_\lambda (1 + |u - v|)^{-2}(1 + |u + v|)^{-\lambda},$$

which in turn implies (4.12). Boundedness of the form associated with (4.11) now follows from Lemma 12 applied with $\Phi = \widecheck{m}$ and by letting $N \to \infty$.

It remains to consider the form associated with the symbol

$$\sum_{k<0} \sum_{j\in\mathbb{Z}} m^{(k)}(2^j(\xi, \eta)). \qquad (4.14)$$

Note that $m^{(k)}$ is supported in the strip $2^{-2} \le |\xi + \eta| \le 1$ for each k. To estimate the form associated with (4.14), it now suffices to show that for each $k < 0$ we have

$$|\widetilde{m^{(k)}}(u, v)| \lesssim_\lambda 2^{k(\lambda-1)}(1 + |u + v|)^{-2}2^{-k}(1 + 2^{-k}|u - v|)^{-2}, \qquad (4.15)$$

with the implicit constant independent of k. Once we have that, boundedness of the form associated with the symbol in (4.14) for a fixed k follows from Lemma 12 applied with $\Phi = \widetilde{m^{(k)}}$ 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 (4.15) 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 $0 \le n \le 2$, $|\alpha| \sim 1$, and $0 < |\beta| \le 1$, we have

$$\left|\partial_\beta \partial_\alpha^n\big(M_0(\alpha + \beta, \beta - \alpha)\big)\right| \lesssim_\lambda |\beta|^{\lambda-2}, \quad \left|\partial_\beta^2 \partial_\alpha^n\big(M_0(\alpha + \beta, \beta - \alpha)\big)\right| \lesssim_\lambda |\beta|^{\lambda-3}. \qquad (4.16)$$

For now let us assume that the estimates in (4.16) hold. For $0 \le n \le 2$, 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 $|\alpha| \sim 1$, $0 < |\beta| \le 1$, and $0 \le n \le 2$, the first estimate in (4.16) implies that

$$\left|\partial_\alpha^n\big(M_0(\alpha + \beta, \beta - \alpha)\big)\right| = \left|\int_0^\beta \partial_2\mu^{(n)}(\alpha, \gamma)\, d\gamma\right| \lesssim_\lambda |\beta|^{\lambda-1}. \qquad (4.17)$$

The estimates (4.16) and (4.17) together imply that for any $0 \le l, n \le 2$ one has

$$\left|\partial_\beta^l \partial_\alpha^n\big(M_0(\alpha + 2^k\beta, 2^k\beta - \alpha)\widehat{\theta}(2\alpha)\widehat{\theta}(2\beta)\big)\right| \lesssim_\lambda 2^{k(\lambda-1)},$$

which is by the homogeneity of M_0 equivalent to

$$\left|\partial_\beta^l \partial_\alpha^n\big(m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha)\big)\right| \lesssim_\lambda 2^{k(\lambda-1)}. \qquad (4.18)$$

We proceed by verifying (4.15). 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

$$\widetilde{m^{(k)}}(u, v) = \int_{\mathbb{R}^2} m^{(k)}(\xi, \eta)e^{2\pi i(u\xi + v\eta)}\, d\xi\, d\eta$$

$$= 2^{-k+1}\int_{\mathbb{R}^2} m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha)e^{2\pi i(2^{-k}(u-v)\alpha + (u+v)\beta)}\, d\alpha\, d\beta.$$

If $|u - v| \le 2^k$ and $|u + v| \le 1$, then we bound

$$|\widetilde{m^{(k)}}(u, v)| \lesssim 2^{-k}\|m^{(k)}\|_{L^\infty(\mathbb{R}^2)} \lesssim_\lambda 2^{k(\lambda-1)}2^{-k},$$

which implies (4.15) in this case. Here we used (4.18) to control the L^∞ norm and observed that

$$\big\{(\alpha, \beta) : m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha) \neq 0\big\} \subseteq ([-2^{-1}, -2^{-3}] \cup [2^{-3}, 2^{-1}])^2. \quad (4.19)$$

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

$$2^{-k}(2^{-k}|u - v|)^{-2}|u + v|^{-2}\bigg| \int_{\mathbb{R}^2} \partial_\beta^2 \partial_\alpha^2 (m^{(k)}(2^{-k}\alpha + \beta, \beta - 2^{-k}\alpha))$$

$$\times e^{2\pi i(2^{-k}(u-v)\alpha + (u+v)\beta)} \, d\alpha \, d\beta\bigg|.$$

Together with (4.18) and (4.19), this shows (4.15) 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 (4.16) and for that we need

$$|\widehat{\rho}'(\xi)| \lesssim_\lambda |\xi|^{\lambda-2}, \quad |\widehat{\rho}''(\xi)| \lesssim_\lambda |\xi|^{\lambda-3} \quad (4.20)$$

for $|\xi| \leq 1$, where ρ is our very particular choice of function (4.9). The following formulae that hold for $\xi > 0$ can be found using [**40**] or [**1**]:

$$\widehat{\rho}(\xi) = 2\pi^{\lambda/2}\xi^{(\lambda-1)/2}\mathcal{K}_{(1-\lambda)/2}(2\pi\xi)/ \Gamma(\lambda/2),$$

$$\widehat{\rho}'(\xi) = -4\pi^{1+\lambda/2}\xi^{(\lambda-1)/2}\mathcal{K}_{(\lambda-3)/2}(2\pi\xi)/ \Gamma(\lambda/2),$$

$$\widehat{\rho}''(\xi) = 4\pi^{1+\lambda/2}\xi^{(\lambda-3)/2}\big(2\pi\xi\mathcal{K}_{(\lambda-5)/2}(2\pi\xi) - \mathcal{K}_{(\lambda-3)/2}(2\pi\xi)\big)/ \Gamma(\lambda/2),$$

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

$$\mathcal{K}_\alpha(z) = \frac{\pi}{2\sin(\alpha\pi)}\bigg(\sum_{n=0}^{\infty} \frac{1}{n!\Gamma(n - \alpha + 1)}\Big(\frac{z}{2}\Big)^{2n-\alpha} - \sum_{n=0}^{\infty} \frac{1}{n!\Gamma(n + \alpha + 1)}\Big(\frac{z}{2}\Big)^{2n+\alpha}\bigg).$$

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 (4.20). 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 (4.20) and the product rule gives (4.16). $\qquad \square$

5. *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 $F, G \in L^4(\mathbb{R}^2)$, denote $A_t(F, G) := A_t^{\mathbb{1}_{[0,1)}}(F, G)$, so that

$$A_t(F, G)(x, y) = \frac{1}{t} \int_{[0,t)} F(x + s, y)G(x, y + s) \, ds$$

$$= \frac{1}{t} \int_{[x+y, x+y+t)} F(u - y, y)G(x, u - x) \, du. \quad (5.1)$$

Applying Theorem 2 to the scales $t_j = n_j$ and arbitrary functions F, $G \in L^4(\mathbb{R}^2)$ normalized as in (2.1) gives

$$\sum_{j=1}^{m} \|A_{n_j}(F, G) - A_{n_{j-1}}(F, G)\|_{L^2(\mathbb{R}^2)}^2 \lesssim 1. \tag{5.2}$$

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

$$\widetilde{A}_n(\widetilde{F}, \widetilde{G})(k, l) = \frac{1}{n} \sum_{\substack{i \in \mathbb{Z} \\ k+l \leq i \leq k+l+n-1}} \widetilde{F}(i - l, l) \, \widetilde{G}(k, i - k). \tag{5.3}$$

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 $F, G : \mathbb{R}^2 \to \mathbb{R}$ as

$$F(x, y) := \sum_{i,l \in \mathbb{Z}} \widetilde{F}(i - l, l) \, \mathbb{1}_{[i,i+1)}(x + y) \, \mathbb{1}_{[l,l+1)}(y),$$

$$G(x, y) := \sum_{i,k \in \mathbb{Z}} \widetilde{G}(k, i - k) \, \mathbb{1}_{[k,k+1)}(x) \, \mathbb{1}_{[i,i+1)}(x + y).$$

Note that F and G are constant on certain skew parallelograms of area 1 and $\|F\|_{L^4(\mathbb{R}^2)} = \|G\|_{L^4(\mathbb{R}^2)} = 1$ as well. Splitting the integral (5.1) into the pieces over $i \leq u < i + 1$, we get

$$A_n(F, G)(k + \alpha, l + \beta) = \frac{1}{n} \sum_{i \in \mathbb{Z}} a_i \widetilde{F}(i - l, l) \widetilde{G}(k, i - k) \tag{5.4}$$

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 (5.4) with (5.3), 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

$$\big\| A_n(F, G)(k + \alpha, l + \beta) - \widetilde{A}_n(\widetilde{F}, \widetilde{G})(k, l) \big\|_{\ell_{(k,l)}^2(\mathbb{Z}^2)} \leq \frac{4}{n}.$$

Observe that this estimate is uniform in $\alpha, \beta \in [0, 1)$. Consequently,

$$\Big| \|A_{n_j}(F, G)(k + \alpha, l + \beta) - A_{n_{j-1}}(F, G)(k + \alpha, l + \beta)\|_{\ell_{(k,l)}^2(\mathbb{Z}^2)}$$

$$- \|\widetilde{A}_{n_j}(\widetilde{F}, \widetilde{G}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}, \widetilde{G})\|_{\ell^2(\mathbb{Z}^2)} \Big| \leq \frac{8}{n_{j-1}},$$

so, taking the $L^2([0, 1)^2)$ norm in (α, β),

$$\left| \|A_{n_j}(F, G) - A_{n_{j-1}}(F, G)\|_{L^2(\mathbb{R}^2)} - \|\widetilde{A}_{n_j}(\widetilde{F}, \widetilde{G}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}, \widetilde{G})\|_{\ell^2(\mathbb{Z}^2)} \right| \leq \frac{8}{n_{j-1}}.$$

Combining this with (5.2) and using $\sum_{j=1}^{m} n_{j-1}^{-2} \leq \sum_{n=1}^{\infty} n^{-2} \lesssim 1$, we conclude that

$$\sum_{j=1}^{m} \|\widetilde{A}_{n_j}(\widetilde{F}, \widetilde{G}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}, \widetilde{G})\|_{\ell^2(\mathbb{Z}^2)}^2 \lesssim 1.$$

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 4.

Finally, we transfer to the measure-preserving system $(X, \mathcal{F}, \mu, S, T)$. Let $f, g \in L^4(X)$ be normalized by $\|f\|_{L^4(X)} = \|g\|_{L^4(X)} = 1$. Take a point $x \in X$ and fix a positive integer $N \geq n_m$. The function $\widetilde{F}_{x,N} : \mathbb{Z}^2 \to \mathbb{R}$ 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

$$M_n(f, g)(S^k T^l x) = \frac{1}{n} \sum_{i=0}^{n-1} f(S^{k+i} T^l x) g(S^k T^{l+i} x) = \widetilde{A}_n(\widetilde{F}_{x,N}, \widetilde{G}_{x,N})(k, l),$$

where we used $ST = TS$ and the definition (1.6). The fact that S and T are measure preserving enables us to write

$$\begin{aligned}
&\|M_{n_j}(f, g) - M_{n_{j-1}}(f, g)\|_{L^2(X)}^2 \\
&= \int_X \left| M_{n_j}(f, g)(x) - M_{n_{j-1}}(f, g)(x) \right|^2 d\mu(x) \\
&= \frac{1}{N^2} \int_X \sum_{k,l=0}^{N-1} \left| M_{n_j}(f, g)(S^k T^l x) - M_{n_{j-1}}(f, g)(S^k T^l x) \right|^2 d\mu(x) \\
&\leq \frac{1}{N^2} \int_X \|\widetilde{A}_{n_j}(\widetilde{F}_{x,N}, \widetilde{G}_{x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{x,N}, \widetilde{G}_{x,N})\|_{\ell^2(\mathbb{Z}^2)}^2 d\mu(x)
\end{aligned}$$

for each $1 \leq j \leq m$. A similar computation as above gives

$$1 = \|f\|_{L^4(X)}^4 = \frac{1}{4N^2} \int_X \sum_{k,l=0}^{2N-1} |f(S^k T^l x)|^4 d\mu(x) = \frac{1}{4N^2} \int_X \|\widetilde{F}_{x,N}\|_{\ell^4(\mathbb{Z}^2)}^4 d\mu(x).$$

Taking $\widetilde{F} = \widetilde{F}_{x,N}, \widetilde{G} = \widetilde{G}_{x,N}$ in Corollary 4 gives

$$\sum_{j=1}^{m} \|\widetilde{A}_{n_j}(\widetilde{F}_{x,N}, \widetilde{G}_{x,N}) - \widetilde{A}_{n_{j-1}}(\widetilde{F}_{x,N}, \widetilde{G}_{x,N})\|_{\ell^2(\mathbb{Z}^2)}^2 \lesssim \|\widetilde{F}_{x,N}\|_{\ell^4(\mathbb{Z}^2)}^4 + \|\widetilde{G}_{x,N}\|_{\ell^4(\mathbb{Z}^2)}^4.$$

Integrating this inequality in x over X and dividing by N^2 yields

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

for any $n_0 < n_1 < \cdots < n_m$. This completes the proof of Theorem 1.

A. *Appendix*

The following inequality (A.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 13. *If $a : [2^i, 2^{i+1}] \to \mathbb{R}$ is a continuously differentiable function, then*

$$\sup_{2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}} \sum_{j=1}^{m} |a(t_j) - a(t_{j-1})|^2$$
$$\lesssim \|a(t)\|_{L_t^2((2^i, 2^{i+1}), dt/t)} \|ta'(t)\|_{L_t^2((2^i, 2^{i+1}), dt/t)}, \tag{A.1}$$

$$\sup_{2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}} \sum_{j=1}^{m} |a(t_j) - a(t_{j-1})|^2 \leq \|ta'(t)\|_{L_t^2((2^i, 2^{i+1}), dt/t)}^2. \tag{A.2}$$

Proof of Lemma 13. To obtain (A.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

$$|a(t_j) - a(t_{j-1})|^2 \lesssim \|a(t)\|_{L_t^2((t_{j-1}, t_j), dt/t)} \|ta'(t)\|_{L_t^2((t_{j-1}, t_j), dt/t)}. \tag{A.3}$$

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_+ and a_- satisfy the required properties and that

$$\|a_+(t)\|_{L_t^2((t_{j-1}, t_j), dt/t)} \leq \|a(t)\|_{L_t^2((t_{j-1}, t_j), dt/t)},$$
$$\|ta'_+(t)\|_{L_t^2((t_{j-1}, t_j), dt/t)} \leq \|ta'(t)\|_{L_t^2((t_{j-1}, t_j), dt/t)}$$

and analogously for a_-, a'_-. Using the triangle inequality and applying (A.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

$$|a(t_j) - a(t_{j-1})|^2 \leq |a(t_j)^2 - a(t_{j-1})^2|$$
$$\leq \left| \int_{t_{j-1}}^{t_j} t(a(t)^2)' \frac{dt}{t} \right| = \left| \int_{t_{j-1}}^{t_j} 2a(t)ta'(t) \frac{dt}{t} \right|.$$

Applying the Cauchy–Schwarz inequality in t, we bound this up to a constant by

$$\left(\int_{t_{j-1}}^{t_j} a(t)^2 \frac{dt}{t} \right)^{1/2} \left(\int_{t_{j-1}}^{t_j} (ta'(t))^2 \frac{dt}{t} \right)^{1/2},$$

which shows (A.3). Summing over j and applying the Cauchy–Schwarz inequality, we obtain

$$\sum_{j=1}^{m} |a(t_j) - a(t_{j-1})|^2$$

$$\lesssim \left(\sum_{j=1}^{m} \|a(t)\|_{L_t^2((t_{j-1},t_j),dt/t)}^2 \right)^{1/2} \left(\sum_{j=1}^{m} \|ta'(t)\|_{L_t^2((t_{j-1},t_j),dt/t)}^2 \right)^{1/2}$$

$$\leq \|a(t)\|_{L_t^2((2^i,2^{i+1}),dt/t)} \|ta'(t)\|_{L_t^2((2^i,2^{i+1}),dt/t)}$$

for any $2^i \leq t_0 < \cdots < t_m \leq 2^{i+1}$, which establishes (A.1).

To see (A.2), we estimate

$$|a(t_j) - a(t_{j-1})|^2 = \left| \int_{t_{j-1}}^{t_j} ta'(t) \frac{dt}{t} \right|^2$$

$$\leq (t_j - t_{j-1}) \int_{t_{j-1}}^{t_j} (ta'(t))^2 \frac{dt}{t^2} \leq 2^i \int_{t_{j-1}}^{t_j} (ta'(t))^2 \frac{dt}{t^2}.$$

The first inequality follows from the Cauchy–Schwarz inequality in t, while for the second inequality we used the crude bound $t_j - t_{j-1} \leq 2^i$. Thus,

$$\sum_{j=1}^{m} |a(t_j) - a(t_{j-1})|^2 \leq 2^i \int_{2^i}^{2^{i+1}} (ta'(t))^2 \frac{dt}{t^2} \leq \int_{2^i}^{2^{i+1}} (ta'(t))^2 \frac{dt}{t},$$

which gives (A.2). □

References

[1] M. Abramowitz and I. A. Stegun (Eds). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1992.

[2] T. Austin. On the norm convergence of non-conventional ergodic averages. *Ergod. Th. & Dynam. Sys.* **30**(2) (2010), 321–338.

[3] J. Avigad and J. Rute. Oscillation and the mean ergodic theorem for uniformly convex Banach spaces. *Ergod. Th. & Dynam. Sys.* **35**(4) (2015), 1009–1027.

[4] F. Bernicot. L^p estimates for non smooth bilinear Littlewood–Paley square functions on $\mathbb{R}$. *Math. Ann.* **351**(1) (2011), 1–49.

[5] F. Bernicot and S. Shrivastava. Boundedness of smooth bilinear square functions and applications to some bilinear pseudo-differential operators. *Indiana Univ. Math. J.* **60**(1) (2011), 233–268.

[6] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Natl. Acad. Sci. USA* **17**(12) (1931), 656–660.

[7] J. Bourgain. Almost sure convergence and bounded entropy. *Israel J. Math.* **63**(1) (1988), 79–97.

[8] J. Bourgain. Pointwise ergodic theorems for arithmetic sets, with an appendix by the author, H. Furstenberg, Y. Katznelson, and D. S. Ornstein. *Publ. Math. Inst. Hautes Études Sci.* **69** (1989), 5–45.

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

[10] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France* **112**(2) (1984), 143–175.

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

[12] Y. Do, R. Oberlin and E. A. Palsson. Variation-norm and fluctuation estimates for ergodic bilinear averages. *Indiana Univ. Math. J.* **66** (2017), 55–99.

[13] S. Donoso and W. Sun. Pointwise multiple averages for systems with two commuting transformations. *Ergod. Th. & Dynam. Sys.* 1–26. doi:10.1017/etds.2016.127. Published online 14 March 2017.

[14] P. Durcik. An L^4 estimate for a singular entangled quadrilinear form. *Math. Res. Lett.* **22**(5) (2015), 1317–1332.

[15] P. Durcik. L^p estimates for a singular entangled quadrilinear form. *Trans. Amer. Math. Soc.* doi:10.1090/tran/6850. Published online 30 March 2017.

[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 and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Anal. Math.* **38**(1) (1978), 275–291.

[18] H. Furstenberg, Y. Katznelson and D. Ornstein. The ergodic theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc. (N.S.)* **7**(3) (1982), 527–552.

[19] R. L. Jones, R. Kaufman, J. M. Rosenblatt and M. Wierdl. Oscillation in ergodic theory. *Ergod. Th. & Dynam. Sys.* **18**(4) (1998), 889–935.

[20] R. L. Jones, I. V. Ostrovskii and J. M. Rosenblatt. Square functions in ergodic theory. *Ergod. Th. & Dynam. Sys.* **16**(2) (1996), 267–305.

[21] R. L. Jones, A. Seeger and J. Wright. Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.* **360**(12) (2008), 6711–6742.

[22] V. Kovač. Bellman function technique for multilinear estimates and an application to generalized paraproducts. *Indiana Univ. Math. J.* **60**(3) (2011), 813–846.

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

[24] V. Kovač. Quantitative norm convergence of double ergodic averages associated with two commuting group actions. *Ergod. Th. & Dynam. Sys.* **36**(3) (2016), 860–874.

[25] V. Kovač and K. A. Škreb. One modification of the martingale transform and its applications to paraproducts and stochastic integrals. *J. Math. Anal. Appl.* **426**(2) (2015), 1143–1163.

[26] V. Kovač and C. Thiele. A $T(1)$ theorem for entangled multilinear dyadic Calderón–Zygmund operators. *Illinois J. Math.* **57**(3) (2013), 775–799.

[27] V. Kovač, C. Thiele and P. Zorin-Kranich. Dyadic triangular Hilbert transform of two general functions and one not too general function. *Forum Math. Sigma* **3** (2015) (e25), 27 pages. doi:10.1017/fms.2015.25.

[28] M. Lacey. On bilinear Littlewood–Paley square functions. *Publ. Mat.* **40**(2) (1996), 387–396.

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

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

[31] M. Mirek, E. Stein and B. Trojan. $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: variational estimates. *Invent. Math.* doi:10.1007/s00222-017-0718-4. Published online 31 January 2017.

[32] P. Mohanty and S. Shrivastava. A note on the bilinear Littlewood–Paley square function. *Proc. Amer. Math. Soc.* **138**(6) (2010), 2095–2098.

[33] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA* **18**(1) (1932), 70–82.

[34] P. K. Ratnakumar and S. Shrivastava. On bilinear Littlewood–Paley square functions. *Proc. Amer. Math. Soc.* **140**(12) (2012), 4285–4293.

[35] J. L. Rubio de Francia. A Littlewood–Paley inequality for arbitrary intervals. *Rev. Mat. Iberoam.* **1**(2) (1985), 1–14.

[36] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergod. Th. & Dynam. Sys.* **28**(2) (2008), 657–688.

[37] T. Tao. Cancellation for the multilinear Hilbert transform. *Collect. Math.* **67**(2) (2016), 191–206.

[38] C. Thiele. *Wave Packet Analysis (CBMS Regional Conference Series in Mathematics, 105)*. American Mathematical Society, Providence, RI, 2006.

[39] M. N. Walsh. Norm convergence of nilpotent ergodic averages. *Ann. of Math. (2)* **175**(3) (2012), 1667–1688.

[40] Wolfram Research, Inc., *Mathematica*, ver. 9.0, Champaign, IL, 2012.

[41] P. Zorin-Kranich. Cancellation for the simplex Hilbert transform. *Math. Res. Lett.*, to appear, *Preprint*, 2015, arXiv:1507.02436.