BILINEAR REPRESENTATION THEOREM

KANGWEI LI, HENRI MARTIKAINEN, YUMENG OU, AND EMIL VUORINEN

Abstract. We represent a general bilinear Calderón–Zygmund operator as a sum of simple dyadic operators. The appearing dyadic operators also admit a simple proof of a sparse bound. In particular, the representation implies a so-called sparse $T_1$ theorem for bilinear singular integrals.

1. Introduction

In this paper we show the exact dyadic structure behind bilinear Calderón–Zygmund operators by representing them using simple dyadic operators, namely some cancellative bilinear shifts and bilinear paraproducts. In the linear case Petermichl [14] first represented the Hilbert transform in this way, and later Hytönen [4] proved a representation theorem for all linear Calderón–Zygmund operators.

The representation theorems were originally motivated by the sharp weighted $A_p$ theory, but they certainly also have other value and intrinsic interest. For example, a representation theorem holds also in the biparameter setting, as shown by Martikainen [10] (the multiparameter extension of this is by Ou [12]), and in this context the representation has proved to be very useful, e.g., in connection with biparameter commutators; see [3], [13].

Outside of the multiparameter context it is true that sparse domination results yield sharp weighted bounds, and that sparse domination can also be proved directly (without going through a representation). Such proofs usually start from the unweighted boundedness assumption, conclude some weak type estimates, and then finally go about proving the sparse domination. However, we think that the idea of a so-called sparse $T_1$, as coined by Lacey and Mena Arias [8], is extremely practical. This amounts to concluding a sparse bound directly from the $T_1$ assumptions (by modifying the probabilistic $T_1$ proof), then noting that the sparse bound implies all of the standard boundedness properties (even the weak type ones). Such a combination gives everything in one blow.

We think that a very efficient way to go about things is to first prove a sharp form of a representation theorem working directly from the $T_1$ assumptions. This is interesting in its own right, entails $T_1$, gives an explicit equality containing

Received by the editors June 19, 2017, and, in revised form, December 3, 2017.

2010 Mathematics Subject Classification. Primary 42B20.

Key words and phrases. Dyadic analysis, Calderón–Zygmund operators, model operators, dyadic shifts, bilinear analysis, representation theorems, $T_1$ theorems, weighted theory.

The first author was supported by Juan de la Cierva—Formación 2015 FJCI-2015-24547, by the Basque Government through the BERC 2014-2017 program, and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323. The second author was supported by the Academy of Finland through Grants No. 294840 and No. 306901. The third author was supported by the National Science Foundation under Grant No. DMS-1440140.

©2018 American Mathematical Society
the full dyadic structure of the operator, and can even be used to transfer sparse bounds, at least in the form sense, from the model dyadic operators to the singular integral. This strategy was employed in the linear setting by Culiuc, Di Plinio, and Ou [2], but of course they were able to cite the linear representation theorem with $T1$ assumptions from the previous literature [5]. It is also to be noted that sparse bounds are remarkably simple to prove for dyadic model operators using the method of [2].

In this paper we, for the first time, prove a representation theorem in the bilinear setting, and we do it starting from the bilinear $T1$ assumptions. Moreover, we carry out the above strategy in the bilinear setting, i.e., we prove sparse domination for our model operators and then transfer them back to the singular integral. In particular, we get a sparse bilinear $T1$ implying directly the boundedness of singular integrals from $\mathcal{L}^p \times \mathcal{L}^q$ to $\mathcal{L}^r$ for all cases in which $1 < p, q < \infty$ and $1/2 < r < \infty$, satisfying $1/p + 1/q = 1/r$, and even the boundedness from $\mathcal{L}^1 \times \mathcal{L}^1$ to $\mathcal{L}^{1/2, \infty}$, just from the $T1$ assumptions. Of course, one can also recover known sharp weighted bounds (see, e.g., [7]) from sparse domination. It is to be noted though that we prove sparse domination in the trilinear form sense, as such bounds are easy to transfer using the representation. A caveat regarding weighted bounds is that outside of the Banach range the literature currently seems to lack an argument giving sharp weighted bounds from form type domination (but such bounds can be derived using pointwise sparse domination [3, 11]).

The proof of the representation entails finding a dyadic-probabilistic proof technique which produces only simple model operators. Some bilinear dyadic-probabilistic methods were studied by Martikainen and Vuorinen in [11] in the nonhomogeneous setting. However, there seem to be a plethora of possible ways to decompose things in the bilinear setting, and one has to be quite careful to really get only nice shifts and nice paraproducts (such that can easily be seen to obey sparse domination). We now move on to formulating some basic definitions and stating our theorems.

A function
\[ K : (\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n) \setminus \Delta \to \mathbb{C}, \quad \Delta := \{(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : x = y = z\} \]
is called a standard bilinear Calderón–Zygmund kernel if for some $\alpha \in (0, 1]$ and $C_K < \infty$ it holds that
\[
|K(x, y, z)| \leq \frac{C_K}{(|x - y| + |x - z|)^2n + 1},
\]
\[
|K(x, y, z) - K(x', y, z)| \leq C_K \frac{|x - x'|^\alpha}{(|x - y| + |x - z|)^2n + \alpha}
\]
whenever $|x - x'| \leq \max(|x - y|, |x - z|)/2$, and
\[
|K(x, y, z) - K(x, y', z)| \leq C_K \frac{|y - y'|^\alpha}{(|x - y| + |x - z|)^2n + \alpha}
\]
whenever $|y - y'| \leq \max(|x - y|, |x - z|)/2$, and
\[
|K(x, y, z) - K(x, y, z')| \leq C_K \frac{|z - z'|^\alpha}{(|x - y| + |x - z|)^2n + \alpha}
\]
whenever $|z - z'| \leq \max(|x - y|, |x - z|)/2$. The best constant $C_K$ is denoted by \(\|K\|_{CZ, \alpha}\).
Given a standard bilinear Calderón–Zygmund kernel $K$, we define

$$T_\varepsilon(f, g)(x) = \int \int_{\max(|x-y|,|x-z|) > \varepsilon} K(x, y, z) f(y) g(z) \ dy \ dz.$$ 

The above is well-defined as an absolutely convergent integral if, e.g., $f \in L^{p_1}$ and $g \in L^{p_2}$ for some $p_1, p_2 \in [1, \infty)$, since then

$$\int \int_{\max(|x-y|,|x-z|) > \varepsilon} |K(x, y, z) f(y) g(z)| \ dy \ dz \lesssim \frac{1}{\varepsilon^{n(1/p_1 + 1/p_2)}} \|f\|_{L^{p_1}} \|g\|_{L^{p_2}}.$$ 

For us a bilinear Calderón–Zygmund operator is essentially the family of truncations $(T_\varepsilon)_\varepsilon$. In particular, this means that boundedness in some $L^p$ spaces is understood in the sense that all $T_\varepsilon$’s are bounded uniformly in $\varepsilon > 0$.

We shall also define some smoother truncations. Suppose that $\varphi \in A$, where $A$ consists of smooth functions $\varphi: [0, \infty) \to [0, 1]$ satisfying $\varphi = 0$ on $[0, 1/2]$, $\varphi = 1$ on $[1, \infty)$, and $\|\varphi\|_{L^\infty} \leq 10$. Define the smoothly truncated singular integrals

$$T^\varphi_\varepsilon(f, g)(x) = \int \int K^\varphi_\varepsilon(x, y, z) f(y) g(z) \ dy \ dz, \quad \varepsilon > 0,$$

where

$$K^\varphi_\varepsilon(x, y, z) = K(x, y, z) \varphi\left(\frac{|x-y| + |x-z|}{\varepsilon}\right).$$

The point is that $T^\varphi_\varepsilon$, $\varepsilon > 0$, are operators with standard bilinear $n$-dimensional kernels (with the kernel bounds being independent of $\varepsilon$). Moreover, we have

$$|T_\varepsilon(f, g)(x) - T^\varphi_\varepsilon(f, g)(x)| \lesssim M(f, g)(x) := \sup_{r > 0} \langle |f| \rangle_{B(x, r)} \langle |g| \rangle_{B(x, r)},$$

where $\langle f \rangle_A := \frac{1}{|A|} \int_A f$. If $0 < \varepsilon_1 < \varepsilon_2$, we denote by $T^\varphi_{\varepsilon_1, \varepsilon_2}$ the operator

$$T^\varphi_{\varepsilon_1, \varepsilon_2}(f, g)(x) = T^\varphi_{\varepsilon_1}(f, g)(x) - T^\varphi_{\varepsilon_2}(f, g)(x)$$

$$= \int \int K^\varphi_{\varepsilon_1, \varepsilon_2}(x, y, z) f(y) g(z) \ dy \ dz,$$

where $K^\varphi_{\varepsilon_1, \varepsilon_2} = K^\varphi_{\varepsilon_1} - K^\varphi_{\varepsilon_2}$.

The notations $T^{1*}$ and $T^{2*}$ stand for the adjoints of a bilinear operator $T$, i.e.,

$$\langle T(f, g), h \rangle = \langle T^{1*}(h, g), f \rangle = \langle T^{2*}(f, h), g \rangle.$$ 

We can now state the main theorem. For the exact definitions of the various objects and notions (random dyadic grids, bilinear cancellative shifts, bilinear para-products, weak boundedness, $T_\delta(1, 1)$, sparse collections, etc.), see the following two sections.

**Theorem 1.1.** Let $K$ be a bilinear Calderón–Zygmund kernel so that $\|K\|_{CZ, n} < \infty$, and let $(T_\varepsilon)_\varepsilon$ be the corresponding bilinear singular integral. Assume that

$$\sup_{\delta > 0} \|T_\delta\|_{WBP} + \|T_\delta(1, 1)\|_{BMO} + \|T^{1*}_\delta(1, 1)\|_{BMO} + \|T^{2*}_\delta(1, 1)\|_{BMO} < \infty.$$
Let also $\varphi \in A$. Then there is a constant $C = C(n, \alpha) < \infty$ so that for all $\varepsilon > 0$ and all compactly supported and bounded functions $f, g, h$ it holds that

$$
\langle T^\varepsilon(f, g), h \rangle = C(\|K\|_{\text{CZ}_n} + \sup_{\delta > 0} \|T_0\|_{\text{WBP}})E_\omega \sum_{k=0}^{\infty} \sum_{i=0}^{k} 2^{-\alpha k/2} \langle U_{\varepsilon,\varphi,\omega}^i(f, g), h \rangle
$$

$$
+ C(\|K\|_{\text{CZ}_n} + \sup_{\delta > 0} \|T_0(1, 1)\|_{\text{BMO}})E_\omega \langle \Pi_0(\varepsilon, \varphi, \omega)(f, g), h \rangle
$$

$$
+ C(\|K\|_{\text{CZ}_n} + \sup_{\delta > 0} \|T_0^1(1, 1)\|_{\text{BMO}})E_\omega \langle \Pi_1(\varepsilon, \varphi, \omega)(f, g), h \rangle
$$

$$
+ C(\|K\|_{\text{CZ}_n} + \sup_{\delta > 0} \|T_0^2(1, 1)\|_{\text{BMO}})E_\omega \langle \Pi_2(\varepsilon, \varphi, \omega)(f, g), h \rangle,
$$

where each $U_{\varepsilon,\varphi,\omega}^i$ is a sum of cancellative bilinear shifts $S_{\varepsilon,\varphi,\omega}^{i,k}$, $S_{\varepsilon,\varphi,\omega}^{i,i+1,k}$ and adjoints of such operators, and where $\Pi_\alpha$ stands for a bilinear paraproduct with $\alpha$ as in \[3.1\]. For a fixed $\omega$ the operators above are defined using the dyadic lattice $D_\omega$.

The following corollary follows from the sparse domination of shifts and paraproducts (see section \[5\]), and the trivial sparse bound for $M$.

**Corollary 1.2.** There exist dyadic grids $D_i$, $i = 1, \ldots, 3^n$, with the following property. Let $\eta \in (0, 1)$. For compactly supported and bounded functions $f, g, h$ and $h$ there is a dyadic grid $D$, and an $\eta$-sparse collection $S = S(f, g, h, \eta) \subset D_1$ such that the following holds.

Let $K$ be any standard bilinear Calderón–Zygmund kernel and $(T^\varepsilon)_{\varepsilon > 0}$ be the corresponding bilinear singular integral. Then we have

$$
\sup_{\varepsilon > 0} |\langle T^\varepsilon(f, g), h \rangle| \leq C_{T,K} \Lambda_S(f, g, h),
$$

where

$$
C_{T,K} := C(\|K\|_{\text{CZ}_n} + \sup_{\varepsilon > 0} \|T^\varepsilon(1, 1)\|_{\text{BMO}} + \sup_{\varepsilon > 0} \|T^1(1, 1)\|_{\text{BMO}} + \sup_{\varepsilon > 0} \|T^2(1, 1)\|_{\text{BMO}} + \sup_{\varepsilon > 0} \|T_0\|_{\text{WBP}})
$$

for some $C = C(n, \alpha, \eta) < \infty$ and

$$
\Lambda_S(f, g, h) := \sum_{Q \in S} |Q| \langle |f| \rangle_Q \langle |g| \rangle_Q \langle |h| \rangle_Q.
$$

1.1. **Additional notation.** We write $A \lesssim B$ if there is an absolute constant $C > 0$ (depending only on some fixed constants like $n, \alpha$, etc.) so that $A \leq CB$. Moreover, $A \lesssim \tau B$ means that the constant $C$ can also depend on some relevant given parameter $\tau > 0$. We may also write $A \sim B$ if $B \lesssim A \lesssim B$.

We then define some notation related to cubes. If $Q$ and $R$ are two cubes, we set the following:

- $\ell(Q)$ is the side length of $Q$.
- If $a > 0$, we denote by $aQ$ the cube that is concentric with $Q$ and has a side length $a\ell(Q)$.
- $d(Q, R) = \text{dist}(Q, R)$ denotes the distance between the cubes $Q$ and $R$.
- $\text{ch}(Q)$ denotes the dyadic children of $Q$.
- If $Q$ is in a dyadic grid, then $Q^{(k)}$ denotes the unique dyadic cube $S$ in the same grid so that $Q \subset S$ and $\ell(S) = 2^k \ell(Q)$.
- If $D$ is a dyadic grid, then $D_k = \{Q \in D \colon \ell(Q) = 2^{-k}\}$.
The notation $\langle f, g \rangle$ stands for the pairing $\int fg$.

The following maximal functions are also used:

$$M_D f(x) = \sup_{Q \in \mathcal{D}} 1_Q(x) ||f||_Q \quad (\mathcal{D} \text{ is a dyadic grid}),$$

$$M f(x) = \sup_{r > 0} ||f||_{B(x,r)}.$$  

Here $B(x,r) = \{ y : |x - y| < r \}$. The bilinear variants are defined in the natural way, e.g.,

$$\mathcal{M}(f,g)(x) = \sup_{r > 0} ||f||_{B(x,r)} ||g||_{B(x,r)}.$$  

2. Basic definitions

2.1. Random dyadic grids, martingales, Haar functions. Let $\omega = (\omega^i)_{i \in \mathbb{Z}}$, where $\omega^i \in \{0,1\}^n$. Let $\mathcal{D}_0$ be the standard dyadic grid on $\mathbb{R}^n$. We define the new dyadic grid

$$\mathcal{D}_\omega = \left\{ I + \sum_{i: 2^{-i} < \ell(I)} 2^{-i} \omega^i : I \in \mathcal{D}_0 \right\} = \left\{ I + \omega : I \in \mathcal{D}_0 \right\},$$

where we simply have defined $I + \omega := I + \sum_{i: 2^{-i} \leq \ell(I)} 2^{-i} \omega^i$. There is a natural product probability measure $\mathbb{P}_\omega = \mathbb{P}$ on $(\{0,1\}^n)^2$; this gives us the notion of random dyadic grids $\omega \mapsto \mathcal{D}_\omega$.

A cube $I \in \mathcal{D} = \mathcal{D}_\omega$ is called bad if there exists such a cube $J \in \mathcal{D}$ that $\ell(J) \geq 2^r \ell(I)$ and

$$d(I, \partial J) \leq \ell(I)^{\gamma} \ell(J)^{1-\gamma}.$$  

Here $\gamma = \alpha/(2[2n + \alpha])$, where $\alpha > 0$ appears in the kernel estimates. Otherwise a cube is called good. We note that $\pi_{\text{good}} := \mathbb{P}_\omega (I + \omega \text{ is good})$ is independent of the choice of $I \in \mathcal{D}_0$. The appearing parameter $r$ is a large enough fixed constant so that $\pi_{\text{good}} > 0$. Moreover, for a fixed $I \in \mathcal{D}_0$ the set $I + \omega$ depends on $\omega^i$ with $2^{-i} < \ell(I)$, while the goodness of $I + \omega$ depends on $\omega^i$ with $2^{-1} \geq \ell(I)$. These notions are thus independent by the product probability structure.

For $I \in \mathcal{D}$ and a locally integrable function $f$ we define the martingale difference

$$\Delta_I f = \sum_{I' \in \text{ch}(I)} \left[ \langle f \rangle_{I'} - \langle f \rangle_I \right] 1_{I'}.$$  

We have the standard estimate

$$\left\| \left( \sum_{I \in \mathcal{D}} |\Delta_I f|^2 \right)^{1/2} \right\|_{L^p} \sim \|f\|_{L^p}, \quad 1 < p < \infty.$$  

Writing $I = I_1 \times \cdots \times I_n$, we can define the Haar function $h_I^\eta$, $\eta = (\eta_1, \ldots, \eta_n) \in \{0,1\}^n$, by setting

$$h_I^\eta = h_{I_1}^{\eta_1} \otimes \cdots \otimes h_{I_n}^{\eta_n},$$

where $h_{I_i}^0 = |I_i|^{-1/2} 1_{I_i}$ and $h_{I_i}^1 = |I_i|^{-1/2} (1_{I_{i,\text{left}}} - 1_{I_{i,\text{right}}})$ for every $i = 1, \ldots, n$. Here $I_{i,\text{left}}$ and $I_{i,\text{right}}$ are the left and right halves of the interval $I_i$, respectively. If $\eta \neq 0$, the Haar function is cancellative: $\int h_I^\eta = 0$. We have

$$\Delta_I f = \sum_{\eta \in \{0,1\}^n \setminus \{0\}} \langle f, h_I^\eta \rangle h_I^\eta,$$
but for convenience we understand that the $\eta$ summation is suppressed and simply write

$$\Delta_I f = \langle f, h_I \rangle h_I.$$  

In this paper $h_I$ always denotes a cancellative Haar function (i.e., $h_I = h_I^0$ for some $\eta \neq 0$). A noncancellative Haar function is explicitly denoted by $h_I^0$.

2.2. Testing conditions: BMO and WBP. Let $K$ be a standard bilinear Calderón–Zygmund kernel, and let $\{T_\varepsilon\}_{\varepsilon > 0}$ be the related family of truncated operators. We recall the usual interpretation of $T_\varepsilon(1,1)$ and what it means that it belongs to BMO (the space of bounded mean oscillation).

Fix some $\varepsilon > 0$. Let $R \subseteq \mathbb{R}^n$ be a closed cube, and let $\phi$ be an $L^\infty$ function supported in $R$ such that $\int \phi = 0$. Let $C = C(\varepsilon) \geq 3$ be any large constant so that $2^{-1}(C - 1)\ell(R) > \varepsilon$, whence $|x - y| > \varepsilon$ for all cases in which $x \in R$ and $y \not\in CR$.

We define

$$\langle T_\varepsilon(1,1), \phi \rangle := \langle T_\varepsilon(1_{CR}, 1_{CR}), \phi \rangle$$

$$+ \int\int\int (K(x,y,z) - K(c_R,y,z)) 1_{(CR \times CR)^c}(y,z) \phi(x)\,dy\,dz\,dx.$$  

Applying the $x$-Hölder estimate of the kernel, it is seen that the integral is absolutely convergent. It is straightforward to check that the right-hand side of (2.1) is independent of the cube $R$ and the constant $C$ as long as $\phi$ is supported in $R$ and $2^{-1}(C - 1)\ell(R) > \varepsilon$, $C \geq 3$.

If $\varphi \in \mathcal{A}$ and $\phi$ is as above, we define

$$\langle T_\varepsilon^\varphi(1,1), \phi \rangle := \langle T_\varepsilon^\varphi(1_{CR}, 1_{CR}), \phi \rangle$$

$$+ \int\int\int (K_\varepsilon^\varphi(x,y,z) - K_\varepsilon^\varphi(c_R,y,z)) 1_{(CR \times CR)^c}(y,z) \phi(x)\,dy\,dz\,dx$$

for any closed cube $R$ containing the support of $\phi$ and any $C \geq 3$, say.

**Definition 2.3.** Let $\varepsilon > 0$. Suppose that $K$ is a standard bilinear Calderón–Zygmund kernel, and let $T_\varepsilon$ be the related truncated operator. We say that $T_\varepsilon(1,1)$ is in BMO, and write $T_\varepsilon(1,1) \in \text{BMO}$, if there exists a constant $C$ so that for all closed cubes $R$ and all functions $\phi$ supported in $R$ such that $\|\phi\|_{L^\infty} \leq 1$ and $\int \phi = 0$, it holds that

$$\left| \frac{\langle T_\varepsilon(1,1), \phi \rangle}{|R|} \right| \leq C.$$  

We denote the smallest constant $C$ in (2.4) by $\|T_\varepsilon(1,1)\|_{\text{BMO}}$.

If $\varphi \in \mathcal{A}$, the corresponding definition for the smoothly truncated operator $T_\varepsilon^\varphi$ is obtained just by replacing $T_\varepsilon$ by $T_\varepsilon^\varphi$.

In the representation theorem we will assume that $T_\varepsilon(1,1) \in \text{BMO}$. The following simple lemma shows that the conditions $T_\varepsilon(1,1) \in \text{BMO}$ and $T_\varepsilon^\varphi(1,1) \in \text{BMO}$ are equivalent.

**Lemma 2.5.** Suppose that $K$ is a standard bilinear Calderón–Zygmund kernel, and let $\varepsilon > 0$ and $\varphi \in \mathcal{A}$. Then

$$\|T_\varepsilon^\varphi(1,1)\|_{\text{BMO}} \leq C\left(\|K\|_{\text{CZ}} + \|T_\varepsilon(1,1)\|_{\text{BMO}}\right)$$
and
\[ \|T_\varepsilon(1,1)\|_{\text{BMO}} \leq C(\|K\|_{CZ_\alpha} + \|T_\varphi^\varepsilon(1,1)\|_{\text{BMO}}). \]

**Proof.** Fix a closed cube \( R \) and a function \( \phi \) supported in \( R \) such that \( \|\phi\|_{L^\infty} \leq 1 \) and \( \int \phi = 0 \). Then, using the definitions (2.1) and (2.2), one sees that
\[
\left| \langle T_\varepsilon(1,1), \phi \rangle - \langle T_\varphi^\varepsilon(1,1), \phi \rangle \right| \leq \langle M(1_{CR},1_{CR}), |\phi| \rangle \leq \int |\phi| dx \leq |R|.
\]

The claim follows from this estimate. \( \square \)

For the convenience of the reader we state the following lemma on the equivalence of some BMO type conditions—although \( T_\varphi^\varepsilon(1,1) \) is not strictly speaking a function, the lemma nevertheless follows from the John–Nirenberg inequality by standard arguments. Therefore, the paraproducts we will encounter can be made to obey the normalization in (3.1).

**Lemma 2.6.** Suppose that \( K \) is a standard bilinear Calderón–Zygmund kernel, and let \( \varepsilon > 0 \) and \( \varphi \in A \). Suppose that \( \mathcal{D} \) is a dyadic lattice. Then
\[
\sup_{R \in \mathcal{D}} \frac{1}{|R|} \sum_{Q \subset R} \left| \langle T_\varphi^\varepsilon(1,1), h_Q \rangle \right|^2 \leq C \|T_\varphi^\varepsilon(1,1)\|_{\text{BMO}}^2
\]
for some absolute constant \( C \).

Next, we give the definition of the weak boundedness property.

**Definition 2.7.** The weak boundedness property constant \( \|T_\varepsilon\|_{\text{WBP}} \) is the best constant \( C \), so the inequality
\[
|\langle T_\varepsilon(1,I), 1_I \rangle| \leq C|I|
\]
holds for all cubes \( I \subset \mathbb{R}^n \).

2.3. **Sparse collections.** A collection \( \mathcal{S} \) of cubes is said to be \( \eta \) sparse (or just sparse), \( 0 < \eta < 1 \), if for any \( Q \in \mathcal{S} \) there exists an \( E_Q \subset Q \) so that \( |E_Q| > \eta|Q| \) and \( \{ E_Q : Q \in \mathcal{S} \} \) are pairwise disjoint. The definition does not require the cubes to be part of some fixed dyadic grid. Although it can be convenient to know that in Corollary 1.2 the sparse family \( \mathcal{S} \) can always be found inside one of the fixed dyadic grids \( \mathcal{D}_i \), \( \#i \lesssim 1 \).

3. Bilinear shifts

In this section all cubes are part of some fixed dyadic grid \( \mathcal{D} \). We will introduce certain cancellative shifts and paraproducts in this section. We will also show their boundedness \( L^p \times L^q \to L^r \) in the simple case \( 1 < p, q, r < \infty \) satisfying \( 1/p + 1/q = 1/r \). The restriction \( r > 1 \) can be lifted after we have shown the sparse domination (see section 5).
3.1. Cancellative bilinear shifts. Define for \(i, j, k \geq 0\) the bilinear shift \((f, g) \mapsto S^{i,j,k}(f, g)\) by setting

\[
S^{i,j,k}(f, g) = \sum_{Q} A^{i,j,k}_{Q}(f, g),
\]

where

\[
A^{i,j,k}_{Q}(f, g) = \sum_{I,J,K \subset Q} \alpha_{I,J,K,Q}\langle f, \tilde{h}_I \rangle \langle g, \tilde{h}_J \rangle h_K
\]

and

\[
(\tilde{h}_I, \tilde{h}_J) \in \{(h_I, h_J), (h^0_I, h_J)(h_I, h^0_J)\}.
\]

We also demand that

\[
|\alpha_{I,J,K,Q}| \leq |I|^{1/2}|J|^{1/2}|K|^{1/2} |Q|^2.
\]

Such a shift will be considered to be a cancellative bilinear shift. Also, the duals of these operators will be used in the representation.

Let \(1 < p, q, r < \infty\) be such that \(1/p + 1/q = 1/r\). We will show that

\[
\|S^{i,j,k}(f, g)\|_{L^r} \lesssim \|f\|_{L^p} \|g\|_{L^q}
\]

with the constant independent of the shift in question, and depending only on \(p, q, r\).

To do this, we may assume without loss of generality that, for example, \(\tilde{h}_I = h_I\) for all \(I\)'s (a general shift can be split into two shifts where \(\tilde{h}_I = h_I\) for all \(I\)'s in one of them and \(\tilde{h}_J = h_J\) for all \(J\)'s in the other). Notice that we have the pointwise estimate

\[
|A^{i,j,k}_{Q}(f, g)| \leq \langle |f| \rangle_Q \langle |g| \rangle_Q 1_Q.
\]

Define also

\[
D^i_Q f = \sum_{I \subset Q} \langle f, h_I \rangle h_I.
\]

Since \(A^{i,j,k}_{Q}(f, g) = A^{i,j,k}_{Q}(D^i_Q f, g)\), we have

\[
|A^{i,j,k}_{Q}(f, g)| \leq M_{D_Q}(|D^i_Q f|)_{Q} 1_Q.
\]

Notice that

\[
\left( \sum_{Q} |D^i_Q f|^2 \right)^{1/2}_{L^p} = \left( \sum_{I} |\Delta_I f|^2 \right)^{1/2}_{L^p} \sim \|f\|_{L^p}, \quad 1 < p < \infty.
\]

Let \(1 < p, q, r < \infty\) be such that \(1/p + 1/q = 1/r\). Using the above, we see that

\[
\|S^{i,j,k}(f, g)\|_{L^r} \sim \left( \sum_{Q} |D^i_Q(S^{i,j,k}(f, g))|^{2} \right)^{1/2}_{L^r} = \left( \sum_{Q} |A^{i,j,k}_{Q}(f, g)|^{2} \right)^{1/2}_{L^r}.
\]
Now we have
\[ \left\| \left( \sum_Q |A_{Q,j,k}^i(f,g)|^2 \right)^{1/2} \right\|_{L^r} \leq \left\| M_D g \left( \sum_Q \langle |D_Q f| \rangle^2_Q \right)^{1/2} \right\|_{L^r} \]
\[ \leq \left\| \left( \sum_Q \langle |D_Q f| \rangle^2_Q \right)^{1/2} \right\|_{L^r} \| M_D g \|_{L^q} \]
\[ \leq \left\| \left( \sum_Q |D_Q f|^2 \right)^{1/2} \right\|_{L^r} \| g \|_{L^q} \lesssim \| f \|_{L^p} \| g \|_{L^q}. \]

### 3.2. Bilinear paraproduct.

Let \( \alpha = \{ \alpha_K \}_{K \in \mathcal{D}} \) be a sequence of complex numbers such that
\[
\frac{1}{|K_0|} \sum_{K : K \subset K_0} |\alpha_K|^2 \leq 1
\]
for all cases in which \( K_0 \in \mathcal{D} \). We define the bilinear paraproduct
\[ \Pi_\alpha(f, g) = \sum_K \alpha_K \langle f \rangle_K \langle g \rangle_K h_K. \]

To deal with this, it is useful to recall the usual (linear) paraproduct
\[ \pi_\alpha f = \sum_K \alpha_K \langle f \rangle_K h_K. \]

It is well known that \( \pi_\alpha : L^r \to L^r \) boundedly for \( 1 < r < \infty \) because of condition (3.1). An elegant way to do this directly in \( L^r \) is in [6]. It follows that \( \Pi_\alpha : L^p \times L^q \to L^r \) boundedly for \( 1 < p, q, r < \infty \), satisfying \( 1/r = 1/p + 1/q \). Indeed, it holds that
\[ \left\| \Pi_\alpha (f, g) \right\|_{L^r} \sim \left\| \left( \sum_K |\alpha_K|^2 \langle f \rangle_K^2 \langle g \rangle_K^2 \frac{1}{|K|} \right)^{1/2} \right\|_{L^r} \]
\[ \leq \left\| \left( \sum_K |\alpha_K|^2 \langle M_D (f, g) \rangle_K^2 \frac{1}{|K|} \right)^{1/2} \right\|_{L^r} \]
\[ \sim \left\| \pi_\alpha (M_D (f, g)) \right\|_{L^r} \lesssim \| M_D (f, g) \|_{L^r} \lesssim \| f \|_{L^p} \| g \|_{L^q}. \]

### 4. Proof of the bilinear representation theorem, Theorem 1.1

Consider an arbitrary \( \varepsilon_1 > 0 \) and let \( f, g, \) and \( h \) be bounded functions with compact support. For the moment, let \( \varepsilon_2 > \varepsilon_1 \) be arbitrary, and write \( T = T_{\varepsilon_1, \varepsilon_2}^* \) and \( K = K_{\varepsilon_1, \varepsilon_2}^* \). This is an a priori bounded operator (for example, in the \( L^4 \times L^4 \to L^2 \) sense), which makes the calculations below legit. We will decompose \( (T(f, g), h) \) first and take the limit \( \varepsilon_2 \to \infty \) at the end.
Begin by decomposing \( \langle T(f, g), h \rangle \) as
\[
\langle T(f, g), h \rangle = E_\omega \sum_{K \in D_\omega} \sum_{I \in D_\omega} \sum_{J \in D_\omega} \langle T(\Delta_I f, \Delta_J g), \Delta_K h \rangle \\
+ E_\omega \sum_{I \in D_\omega} \sum_{J \in D_\omega} \sum_{K \in D_\omega} \langle T^{1*}(\Delta_K h, \Delta_J g), \Delta_I f \rangle \\
+ E_\omega \sum_{J \in D_\omega} \sum_{I \in D_\omega} \sum_{K \in D_\omega} \langle T^{2*}(\Delta_I f, \Delta_K h), \Delta_J g \rangle =: \Sigma^1 + \Sigma^2 + \Sigma^3.
\]

We focus on the first sum \( \Sigma^1 \) and at this point write
\[
\sum_{K \in D_\omega} \sum_{I \in D_\omega} \sum_{J \in D_\omega} \langle T(\Delta_I f, \Delta_J g), \Delta_K h \rangle \\
= \sum_{K \in D_\omega} \langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K h \rangle,
\]
where
\[
E^\omega_{\ell(K)/2} f = \sum_{I \in D_\omega} 1_I(f) 1_{I=\ell(K)/2}.
\]

The point of doing this is to gain the needed independence for the argument below (this seems to be a new simpler way to add goodness than in [5], and it is straightforward to use also in this bilinear setting). We can use now \( D_\omega = D_0 + \omega \) to the end in which
\[
\sum_{K \in D_\omega} \langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K h \rangle = \sum_{K \in D_0} \langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K + \omega h \rangle.
\]

Next, we write
\[
\Sigma^1 = E_\omega \sum_{K \in D_0} \langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K + \omega h \rangle \\
= \frac{1}{\pi_{\text{good}}} \sum_{K \in D_0} E_\omega [1_{\text{good}}(K + \omega)] E_\omega [\langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K + \omega h \rangle] \\
= \frac{1}{\pi_{\text{good}}} E_\omega \sum_{K \in D_\omega, \text{good}} \langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K h \rangle =: \frac{1}{\pi_{\text{good}}} E_\omega \Sigma_1^1(\omega),
\]
where we used independence: \( 1_{\text{good}}(K + \omega) \) depends on \( \omega_j \) for \( 2^{-j} \geq \ell(K) \), while \( E^\omega_{\ell(K)/2} f \) depends on \( \omega_j \) for \( 2^{-j} < \ell(K) < \ell(K)/2 < \ell(K) \), and the same for \( E^\omega_{\ell(K)/2} g \), and \( \Delta_K h \) depends on \( \omega_j \) for \( 2^{-j} < \ell(K) \).

Fix \( \omega \), and let \( D_\omega = D \). We will now start finding the shift structure in the sum \( \Sigma_1^1(\omega) \), i.e.,
\[
\sum_{K \in D_\omega, \text{good}} \sum_{I \in D} \sum_{J \in D} \langle T(\Delta_I f, \Delta_J g), \Delta_K h \rangle.
\]
The double sum \( \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} \) can be organized as
\[
\sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} + \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} .
\]
This leads to the fact that
\[
\sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} \langle T(\Delta_I f, \Delta_J g), \Delta_K h \rangle
\]
\[
= \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} \langle T(\Delta_I f, E_{\ell(I)/2} g), \Delta_K h \rangle
\]
\[
+ \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} \langle T(E_{\ell(J)} f, \Delta_J g), \Delta_K h \rangle =: \sigma^1 + \sigma^2.
\]
We will now mostly focus on the part
\[
\sigma^1 = \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} \langle T(\Delta_I f, E_{\ell(I)/2} g), \Delta_K h \rangle
\]
\[
= \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I \in \mathcal{D}} \sum_{J \in \mathcal{D}} \langle T(\Delta_I f, 1_J \langle g \rangle_J), \Delta_K h \rangle.
\]
However, to get a simple paraproduct, it is crucial to combine, i.e., sum up the paraproduct parts from these two parts, \( \sigma^1 \) and \( \sigma^2 \).

4.1. Step 1: Separated part. In this section we consider
\[
\sigma^1_1 := \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I, J \in \mathcal{D}: \ell(K) \leq \ell(I) = 2\ell(J)} \langle T(\Delta_I f, 1_J \langle g \rangle_J), \Delta_K h \rangle
\]
\[
= \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{I, J \in \mathcal{D}: \ell(K) \leq \ell(I) = 2\ell(J)} \langle T(h_I, h_J^0), h_K \rangle \langle f, h_I \rangle \langle g, h_J^0 \rangle \langle h, h_K \rangle.
\]
We need the existence of certain nice parents, and the proof in the bilinear setting is essentially the same as in \([5]\).

Lemma 4.1. For \( I, J, K \) as in \( \sigma^1_1 \), there exists a cube \( Q \in \mathcal{D} \) such that \( I \cup J \cup K \subset Q \) and
\[
\max(d(K, I), d(K, J)) \gtrsim \ell(K)^{\gamma} \ell(Q)^{1-\gamma}.
\]

Proof. Let \( Q \in \mathcal{D} \) be the minimal parent of \( K \) for which both of the following two conditions hold:
\begin{itemize}
  \item \( \ell(Q) \geq 2^\gamma \ell(K) \),
  \item \( \max(d(K, I), d(K, J)) \leq \ell(K)^{\gamma} \ell(Q)^{1-\gamma} \).
\end{itemize}
Since \( \ell(Q) \geq 2^\gamma \ell(K) \), the goodness of \( K \) gives
\[
\ell(K)^{\gamma} \ell(Q)^{1-\gamma} < d(K, Q^c).
\]
If we would have $I \subset Q^c$ or $J \subset Q^c$, we would get
\[ \ell(K)^\gamma \ell(Q)^{1-\gamma} < \max(d(K, I), d(K, J)) \leq \ell(K)^\gamma \ell(Q)^{1-\gamma}, \]
which is a contradiction. Therefore, we have $I \cap Q \neq \emptyset$ and $J \cap Q \neq \emptyset$. Moreover, we have
\[ \ell(K)^\gamma \ell(J)^{1-\gamma} < \max(d(K, I), d(K, J)) \leq \ell(K)^\gamma \ell(Q)^{1-\gamma}, \]
implying that $\ell(Q) > \ell(J)$, and also that $\ell(Q) \geq \ell(I)$. This implies that $I \cup J \cup K \subset Q$.

It remains to note that the estimate $\max(d(K, I), d(K, J)) \geq \ell(K)^\gamma \ell(Q)^{1-\gamma}$ is a trivial consequence of the minimality of $Q$. Indeed, there is something to check only if $Q$ is minimal because $\ell(Q) \leq \ell(K)$. But then $\ell(Q) \leq \ell(J)$ and we get
\[ \ell(K)^\gamma \ell(Q)^{1-\gamma} \leq \ell(K)^\gamma \ell(J)^{1-\gamma} \leq \max(d(K, I), d(K, J)). \]

\[ \square \]

For $I, J, K$ as in $\sigma^1_I$, we let $Q = I \cup J \cup K$ be the minimal cube $Q \in D$ so that
\[ I \cup J \cup K \subset Q. \]
We then know that
\[ \max(d(K, I), d(K, J)) \geq \ell(K)^\gamma \ell(Q)^{1-\gamma}. \]

Let us write
\[
\sigma^1_I = \sum_{k=0}^{\infty} \sum_{i=0}^{2^k} \sum_{Q \in D} \sum_{\substack{I, J \in D, K \in D_{good} \\
max(d(K, I), d(K, J)) \geq \ell(K)^\gamma \ell(J)^{1-\gamma} \\
2\ell(J) = \ell(I) = 2^{-i} \ell(Q), \ell(K) = 2^{-i} \ell(Q) \\
I \cup J \cup K = Q}} \langle T(h_I, h_J^0), h_K \rangle \langle f, h_I \rangle \langle g, h_J^0 \rangle \langle h, h_K \rangle.
\]

Next, we define
\[ \alpha_{I, J, K, Q} = \frac{\langle T(h_I, h_J^0), h_K \rangle}{C(\ell(K)/\ell(Q))^{\alpha/2}}, \]
if $I, J \in D$, $K \in D_{good}$, $\max(d(K, I), d(K, J)) > \ell(K)^\gamma \ell(J)^{1-\gamma}$, $\ell(K) \leq \ell(I) = 2\ell(J)$, and $I \cup J \cup K = Q$, or $\alpha_{I, J, K, Q} = 0$ otherwise. We can then write for a fixed $\sum_{Q \in D} \sum_{\substack{I, J \in D, K \in D_{good} \\
max(d(K, I), d(K, J)) > \ell(K)^\gamma \ell(J)^{1-\gamma} \\
2\ell(J) = \ell(I) = 2^{-i} \ell(Q), \ell(K) = 2^{-i} \ell(Q) \\
I \cup J \cup K = Q}} \langle T(h_I, h_J^0), h_K \rangle \langle f, h_I \rangle \langle g, h_J^0 \rangle h_K
\]
\[ = C2^{-\alpha k/2} \sum_{Q \in D} \sum_{\substack{I, J, K \subset Q \\
\ell(I) = 2^{-i} \ell(Q), \ell(J) = 2^{-i-1} \ell(Q), \ell(K) = 2^{-i} \ell(Q) \\
I \cup J \cup K = Q}} \langle T(h_I, h_J^0), h_K \rangle \langle f, h_I \rangle \langle g, h_J^0 \rangle h_K
\]
\[ =: C2^{-\alpha k/2} S^{I, j, k}(f, g), \]
which gives
\[ \sigma^1_I = C \sum_{k=0}^{\infty} \sum_{i=0}^{k} 2^{-\alpha i/2} \langle S^{I, j, k}(f, g), h \rangle. \]
It remains for us to verify that
\[ |α_{I,J,K,Q}| \leq \frac{|I|^{1/2} |J|^{1/2} |K|^{1/2}}{|Q|^2} \]
for an appropriate choice of the constant \( C \) depending on the kernel estimates. We fix \( I, J, K, Q \) so that \( α_{I,J,K,Q} \neq 0 \). Notice that \( |x - c_K| \leq ℓ(K)/2 \) (we are using the \( ℓ^∞ \) distance) for \( x \in K \), while
\[ \max(|x - y|, |x - z|) \geq \max(d(K,I),d(K,J)) > ℓ(K)^γ ℓ(J)^{1-γ} \geq 2^γ \frac{ℓ(K)}{2} \geq 2^γ |x - c_K| \]
for \( x \in K, y \in I, \) and \( z \in J \). Therefore, we have by the Hölder estimate in the \( x \) variable and the estimate (4.2) that
\[ |\langle T(h_I, h_J^0), h_K \rangle| \lesssim \|h_I\|_{L^1}\|h_J^0\|_{L^1}\|h_K\|_{L^1} \frac{ℓ(K)^α}{\max(d(K,I),d(K,J))^{2n+α}} \]
\[ \lesssim \frac{|I|^{1/2} |J|^{1/2} |K|^{1/2}}{|Q|^2} \frac{ℓ(K)^α}{ℓ(Q)^{1-γ}2n+α} \]
\[ = \frac{|I|^{1/2} |J|^{1/2} |K|^{1/2}}{|Q|^2} \frac{ℓ(K)^α}{ℓ(Q)^α(2n+α)} \]
\[ = \frac{|I|^{1/2} |J|^{1/2} |K|^{1/2}}{|Q|^2} \frac{ℓ(K)^α}{ℓ(Q)^α/2} \]
This establishes the desired normalization, and therefore we are done with \( σ^1 \).

4.2. Step II: Diagonal. Here we look at the sum
\[ σ^2 := \sum_{K ∈ D_{good}} \sum_{I,J ∈ D : ℓ(K) = 2ℓ(J)} \langle T(Δ_I f, 1_J(g)g), Δ_K h) \rangle \]
\[ = \sum_{K ∈ D_{good}} \sum_{I,J ∈ D : ℓ(K) = 2ℓ(J)} \langle T(h_I, h_J^0), h_K \rangle \langle f, h_I \rangle \langle g, h_J^0 \rangle \langle h, h_K \rangle. \]
The goodness of the cube \( K \) was used to conclude that we cannot have \( ℓ(I) > 2^ε ℓ(K) \). Indeed, in the case \( K ∩ I = Φ \) this would imply that \( d(K,I) > ℓ(K)^γ ℓ(I)^{1-γ} \geq ℓ(K)^γ ℓ(J)^{1-γ} \), a contradiction. In the case \( K ∩ J = Φ \) we would have (as \( ℓ(J) ≥ 2^ε ℓ(K) \)) \( d(K,J) > ℓ(K)^γ ℓ(J)^{1-γ} \), a contradiction.

Lemma 4.3. For \( I, J, K \) as in \( σ^2 \), there exists a cube \( Q ∈ D \) such that \( I ∪ J ∪ K ⊂ Q \) and \( ℓ(Q) ≤ 2^ε ℓ(K) \).

Proof. Define \( Q = K^{(r)} \). Then \( ℓ(Q) = 2^ε ℓ(K) ≥ ℓ(I) > ℓ(J) \). Therefore, it suffices to show that \( I ∩ Q ≠ Φ \) and \( J ∩ Q ≠ Φ \). But this is essentially the same argument as previously: If we would have \( I ⊂ Q^c \) or \( J ⊂ Q^c \), we would get
\[ ℓ(K)^γ ℓ(Q)^{1-γ} < d(K,Q^c) ≤ \max(d(K,I),d(K,J)) ≤ ℓ(K)^γ ℓ(J)^{1-γ}, \]
which implies \( ℓ(J) > ℓ(Q) \), a contradiction. \( \square \)
We can now write
\[
\sigma_2^1 = \sum_{k=0}^{r} \sum_{i=0}^{k} \left( \sum_{I,J \in D, K \in D_{\text{good}}:} \langle T(h_I, h_J^0), h_K \rangle \langle f, h_I \rangle \langle g, h_J^0 \rangle \langle h, h_K \rangle. \right)
\]

Notice that if \( K \cap I = \emptyset \), then
\[
|\langle T(h_I, h_J^0), h_K \rangle| \lesssim |J|^{-1/2}|J|^{-1/2}|K|^{-1/2} \int_{10I \setminus J} \int \frac{dy dx}{|x - y|^n}
\]
\[
\lesssim |J|^{1/2}|J|^{-1/2}|K|^{-1/2} \sim \frac{|J|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2} \left( \frac{\ell(K)}{\ell(Q)} \right)^{\alpha/2}.
\]

We get the same bound also if \( K \cap J = \emptyset \) with an analogous calculation. So we only need to estimate for the case \( K = I \) and \( J \in \text{ch}(K) \). Then we have
\[
|\langle T(h_K, h_J^0), h_K \rangle| \lesssim |K|^{-3/2} \sum_{K', K'' \in \text{ch}(K)} |\langle T(1_{K'}, 1_J), 1_{K''} \rangle|.
\]

If \( K' \neq J \) or \( K'' \neq J \), then \(|\langle T(1_{K'}, 1_J), 1_{K''} \rangle| \lesssim |K|\) simply by the size estimate of the kernel. In the case \( K' = K'' = J \) we have used the weak boundedness property in which \(|\langle T(1_J, 1_J), 1_J \rangle| \lesssim |K|\). So in the case \( K = I \) and \( J \in \text{ch}(K) \) we also have
\[
|\langle T(h_K, h_J^0), h_K \rangle| \lesssim |K|^{-1/2} \sim \frac{|J|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2} \left( \frac{\ell(K)}{\ell(Q)} \right)^{\alpha/2}.
\]

The above lets us write
\[
\sigma_2^1 = C \sum_{k=0}^{r} \sum_{i=0}^{k} 2^{-k\alpha/2} \langle S^{i,i+1,k}(f, g), h \rangle
\]
for cancellative bilinear shifts \( S^{i,i+1,k} \), where \( C \) depends on the kernel estimates and the weak boundedness property. We point out at this point that since \( T = T_{\varepsilon_1, \varepsilon_2} = T_{\varepsilon_1} - T_{\varepsilon_2} \), it holds that
\[
\|T\|_{\text{WBP}} \leq C'\|K\|_{CZ_n} + \sup_{\delta > 0} \|T_{\delta}\|_{\text{WBP}}.
\]

4.3. Step III: Error terms. Here we start working with the sum
\[
\sigma_3^1 := \sum_{I,J \in D, K \in D_{\text{good}}:} \langle T(\Delta_J f, 1_J \langle g \rangle_J), \Delta_K h \rangle
\]
\[
= \sum_{J \in D, K \in D_{\text{good}}:} \langle T(\Delta_{J(1)} f, 1_J), \Delta_K h \rangle \langle g \rangle_J.
\]

We split
\[
\langle T(\Delta_{J(1)} f, 1_J), \Delta_K h \rangle = \langle T(1_J, (\Delta_{J(1)} f) - \langle \Delta_{J(1)} f \rangle_J), 1_J), \Delta_K h \rangle
\]
\[
- \langle \Delta_{J(1)} f \rangle_J \langle T(1, 1_J), \Delta_K h \rangle + \langle \Delta_{J(1)} f \rangle_J \langle T(1, 1), \Delta_K h \rangle.
\]

This gives us the decomposition \( \sigma_3^1 = \sigma_3^1 + \sigma_3^1 + \sigma_3^1 \), where the first two terms of the above decomposition are part of \( \sigma_3^1 \).
In this section we deal only with the error term \( \sigma_{3,e}^1 \). Notice that
\[
\sigma_{3,e}^1 = \sum_{J \in \mathcal{D}, K \in \mathcal{D}_{\text{good}}, K \subseteq J} |J|^{-1/2} |\langle T(s_J, 1_J), h_K \rangle | - |\langle h_{J(1)} \rangle_J \langle T(1, 1_{J^c}), h_K \rangle | \langle f, h_{J(1)} \rangle | \langle g, h^0_J \rangle | \langle h, h_K \rangle |,
\]
where \( s_J := 1_{J^c}(h_{J(1)} - \langle h_{J(1)} \rangle_J) \) satisfies \( |s_J| \lesssim |J|^{-1/2} \) and is supported in \( J^c \).

We will first bound \( |\langle T(s_J, 1_J), h_K \rangle | \). In the case \( \ell(J) \sim \ell(K) \) we are looking for the bound \( |\langle T(s_J, 1_J), h_K \rangle | \lesssim 1 \). This follows by writing
\[
|\langle T(s_J, 1_J), h_K \rangle | \lesssim |K|^{-1/2} |J|^{-1/2} \ell(K) \alpha \int_K \int_{K^c} \frac{dy}{|x-y|^{n+\alpha}} dx \lesssim |K|^{1/2} |J|^{-1/2} \left( \frac{\ell(K)}{\ell(J)} \right)^{\alpha/2}.
\]
Notice that this is \( \sim 1 \) if \( \ell(J) \sim \ell(K) \), so the same estimate holds in both cases. Also, using almost exactly the same calculations as above, that
\[
|\langle T(1, 1_{J^c}), h_K \rangle | \lesssim |K|^{1/2} \left( \frac{\ell(K)}{\ell(J)} \right)^{\alpha/2}.
\]

But as \( |\langle h_{J(1)} \rangle_J| \lesssim |J|^{-1/2} \), we have the same bound as above. Therefore, we can write
\[
\sigma_{3,e}^1 = C \sum_{k=1}^{\infty} 2^{-\alpha k/2} \langle S^{0,1,k}(f, g), h \rangle
\]
for some cancellative bilinear shifts and for some \( C \) depending on the kernel estimates.

4.4. Part IV: Paraproduct. Here we combine
\[
\sigma_{3,\pi}^1 = \sum_{J \in \mathcal{D}, K \in \mathcal{D}_{\text{good}}, K \subseteq J} \langle T(1, 1), \Delta_K h \rangle | \langle \Delta_{J(1)} f \rangle_J \langle g \rangle_J |
\]
with the relevant paraproduct type term coming from \( \sigma^2 \), namely
\[
\sigma_{3,\pi}^2 = \sum_{J \in \mathcal{D}, K \in \mathcal{D}_{\text{good}}, K \subseteq J} \langle T(1, 1), \Delta_K h \rangle | \langle f \rangle_{J(1)} \langle \Delta_{J(1)} g \rangle_J |.
\]

Notice the key cancellation
\[
| \langle \Delta_{J(1)} f \rangle_J \langle g \rangle_J | + | \langle f \rangle_{J(1)} \langle \Delta_{J(1)} g \rangle_J | = | \langle f \rangle_J \langle g \rangle_J | - | \langle f \rangle_{J(1)} \langle g \rangle_{J(1)} |.
\]

Therefore, we get
\[
\sigma_{3,\pi}^1 + \sigma_{3,\pi}^2 = \sum_{K \in \mathcal{D}_{\text{good}}} \langle T(1, 1), \Delta_K h \rangle | \langle f \rangle_K \langle g \rangle_K |.
\]
Define
\begin{equation}
\alpha_K = \frac{\langle T(1, 1), h_K \rangle}{C \langle \|K\|_{\text{CZ}_0}, \sup_{\delta > 0} \|T_{\delta}(1, 1)\|_{\text{BMO}} \rangle}
\end{equation}
if \(K\) is good, where \(C\) is a large enough absolute constant, and otherwise set \(\alpha_K = 0\).

Recall that \(T = T_{\varepsilon_{1}, \varepsilon_{2}} = T_{\varepsilon_{1}} - T_{\varepsilon_{2}}\), whence in view of Lemmas 2.5 and 2.6 the numbers \(\alpha_K\) satisfy the correct normalization (4.1). Hence, we can write
\[
\alpha_{\sigma_{\beta_{1, \varepsilon_{1}, \varepsilon_{2}}}}^{1} + \alpha_{\sigma_{\beta_{1, \varepsilon_{1}, \varepsilon_{2}}}}^{2} = C \langle \|K\|_{\text{CZ}_0}, \sup_{\delta > 0} \|T_{\delta}(1, 1)\|_{\text{BMO}} \rangle \langle \Pi_{\alpha}(f, g), h \rangle.
\]

4.5. Synthesis. Let us collect the pieces of the above steps. Recall that the operator \(T\) is actually \(T_{\varepsilon_{1}, \varepsilon_{2}}\). We have shown that
\[
\Sigma^{1}(\omega) = C \langle \|K\|_{\text{CZ}_0}, \sup_{\delta > 0} \|T_{\delta}\|_{\text{BMO}} \rangle \sum_{k=0}^{\infty} \sum_{i=0}^{k} 2^{-\alpha k/2} \langle U^{i,k}_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}(f, g), h \rangle + C \langle \|K\|_{\text{CZ}_0}, \sup_{\delta > 0} \|T_{\delta}(1, 1)\|_{\text{BMO}} \rangle \langle \Pi_{\alpha}(f, g), h \rangle,
\]
where each \(U^{i,k}_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}\) is a sum of cancellative shifts \(S_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}^{i,k}\) and \(S_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}^{i+1,k}\), and where \(\Pi_{\alpha}(f, g)\) is the paraproduct related to the sequence defined around equation (4.4). Collecting together the symmetric parts, we get the result of Theorem 1.1 except that we have the dependence on \(\varepsilon_{2}\) on both sides. However, it is clear that \(\langle T_{\varepsilon_{1}, \varepsilon_{2}}(f, g), h \rangle = \langle T_{\varepsilon_{1}}(f, g), h \rangle\) if \(\varepsilon_{2}\) is large enough (depending on the supports of \(f, g\), and \(h\)). Thus, it is enough to do some limiting argument \(\varepsilon_{2} \to \infty\) on the right-hand side also.

The operators \(U^{i,k}_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}\) depend on \(\varepsilon_{1}\), \(\varepsilon_{2}\), and \(\varphi\) because the coefficients of the shifts are defined using the operator \(T_{\varepsilon_{1}, \varepsilon_{2}}\). Let \(U^{i,k}_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}\) be the corresponding operator, but where the coefficients of the shifts are defined with the operator \(T_{\varepsilon_{1}}\) instead. Do a similar thing with the paraproducts. The dominated convergence theorem shows that it is enough to show that
\[
\langle U^{i,k}_{\varepsilon_{1}, \varepsilon_{2}, \varphi, \omega}(f, g), h \rangle \to \langle U^{i,k}_{\varepsilon_{1}, \varphi, \omega}(f, g), h \rangle,
\]
when \(\varepsilon_{2} \to \infty\), and similarly for the paraproducts. The convergence of the above pairings is based simply on the fact that the coefficients of the shifts defined with \(T_{\varepsilon_{1}, \varepsilon_{2}}\) approach the ones defined with \(T_{\varepsilon_{1}}\). Let us quickly show the argument for the paraproduct. The same reasoning applies to the cancellative shifts.

It is enough to show that
\[
\lim_{\varepsilon_{2} \to \infty} \sum_{K \in D_{\text{good}}} \langle T_{\varepsilon_{2}}(1, 1), h_K \rangle \langle f \rangle \langle g \rangle \langle h, h_K \rangle = 0.
\]
Fix \(M > 0\). Notice that using \(\sup_{\delta > 0} \|T_{\delta}(1, 1)\|_{\text{BMO}} < \infty\) and the boundedness of the paraproduct, there holds for every \(\varepsilon_{2} > 0\) that
\[
\sum_{K: \ell(K) < 1/M \text{ or } \ell(K) > M} \langle T_{\varepsilon_{2}}(1, 1), h_K \rangle \langle f \rangle \langle g \rangle \langle h, h_K \rangle \lesssim \|f\|_{L^4} \|g\|_{L^4} \left( \sum_{K: \ell(K) < 1/M \text{ or } \ell(K) > M} \|\Delta_K h\|_{L^2}^2 \right)^{1/2} = o(M),
\]
where \( c(M) \to 0 \) when \( M \to \infty \). This gives

\[
\left| \sum_K \langle T_{\varepsilon_2}^\phi (1,1), h_K \rangle_K \langle f \rangle_K \langle g \rangle_K \langle h, h_K \rangle \right| \\
\leq c(M) + \left| \sum_{K: 1/M \leq \ell(K) \leq M} \langle T_{\varepsilon_2}^\phi (1,1), h_K \rangle_K \langle f \rangle_K \langle g \rangle_K \langle h, h_K \rangle \right|.
\]

The latter sum is finite, as \( h \) has compact support. Since \( \langle T_{\varepsilon_2}^\phi (1,1), h_K \rangle \to 0 \) when \( \varepsilon_2 \to \infty \), we have

\[
\lim_{\varepsilon_2 \to \infty} \left| \sum_{K \in \text{D}_{\text{good}}} \langle T_{\varepsilon_2}^\phi (1,1), h_K \rangle_K \langle f \rangle_K \langle g \rangle_K \langle h, h_K \rangle \right| \leq c(M).
\]

The claim follows by letting \( M \to \infty \).

We are done with the proof of Theorem 1.1.

### 5. Sparse form domination for shifts

Let us first introduce a general framework of trilinear forms. Let \( D \) be a fixed dyadic grid on \( \mathbb{R}^n \), and let \( i, j, k \) be nonnegative integers. Define the trilinear form

\[
S^\rho(f_1, f_2, f_3) := \sum_{Q \in D} S_Q(f_1, f_2, f_3)
\]

\[
:= \sum_{Q \in D} \int \int \int K_Q(x_1, x_2, x_3) \prod_{j=1}^3 f_j(x_j) \, dx_1 \, dx_2 \, dx_3,
\]

where \( \rho \geq 0 \). Assume that it satisfies the following:

A. The kernels \( K_Q : Q \times Q \times Q \to \mathbb{C} \) satisfy \( \|K_Q\|_{L^\infty} \leq |Q|^{-2} \).

B. There exist exponents \( p, q, r \in (1, \infty) \) such that \( 1/p + 1/q = 1/r \) and a constant \( B \) such that for every subcollection \( Q \subset D \) of dyadic cubes the truncated form

\[
S_Q^\rho(f_1, f_2, f_3) := \sum_{Q \in \mathcal{Q}} S_Q(f_1, f_2, f_3)
\]

satisfies

\[
|S_Q^\rho(f_1, f_2, f_3)| \leq B \|f_1\|_{L^p} \|f_2\|_{L^q} \|f_3\|_{L^r}'.
\]

C. \( K_Q \) is constant on sets of the form \( Q_1 \times Q_2 \times Q_3 \), where \( Q_i^{(\rho+1)} = Q_i \).

It can easily be seen that trilinear forms associated with both cancellative bilinear shifts and paraproducts fall into the above class of forms. Corollary 1.2 follows from Theorem 1.1 by using two results from this section, namely Proposition 5.1 and Corollary 5.8.

We state the next proposition for only dyadic grids without quadrants—these are dyadic grids where every sequence of cubes \( I_k \) with \( I_k \subseteq I_{k+1} \) satisfy \( \mathbb{R}^n = \bigcup I_k \). Since almost every dyadic grid has this property, this generality is already enough for us to conclude everything we need. Of course, the proposition would hold in every grid, but since this is not needed, we prefer this technical simplification.

**Proposition 5.1.** Let \( \eta \in (0, 1) \), let \( D \) be a dyadic grid without quadrants, and let \( f_1, f_2, f_3 \) be compactly supported and bounded functions. Then there exists an
The cube $Q \in D$ is such that it contains the supports of all three functions $f_j$. Define $\mathcal{E}$ to be the collection of maximal cubes $Q \in D$, $Q \subset Q_0$ such that

\[ \max \left( \frac{\langle |f_1| \rangle_Q}{\langle |f_1| \rangle_{Q_0}}, \frac{\langle |f_2| \rangle_Q}{\langle |f_2| \rangle_{Q_0}}, \frac{\langle |f_3| \rangle_Q}{\langle |f_3| \rangle_{Q_0}} \right) > C_0. \]

For a large enough $C_0 = C_{0}(\eta)$ it holds that

\[ \sum_{Q \in \mathcal{E}} |Q| \leq (1 - \eta)|Q_0|. \]

The cube $Q_0$ is the first cube to be included in $\mathcal{S}$, and $E_{Q_0} := Q_0 \setminus \bigcup_{Q \in \mathcal{E}} Q$.

Let $\mathcal{G} = \mathcal{G}(Q_0) := \{ Q \in D : Q \subset Q_0 \text{ and } Q \not\subset Q' \text{ for every } Q' \in \mathcal{E} \}$, and for $Q \in D$ write $\mathcal{D}(Q) = \{ R \in D : R \subset Q \}$. Then we have the decomposition

\[ \mathcal{S}^p(f_1, f_2, f_3) = \sum_{Q \in \mathcal{D}} S_Q(f_1, f_2, f_3) + \mathcal{S}'(f_1, f_2, f_3) \]

\[ + \sum_{Q \in \mathcal{E}} \mathcal{S}_{D(Q)}(f_1 1_Q, f_2 1_Q, f_3 1_Q), \]

where we applied the fact that the functions are supported in $Q_0$. The size property $\|K_0\|_{L^\infty} \leq |Q|^{-2}$ of the kernels implies that

\[ \left| \sum_{Q \in \mathcal{D}, Q \supset Q_0} S_Q(f_1, f_2, f_3) \right| \leq \sum_{Q \in \mathcal{D}, Q \supset Q_0} \frac{\|f_1\|_{L^1} \cdot \|f_2\|_{L^1} \cdot \|f_3\|_{L^1}}{|Q|^2} \sim |Q_0| \prod_j \langle |f_j| \rangle_{Q_0}. \]

We will prove the estimate

\[ \mathcal{S}'(f_1, f_2, f_3) \lesssim_{\eta} (B + \rho) |Q_0| \prod_j \langle |f_j| \rangle_{Q_0}. \]

From (5.3) and (5.4) it is then seen that the collection $\mathcal{S}$ can be obtained by iterating this process, in the second step beginning with $\mathcal{S}_{D(Q)}^p(f_1 1_Q, f_2 1_Q, f_3 1_Q)$ for some $Q \in \mathcal{E}$. Hence, to conclude the proof, it remains to show (5.4).

We prove (5.4) by performing a Calderón–Zygmund decomposition to $f_j$ with respect to the collection $\mathcal{E}$, obtaining for each $j = 1, 2, 3$ that

\[ f_j = g_j + b_j := g_j + \sum_{Q \in \mathcal{E}} b_{j,Q}, \quad b_{j,Q} := \left( f_j - \langle f_j \rangle_Q \right) 1_Q. \]

For every $Q \in \mathcal{E}$ there hold the standard properties

\[ \|g_j\|_{L^\infty} \lesssim_{\eta} \langle |f_j| \rangle_{Q_0}, \quad \int_Q b_{j,Q} = 0, \quad \|b_{j,Q}\|_{L^1} \lesssim_{\eta} |Q| \langle |f_j| \rangle_{Q_0}. \]

Decompose the left-hand side of (5.4) into eight parts:

\[ \mathcal{S}_p'((g_1, g_2, g_3), \mathcal{S}_p'((b_1, b_2, b_3), \mathcal{S}_p'((g_1, g_2, b_3), \mathcal{S}_p'((g_1, b_2, g_3), \ldots. \]
The part with three good functions can be directly estimated via the boundedness of \( S_G^p \) and the estimates \( \|g_j\|_{L^\infty} \lesssim_\eta \langle |f_j| \rangle_{Q_0} \):

\[
|S_G^p(g_1, g_2, g_3)| \leq B\|g_1\|_{L^p}\|g_2\|_{L^p}\|g_3\|_{L^p} \lesssim_\eta B|Q_0| \prod_j \langle |f_j| \rangle_{Q_0}.
\]

In all of the other parts, there is at least one bad function involved. All of these terms vanish by assumption C if \( \rho = 0 \), so assume now that \( \rho \geq 1 \). By symmetry we consider a term of the form \( S_G^p(b_1, h_2, h_3) \), where \( h_j \) can be either \( g_j \) or \( b_j \). We further decompose \( G \) into \( \rho \) subcollections, each of which, denoted by \( G' \), satisfies \( \ell(I_1) \geq 2^\rho \ell(I_2) \) whenever \( I_1, I_2 \in G' \). It suffices to show that

\[
|S_{G'}^p(b_1, h_2, h_3)| \lesssim_\eta |Q_0| \prod_j \langle |f_j| \rangle_{Q_0}.
\]

Because of assumption C, the defining property of \( G' \) and the fact that \( \int b_{1,Q} = 0 \) for every \( Q \in \mathcal{E} \), we find that for every \( Q \in \mathcal{E} \) there exists at most one \( R \in \mathcal{G}' \) such that \( Q \subseteq R \) and \( S_R(b_{1,Q}, h_2, h_3) \neq 0 \). If such a cube \( R \) exists, we denote it by \( R(Q) \). Therefore,

\[
|S_{G'}^p(b_1, h_2, h_3)| \leq \sum_{R \in G'} \sum_{Q \in \mathcal{E}} |S_R(b_{1,Q}, h_2, h_3)|
\]

\[
\leq \sum_{R \in G'} \sum_{Q \in \mathcal{E}} \frac{\|b_{1,Q}\|_{L^1} \|b_21_R\|_{L^1} \|b_31_R\|_{L^1}}{|R|^2},
\]

where the size estimate \( \|K_R\|_{L^\infty} \leq |R|^{-2} \) was applied.

Let \( j = 2,3 \) and fix some \( R \in \mathcal{G}' \) for the moment. We will prove \( \|b_{1,R}\|_{L^1} \lesssim_\eta |R|\langle |f_j| \rangle_{Q_0} \). The \( L^\infty \) property of \( g_j \) implies that \( \|g_j1_R\|_{L^1} \lesssim_\eta |R|\langle |f_j| \rangle_{Q_0} \). The estimates \( \|b_{j,Q}\|_{L^1} \lesssim_\eta |Q|\langle |f_j| \rangle_{Q_0} \) give

\[
\|b_{1,R}\|_{L^1} = \sum_{Q \in \mathcal{E} : Q \subseteq R} \|b_{j,Q}\|_{L^1} \lesssim_\eta \sum_{Q \in \mathcal{E} : Q \subseteq R} |Q|\langle |f_j| \rangle_{Q_0} \leq |R|\langle |f_j| \rangle_{Q_0}.
\]

Now we proceed from (5.6) as

\[
|S_{G'}^p(b_1, h_2, h_3)| \lesssim_\eta \sum_{R \in G'} \sum_{Q \in \mathcal{E}} \|b_{1,Q}\|_{L^1} \langle |f_2| \rangle_{Q_0} \langle |f_3| \rangle_{Q_0}
\]

\[
\lesssim_\eta |Q_0| \langle |f_1| \rangle_{Q_0} \langle |f_2| \rangle_{Q_0} \langle |f_3| \rangle_{Q_0}.
\]

This completes the proof of (5.5), and hence the proof of the proposition. \( \square \)

For clarity we give the proof of the following lemma—it is a simple argument that can be extracted from the proof of Lemma 4.7 by Lacey and Mena Arias [8].

**Lemma 5.7.** Let \( 0 < \eta_1, \eta_2 < \infty \). Suppose that \( D \) is a dyadic grid and \( f_1, f_2, f_3 \in L^1 \). Then there is an \( \eta_2 \)-sparse family \( U = U(f_1, f_2, f_3, \eta_2) \subset D \) so that for all \( \eta_1 \)-sparse \( S's \subset D \) it holds that

\[
\Lambda_S(f_1, f_2, f_3) \lesssim_{\eta_1, \eta_2} \Lambda_U(f_1, f_2, f_3).
\]
Proof. We first construct the family $\mathcal{U}$. Let $C = C(\eta_2) \geq 8^n$ be a large enough constant depending on $\eta_2$. For each $k \in \mathbb{Z}$ define

$$U_k = \left\{ \text{maximal cubes } Q \in \mathcal{D} \text{ so that } \prod_j \langle |f_j| \rangle_Q > C^k \right\}.$$ 

Notice that if $Q \in U_k$, then

$$C^k < \prod_j \langle |f_j| \rangle_Q \leq 8^n \prod_j \langle |f_j| \rangle_{Q(1)} \leq 8^n C^k \leq C^{k+1}.$$ 

This means that a given $Q \in \mathcal{D}$ can belong to at most one of the collections $U_k$. Define

$$U = \bigcup_{k \in \mathbb{Z}} U_k.$$ 

Let us show that this is an $\eta_2$-sparse collection. Let $Q \in U$ and fix $k$ so that $Q \in U_k$. Notice first that

$$\left| \bigcup_{R \in U} R \right| = \left| \bigcup_{R \in U_{k+1}} R \right| = \sum_{R \in U_{k+1}} |R|.$$ 

If $R \in U_{k+1}$ is such that $R \subset Q$, then

$$\prod_j \langle |f_j| \rangle_R > C^{k+1} \geq \frac{C}{8^n} \prod_j \langle |f_j| \rangle_Q,$$

so

$$\max_j \frac{\langle |f_j| \rangle_R}{\langle |f_j| \rangle_Q} \geq \frac{C^{1/3} \eta_2}{2^n}.$$ 

This implies that

$$\left| \bigcup_{R \in U} R \right| \leq 3 \cdot 2^n \frac{C^{1/3}}{\eta_2^{1/3}} |Q| \leq (1 - \eta_2) |Q|$$

provided that $C = C(\eta_2)$ is large enough. It is now clear that the sets

$$E_Q := Q \setminus \bigcup_{R \in U} R, \quad Q \in U$$

are disjoint and satisfy $|E_Q| \geq \eta_2 |Q|$, which proves that $U$ is $\eta_2$ sparse.

Consider an arbitrary $S \subset \mathcal{D}$, which is $\eta_1$ sparse. If $Q \in S$ satisfies $\prod_j \langle |f_j| \rangle_Q \neq 0$, then there is a cube $R \in U$ such that $Q \subset R$. Let $\pi_u Q$ denote the minimal $R \in U$ so that $Q \subset R$. Suppose that $\pi_u Q \in U_k$. Then we cannot have $\prod_j \langle |f_j| \rangle_Q > C^{k+1}$ (as otherwise $\pi_u Q$ would not be minimal), so

$$\prod_j \langle |f_j| \rangle_Q \leq C^{k+1} \leq C \prod_j \langle |f_j| \rangle_{\pi_u Q} \lesssim_{\eta_2} \prod_j \langle |f_j| \rangle_{\pi_u Q}.$$
Finally, we get
\[
\Lambda_S(f_1, f_2, f_3) = \sum_{R \in \mathcal{U}} \sum_{Q \in S} |Q| \prod_j \langle |f_j| \rangle_Q
\]
\[
\lessapprox_{\eta_2} \sum_{R \in \mathcal{U}} \prod_j \langle |f_j| \rangle_R \sum_{Q \in S} |Q| \quad (5.9)
\]
\[
\lessapprox_{\eta_1} \sum_{R \in \mathcal{U}} |R| \prod_j \langle |f_j| \rangle_R = \Lambda_U(f_1, f_2, f_3). \quad \square
\]

**Corollary 5.8.** There exist dyadic grids $\mathcal{D}_i$, $i = 1, \ldots, 3^n$ with the following property. Let $\eta_1, \eta_2 \in (0,1)$. Suppose that $f_1, f_2, f_3 \in L^1$. Then for some $i$ there exists an $\eta_2$-sparse collection $\mathcal{U} = \mathcal{U}(f_1, f_2, f_3, \eta_2) \subset \mathcal{D}_i$ such that for all $\eta_1$-sparse collections of cubes $S$ we have

\[
(5.9) \quad \Lambda_S(f_1, f_2, f_3) \lessapprox_{\eta_1, \eta_2} \Lambda_U(f_1, f_2, f_3).
\]

**Proof.** We can let $(\mathcal{D}_i)_i$ be any collection of $3^n$ dyadic grids with the property that for any cube $P \subset \mathbb{R}^n$ there exists an $R \in \bigcup \mathcal{D}_i$ such that $P \subset R$ and $\ell(R) \leq 6\ell(P)$. Then it is easy to find a $6^{-n}\eta_1$-sparse collections $S_i \subset \mathcal{D}_i$ (depending on $S$) such that

\[
\Lambda_S(f_1, f_2, f_3) \lessapprox_{\eta_1} \sum_i \Lambda_{S_i}(f_1, f_2, f_3).
\]

Let $\mathcal{U}_i = \mathcal{U}_i(f_1, f_2, f_3, \eta_2) \subset \mathcal{D}_i$ be the universal sparse collections given by Lemma 5.7. Then we have

\[
\Lambda_S(f_1, f_2, f_3) \lessapprox_{\eta_1} \sum_i \Lambda_{S_i}(f_1, f_2, f_3) \lessapprox_{\eta_1, \eta_2} \sum_i \Lambda_{\mathcal{U}_i}(f_1, f_2, f_3) \lessapprox_{\eta_1} \Lambda_{\mathcal{U}_i}(f_1, f_2, f_3)
\]

for some $i_0$. We are done. \quad \square

**Acknowledgments**

The second and fourth authors are members of the Finnish Centre of Excellence in Analysis and Dynamics Research. The third author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, when this work was carried out.

**References**


BCAM (Basque Center for Applied Mathematics), Alameda de Mazarredo 14, 48009 Bilbao, Spain

*Email address*: kli@bcamath.org

Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, FI-00014 University of Helsinki, Finland

*Email address*: henri.martikainen@helsinki.fi

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

*Email address*: yumengou@mit.edu

Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, FI-00014 University of Helsinki, Finland

*Email address*: emil.vuorinen@helsinki.fi