ANALYSIS & PDE

Volume 11 No. 7 2018

IRINA HOLMES, STEFANIE PETERMICHL AND BRETT D. WICK

WEIGHTED LITTLE BMO AND TWO-WEIGHT INEQUALITIES FOR JOURNÉ COMMUTATORS

msp

ANALYSIS AND PDE
Vol. 11, No. 7, 2018
dx.doi.org/10.2140/apde.2018.11.1693

msp

WEIGHTED LITTLE BMO AND TWO-WEIGHT INEQUALITIES FOR JOURNÉ COMMUTATORS

IRINA HOLMES, STEFANIE PETERMICHL AND BRETT D. WICK

We characterize the boundedness of the commutators $[b, T]$ with biparameter Journé operators T in the two-weight, Bloom-type setting, and express the norms of these commutators in terms of a weighted little bmo norm of the symbol b. Specifically, if μ and λ are biparameter A_p weights, $\nu := \mu^{1/p}\lambda^{-1/p}$ is the Bloom weight, and b is in $\text{bmo}(\nu)$, then we prove a lower bound and testing condition $\|b\|_{\text{bmo}(\nu)} \lesssim \sup \|[b, R_k^1 R_l^2] : L^p(\mu) \to L^p(\lambda)\|$, where R_k^1 and R_l^2 are Riesz transforms acting in each variable. Further, we prove that for such symbols b and any biparameter Journé operators T, the commutator $[b, T] : L^p(\mu) \to L^p(\lambda)$ is bounded. Previous results in the Bloom setting do not include the biparameter case and are restricted to Calderón–Zygmund operators. Even in the unweighted, $p = 2$ case, the upper bound fills a gap that remained open in the multiparameter literature for iterated commutators with Journé operators. As a by-product we also obtain a much simplified proof for a one-weight bound for Journé operators originally due to R. Fefferman.

1. Introduction and statement of main results

Bloom [1985] proved a two-weight version of the celebrated commutator theorem of Coifman, Rochberg and Weiss [Coifman et al. 1976]. Specifically, Bloom characterized the two-weight norm of the commutator $[b, H]$ with the Hilbert transform in terms of the norm of b in a certain weighted BMO space:

$$\|[b, H] : L^p(\mu) \to L^p(\lambda)\| \simeq \|b\|_{\text{BMO}(\nu)},$$

This material is based upon work of Holmes supported by the National Science Foundation under award no. 1606270. The research of Petermichl is supported in part by ERC project CHRiSHarMa no. DLV-682402. The research of Wick is supported in part by a National Science Foundation DMS grant no. 1560955.
MSC2010: 42B20, 42B25, 42A50.
Keywords: commutators, Calderón–Zygmund operators, bounded mean oscillation, weights, Journé operators, little BMO, multiparameter harmonic analysis, singular integrals, weighted inequalities.

where μ, λ are A_p weights, $1 < p < \infty$, and $\nu := \mu^{1/p}\lambda^{-1/p}$. Recently, this was extended to the n-dimensional case of Calderón–Zygmund operators in [Holmes et al. 2017], using the modern dyadic methods started by [Petermichl 2000] and continued in [Hytönen 2012]. The main idea in these methods is to represent continuous operators like the Hilbert transform in terms of dyadic shift operators. This theory was recently extended to biparameter singular integrals in [Martikainen 2012].

In this paper we extend the Bloom theory to commutators with biparameter Calderón–Zygmund operators, also known as Journé operators, and characterize their norms in terms of a weighted version of the little bmo space of [Cotlar and Sadosky 1996]. The main results are:

Theorem 1.1 (upper bound). *Let T be a biparameter Journé operator on $\mathbb{R}^{\vec{n}} = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, as defined in Section 7A. Let μ and λ be $A_p(\mathbb{R}^{\vec{n}})$ weights, $1 < p < \infty$, and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|[b, T] : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{bmo}(\nu)},$$

where $\|b\|_{\mathrm{bmo}(\nu)}$ denotes the norm of b in the weighted little $\mathrm{bmo}(\nu)$ *space on $\mathbb{R}^{\vec{n}}$.*

We make a few remarks about the proof of this result. At its core, the strategy is the same as in [Holmes et al. 2017], and may be roughly stated as:

(1) Use a representation theorem to reduce the problem from bounding the norm of $[b, T]$ to bounding the norm of $[b, \text{dyadic shift}]$.

(2) Prove the two-weight bound for $[b, \text{dyadic shift}]$ by decomposing into paraproducts.

However, the biparameter case presents some significant new obstacles. In [Holmes et al. 2017], T was a Calderón–Zygmund operator on $\mathbb{R}^n$, and the representation theorem was that of [Hytönen 2012]. In the present paper, T is a biparameter Journé operator on $\mathbb{R}^{\vec{n}} = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ (see Section 7A) and we use Martikainen's representation theorem [2012] to reduce the problem to commutators $[b, \mathbb{S}_{\boldsymbol{\mathcal{D}}}]$, where $\mathbb{S}_{\boldsymbol{\mathcal{D}}}$ is now a *biparameter* dyadic shift. These can be cancellative, i.e., all Haar functions have mean zero (defined in Section 7C), or noncancellative (defined in Section 7D). The strategy is summarized in Figure 1.

The main difficulty arises from the structure of the biparameter dyadic shifts. At first glance, the cancellative shifts are "almost" compositions of two one-parameter shifts $\mathbb{S}_{\mathcal{D}_1}$ and $\mathbb{S}_{\mathcal{D}_2}$ applied in each variable — if this were so, many of the results would follow trivially by iteration of the one-parameter results. Unfortunately, there is no reason for the coefficients $a_{P_1Q_1R_1P_2Q_2R_2}$ in the biparameter shifts to "separate" into a product $a_{P_1Q_1R_1} \cdot a_{P_2Q_2R_2}$, as would be required in a composition of two one-parameter shifts. Therefore, many of the inequalities needed for biparameter shifts must be proved from scratch.

Even more difficult is the case of noncancellative shifts. As outlined in Section 7D, these are really paraproducts, and there are three possible types that arise from the representation theorem:

(1) full standard paraproducts;

(2) full mixed paraproducts;

(3) partial paraproducts.

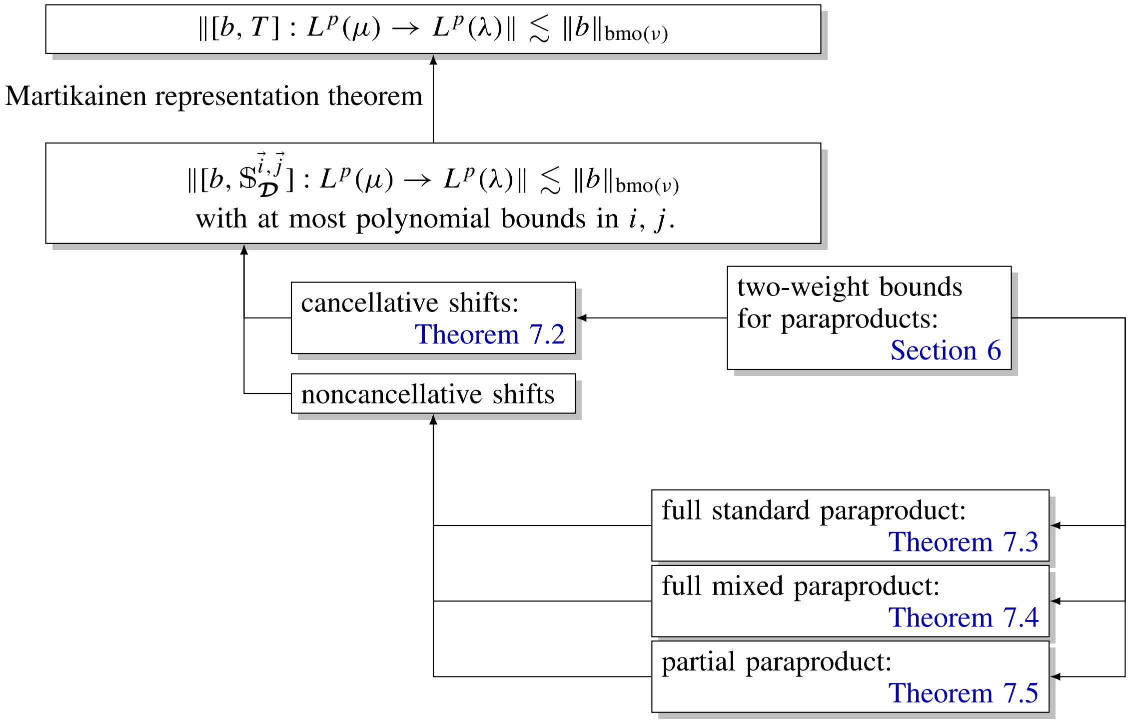

Figure 1. Strategy for Theorem 1.1.

These methods were considered previously in [Ou et al. 2016; Ou and Petermichl 2018] for the unweighted, $p = 2$ case. In [Ou et al. 2016] it was shown that

$$\|[b, T] : L^2(\mathbb{R}^{\vec{n}}) \to L^2(\mathbb{R}^{\vec{n}})\| \lesssim \|b\|_{\mathrm{bmo}(\mathbb{R}^{\vec{n}})},$$

where T is a *paraproduct-free* Journé operator. This restriction essentially means that all the dyadic shifts in the representation of T are *cancellative*, so the case of noncancellative shifts remained open. This gap was partially filled in [Ou and Petermichl 2018], which treats the case of noncancellative shifts of standard paraproduct type. So the case of general Journé operators, which includes noncancellative shifts of mixed and partial type in the representation, remained open even in the unweighted, $p = 2$ case. These types of paraproducts are notoriously difficult — see also [Martikainen and Orponen 2016] for a wonderful discussion of this issue. We fill this gap in Section 7D, where we prove two-weight bounds of the type

$$\|[b, \mathbb{S}_{\mathcal{D}}] : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{bmo}(\nu)},$$

where $\mathbb{S}_{\mathcal{D}}$ is a noncancellative shift. The same is proved for cancellative shifts in Section 7C.

At the backbone of all these proofs will be the biparameter paraproducts, developed in Section 6, and a variety of biparameter square functions, developed in Section 3. For instance, in the case of the cancellative shifts, one can decompose the commutator as

$$[b, \mathbb{S}_{\mathcal{D}}^{\vec{i},\vec{j}}]f = \sum [\mathsf{P}_{\mathsf{b}}, \mathbb{S}_{\mathcal{D}}^{\vec{i},\vec{j}}]f + \sum [\mathsf{p}_{\mathsf{b}}, \mathbb{S}_{\mathcal{D}}^{\vec{i},\vec{j}}]f + \mathcal{R}_{\vec{i},\vec{j}}f.$$

Here P_{b} runs through nine paraproducts associated with *product BMO*, and p_{b} runs through six paraproducts associated with *little bmo*, so we are dealing with fifteen paraproducts in total in the biparameter case. Some of these are straightforward generalizations of the one-parameter paraproducts, while some are more complicated "mixed" paraproducts. Two-weight bounds are proved for all these paraproducts in Section 6, building on two essential blocks: the biparameter square functions in Section 3, and the weighted H^1-BMO duality in the product setting, developed in Section 4. In fact, Section 4 is a self-contained presentation of large parts of the weighted biparameter BMO theory.

Once the paraproducts are bounded, all that is left is to bound the so-called "remainder term" $\mathcal{R}_{\vec{i},\vec{j}}f$, of the form $\Pi_{\mathbb{S}f}b - \mathbb{S}\Pi_f b$, where one can no longer appeal directly to the paraproducts. At this point, however, things become very technical, so bounding the remainder terms is no easy task. To help guide the reader, we outline below the general strategy we will employ. This applies to Theorem 7.2, and in large part to Theorems 7.3, 7.4, and 7.5:

(1) We break up the remainder term into more convenient sums of operators of the type $\mathcal{O}(b, f)$, involving both $b \in \mathrm{bmo}(\nu)$ and $f \in L^p(\mu)$. We want to show $\|\mathcal{O}(b, f) : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{bmo}(\nu)}$. Using duality this amounts to showing that

$$|\langle \mathcal{O}(b, f), g\rangle| \lesssim \|b\|_{\mathrm{BMO}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')}.$$

(2) Some of these operators $\mathcal{O}(b, f)$ involve full Haar coefficients $\hat{b}(Q_1 \times Q_2)$ of b, while others involve a Haar coefficient in one variable and averaging in the other variable, such as $\langle b, h_{Q_1} \times \mathbb{1}_{Q_2}/|Q_2|\rangle$. Since, ultimately, we wish to use some type of H^1-BMO duality, the goal will be to "separate out" b from the inner product $\langle \mathcal{O}(b, f), g\rangle$. If $\mathcal{O}(b, f)$ involves full Haar coefficients of b, we use duality with *product BMO* and obtain

$$|\langle \mathcal{O}(b, f), g\rangle| \lesssim \|b\|_{\mathrm{BMO}(\nu)} \|S_{\boldsymbol{\mathcal{D}}}\phi(f, g)\|_{L^1(\nu)},$$

where $\phi(f, g)$ is the operator we are left with after separating out b, and $S_{\boldsymbol{\mathcal{D}}}$ is the full biparameter dyadic square function. If $\mathcal{O}(b, f)$ involves terms of the form $\langle b, h_{Q_1} \times \mathbb{1}_{Q_2}/|Q_2|\rangle$, we use duality with *little bmo*, and obtain something of the form

$$|\langle \mathcal{O}(b, f), g\rangle| \lesssim \|b\|_{\mathrm{bmo}(\nu)} \|S_{\mathcal{D}_1}\phi(f, g)\|_{L^1(\nu)},$$

where $S_{\mathcal{D}_1}$ is the dyadic square function in the first variable. Obviously this is replaced with $S_{\mathcal{D}_2}$ if the Haar coefficient on b is in the second variable.

(3) Then the next goal is to show that

$$S_{\boldsymbol{\mathcal{D}}}\phi(f, g) \lesssim (\mathcal{O}_1 f)(\mathcal{O}_2 g),$$

where $\mathcal{O}_{1,2}$ will be operators satisfying a *one-weight* bound of the type $L^p(w) \to L^p(w)$. These operators will usually be a combination of the biparameter square functions in Section 3. Once we have this, we are done.

In Theorem 7.2, dealing with cancellative shifts, the crucial part is really step (1). At first glance, the remainder term $\mathcal{R}_{\vec{i},\vec{j}}f$ seems intractable using this method, since it involves average terms $\langle b\rangle_{Q_1 \times Q_2}$ instead of Haar coefficients of b. So the key here is to decompose these terms in some convenient form.

In Section 7D, dealing with noncancellative shifts, the proofs follow this strategy in spirit, but deviate as we advance through the more and more difficult operators. The main issue here is that we are really dealing with terms of the form $|\langle \mathcal{O}(a, b, f), g\rangle|$, where now the operator $\mathcal{O}$ involves a function b in the *weighted little* $\mathrm{bmo}(\nu)$, and a function a in *unweighted product* BMO. In the most difficult case of partial paraproducts, a is even more complicated because it is essentially a *sequence* of *one-parameter* unweighted BMO functions. In all these cases, the creature ϕ in the last step is really $\phi(a, f, g)$. While in the previous case involving $\phi(f, g)$ it was straightforward to see the correct operators $\mathcal{O}_{1,2}$ to achieve step (3), in this case nothing straightforward seems to work.

There are two key new ideas in these cases: one is to combine the cumbersome remainder term with a cleverly chosen third term, which will make the decompositions easier to handle. The other is to temporarily employ martingale transforms — which works for us because this does not increase the BMO norms. We briefly describe the three situations below. As above, we will be rather nonrigorous about the notation in this expository section. There is plenty of notation later, and the purpose here is just to explain the main ideas and guide the reader through the technical proofs in Section 7D:

(1) *The full standard paraproduct*: Theorem 7.3. This case only requires simple martingale transforms (a_τ and g_τ, which have all nonnegative Haar coefficients), and otherwise follows the strategy outlined above. However, we already start to see the operators $\mathcal{O}_{1,2}$ becoming strange compositions of "standard" operators and unweighted paraproducts, such as

$$S_{\boldsymbol{\mathcal{D}}}\phi \leq (M_S \Pi^*_{a_\tau} g_\tau)(S_{\boldsymbol{\mathcal{D}}} f).$$

(2) *The full mixed paraproduct*: Theorem 7.4. Here we introduce the idea of combining the remainder term $\Pi_{\mathbb{S}f} b - \mathbb{S}\Pi_f b$ with a third term T, and we analyze $(\Pi_{\mathbb{S}f} b - T)$ and $(T - \mathbb{S}\Pi_f b)$ separately. This allows us to express the remainder as

$$\sum [\mathsf{P}_{\mathsf{a}}, \mathsf{p}_{\mathsf{b}}] f + T^{(1,0)}_{a,b} f - T^{(0,1)}_{a,b} f,$$

a sum of *commutators of paraproduct operators*, and a new remainder term. The new remainder has no cancellation properties, so we prove separately that the $T_{a,b}$ operators satisfy

$$|\langle T_{a,b} f, g\rangle| \lesssim \|b\|_{\mathrm{bmo}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')}.$$

Here is where we employ the strategy outlined earlier, combined with a martingale transform a_τ applied to a. Interestingly, this transform depends on the particular argument f of $[b, \mathbb{S}_{\boldsymbol{\mathcal{D}}}] f$. This will be absorbed in the end by the BMO norm of the symbol for $\mathbb{S}_{\boldsymbol{\mathcal{D}}}$, so ultimately the choice of f will not matter.

(3) *The partial paraproducts*: Theorem 7.5. Here we again combine the remainder terms with a third term T, and this time end up with terms of the form $\mathsf{p}_{\mathsf{b}} F$, where F is a term depending on a and f. So we are done if we can show that $\|F\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)}$. Without getting too technical about the notation, we reiterate that here a is not *one function* but rather a *sequence* a_{PQR} of one-parameter unweighted BMO functions. So the difficulty here is that the inner products look something like

$$\langle F, g\rangle = \sum \langle \Pi^*_{a_{PQR}} \tilde{f}, \tilde{g}\rangle,$$

where each summand has its own BMO function! The trick is then to write this as $\sum\langle a_{PQR}, \phi_{PQR}(f,g)\rangle$. The happy ending is that these functions a_{PQR} have uniformly bounded BMO norms, so at this point we apply unweighted one-parameter H^1-BMO duality and we are left to work with $\|S_{\mathcal{D}}\phi(f,g)\|_{L^1(\mathbb{R}^n)}$; this is manageable. In one case, we do have to work with F_τ instead, which is again obtained by applying martingale transforms chosen in terms of f — only this time to each function a_{PQR}.

Finally, we see no reason why this result cannot be generalized to k-parameter Journé operators. The main trouble in such a generalization should be strictly computational, as the number of paraproducts will blow up.

In Section 8 we recall the definition of the mixed $\mathrm{BMO}_{\mathcal{I}}$ classes in between Chang and Fefferman's product BMO and Cotlar and Sadosky's little BMO. In the same way as in [Ou et al. 2016] we deduce a corollary from Theorem 1.1:

Theorem 1.2 (upper bound, iterated, unweighted case). *Let us consider $\mathbb{R}^{\vec{d}}$, $\vec{d}=(d_1,\dots,d_t)$, with a partition $\mathcal{I}=(I_s)_{1\le s\le l}$ of $\{1,\dots,t\}$. Let $b\in\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ and let T_s denote a multiparameter Journé operator acting on functions defined on $\bigotimes_{k\in I_s}\mathbb{R}^{d_k}$. Then we have the estimate*

$$\|[T_1,\dots,[T_l,b],\dots]\|_{L^p(\mathbb{R}^{\vec{d}})\to L^p(\mathbb{R}^{\vec{d}})}\lesssim\|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Coming back to the Bloom setting, we prove the lower estimate below, via a modification of the unweighted one-parameter argument of Coifman, Rochberg and Weiss.

Theorem 1.3 (lower bound). *Let μ,λ be $A_p(\mathbb{R}^n\times\mathbb{R}^n)$ weights, and set $\nu=\mu^{1/p}\lambda^{-1/p}$. Then*

$$\|b\|_{\mathrm{bmo}(\nu)}\lesssim\sup_{1\leqslant k,l\leqslant n}\|[b,R_k^1R_l^2]\|_{L^p(\mu)\to L^p(\lambda)},$$

where R_k^1 and R_l^2 are the Riesz transforms acting in the first and second variables, respectively.

This lower estimate allows us to see the tensor products of Riesz transforms as a representative testing class for all Journé operators.

We point out that in our quest to prove Theorem 1.1, we also obtain a much simplified proof of the following one-weight result for Journé operators, originally due to R. Fefferman:

Theorem 1.4 (weighted inequality for Journé operators). *Let T be a biparameter Journé operator on $\mathbb{R}^{\vec{n}}=\mathbb{R}^{n_1}\otimes\mathbb{R}^{n_2}$. Then T is bounded $L^p(w)\to L^p(w)$ for all $w\in A_p(\mathbb{R}^{\vec{n}})$, $1<p<\infty$.*

A version of Theorem 1.4 was first introduced by R. Fefferman and E. M. Stein [1982], with restrictive assumptions on the kernel. Subsequently the kernel assumptions were weakened significantly by R. Fefferman [1987], at the cost of assuming the weight belongs to the more restrictive class $A_{p/2}$. This was due to the use of his sharp function $T^{\#}f=M_S(f^2)^{1/2}$, where M_S is strong maximal function. Finally, he improved his own result in [Fefferman 1988], where he showed that the A_p class sufficed and obtained the full statement of Theorem 1.4. This was achieved by an involved bootstrapping argument based on his previous result [Fefferman 1987].

Our proof in Section 7E of Theorem 1.4 is significantly simpler. This may seem like a "rough sell" in light of the many pages of highly technical calculations that precede it. However, our proof of Section 7E is only based on one-weight bounds for the biparameter dyadic shifts of the form

$$\|\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} : L^p(w) \to L^p(w)\| \lesssim 1. \tag{1-1}$$

These had to be proved along the way, as part of our proof of the two-weight upper bound for commutators, Theorem 1.1. These one-weight bounds are useful in themselves, and their proofs are not that long: the proof for cancellative shifts, given in (7-2), is easy, and the proof for the noncancellative shifts of partial paraproduct type is given in Proposition 7.6. Once we have (1-1), the proof of Theorem 1.4 follows immediately from Martikainen's representation theorem — just as in the one-parameter case, a weighted bound for Calderón–Zygmund operators follows trivially from Hytönen's representation theorem, once one has the one-weight bounds for the one-parameter dyadic shifts.

The paper is organized as follows. In Section 2 we review the necessary background, both one-parameter and biparameter, and set up the notation. In Section 3 we set up the types of dyadic square functions we will need throughout the rest of the paper. In Section 4, we discuss the weighted and Bloom BMO spaces in the biparameter setting, and use some of these results in Section 5 to prove the lower bound result. Section 6 is dedicated to biparameter paraproducts, which will be crucial in Section 7, which proves the upper bound by an appeal to Martikainen's representation theorem [2012]. Finally, we prove Theorem 1.4.

2. Background and notation

We review some of the basic building blocks of one-parameter dyadic harmonic analysis on $\mathbb{R}^n$, followed by their biparameter versions for $\mathbb{R}^{\vec{n}} := \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$.

2A. ***Dyadic grids on*** $\mathbb{R}^n$**.** Let $\mathcal{D}_0 := \{2^{-k}([0,1)^n + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^n\}$ denote the standard dyadic grid on $\mathbb{R}^n$. For every $\omega = (\omega_j)_{j\in\mathbb{Z}} \in (\{0,1\}^n)^{\mathbb{Z}}$ define the shifted dyadic grid $\mathcal{D}_\omega$:

$$\mathcal{D}_\omega := \{Q \dot{+} \omega : Q \in \mathcal{D}_0\}, \quad \text{where } Q \dot{+} \omega := Q + \sum_{j:2^{-j} < l(Q)} 2^{-j}\omega_j,$$

and $l(Q)$ denotes the side length of a cube Q. The indexing parameter ω is rarely relevant in what follows: it only appears when we are dealing with $\mathbb{E}_\omega$ — expectation with respect to the standard probability measure on the space of parameters ω. In fact, an important feature of the (by now standard) methods we employ in this paper is obtaining upper bounds for dyadic operators that are independent of the choice of dyadic grid. The focus therefore is on the geometrical properties shared by all dyadic grids $\mathcal{D}$ on $\mathbb{R}^n$:

- $P \cap Q \in \{P, Q, \varnothing\}$ for every $P, Q \in \mathcal{D}$.
- The cubes $Q \in \mathcal{D}$ with $l(Q) = 2^{-k}$, for some fixed integer k, partition $\mathbb{R}^n$.

For every $Q \in \mathcal{D}$ and every nonnegative integer k we define:

- $Q^{(k)}$ — the k-th generation ancestor of Q in $\mathcal{D}$, i.e., the unique element of $\mathcal{D}$ which contains Q and has side length $2^k l(Q)$.
- $(Q)_k$ — the collection of k-th generation descendants of Q in $\mathcal{D}$, i.e., the 2^{kn} disjoint subcubes of Q with side length $2^{-k} l(Q)$.

2B. *The Haar system on* $\mathbb{R}^n$. Recall that every dyadic interval I in $\mathbb{R}$ is associated with two Haar functions,

$$h_I^0 := \frac{1}{\sqrt{|I|}}(\mathbb{1}_{I_-} - \mathbb{1}_{I_+}) \quad \text{and} \quad h_I^1 := \frac{1}{\sqrt{|I|}}\,\mathbb{1}_I,$$

the first one being cancellative (it has mean 0). Given a dyadic grid $\mathcal{D}$ on $\mathbb{R}^n$, every dyadic cube $Q = I_1 \times \cdots \times I_n$, where all I_i are dyadic intervals in $\mathbb{R}$ with common length $l(Q)$, is associated with $2^n - 1$ cancellative Haar functions:

$$h_Q^{\epsilon}(x) := h_{I_1 \times \cdots \times I_n}^{(\epsilon_1, \dots, \epsilon_n)}(x_1, \dots, x_n) := \prod_{i=1}^{n} h_{I_i}^{\epsilon_i}(x_i),$$

where $\epsilon \in \{0, 1\}^n \setminus \{(1, \dots, 1)\}$ is the signature of h_Q^{ϵ}. To simplify notation, we assume that signatures are never the identically 1 signature, in which case the corresponding Haar function would be noncancellative. The cancellative Haar functions form an orthonormal basis for $L^2(\mathbb{R}^n)$. We write

$$f = \sum_{Q \in \mathcal{D}} \hat{f}(Q^{\epsilon}) h_Q^{\epsilon},$$

where $\hat{f}(Q^{\epsilon}) := \langle f, h_Q^{\epsilon} \rangle$, $\langle f, g \rangle := \int_{\mathbb{R}^n} fg\,dx$, and summation over ϵ is assumed. We list here some other useful facts which will come in handy later:

- $h_P^{\epsilon}(x)$ is constant on any subcube $Q \in \mathcal{D}$, $Q \subsetneq P$. We denote this value by $h_P^{\epsilon}(Q)$.
- The average of f over a cube $Q \in \mathcal{D}$ may be expressed as

$$\langle f \rangle_Q = \sum_{P \in \mathcal{D}, P \supsetneq Q} \hat{f}(P^{\epsilon}) h_P^{\epsilon}(Q). \tag{2-1}$$

- Then, if $Q \subsetneq R \in \mathcal{D}$,

$$\langle f \rangle_Q - \langle f \rangle_R = \sum_{P \in \mathcal{D}, Q \subsetneq P \subset R} \hat{f}(P^{\epsilon}) h_P^{\epsilon}(Q). \tag{2-2}$$

- For $Q \in \mathcal{D}$,

$$\mathbb{1}_Q(f - \langle f \rangle_Q) = \sum_{P \in \mathcal{D}, P \subset Q} \hat{f}(P^{\epsilon}) h_P^{\epsilon}. \tag{2-3}$$

- For two *distinct* signatures $\epsilon \neq \delta$, define the signature $\epsilon + \delta$ by letting $(\epsilon + \delta)_i$ be 1 if $\epsilon_i = \delta_i$ and 0 otherwise. Note that $\epsilon + \delta$ is distinct from both ϵ and δ, and is not the identically $\vec{1}$ signature. Then

$$h_Q^{\epsilon} h_Q^{\delta} = \frac{1}{\sqrt{Q}} h_Q^{\epsilon+\delta} \qquad \text{if } \epsilon \neq \delta \quad \text{and} \quad h_Q^{\epsilon} h_Q^{\epsilon} = \frac{\mathbb{1}_Q}{|Q|}.$$

Again to simplify notation, we assume throughout this paper that we only write $h_Q^{\epsilon+\delta}$ for *distinct* signatures ϵ and δ.

Given a dyadic grid $\mathcal{D}$, we define the dyadic square function on $\mathbb{R}^n$ by

$$S_{\mathcal{D}} f(x) := \left(\sum_{Q \in \mathcal{D}} |\hat{f}(Q^{\epsilon})|^2 \frac{\mathbb{1}_Q(x)}{|Q|} \right)^{1/2}.$$

Then $\|f\|_p \simeq \|S_{\mathcal{D}} f\|_p$ for all $1 < p < \infty$. We also define the dyadic version of the maximal function:

$$M_{\mathcal{D}} f(x) = \sup_{Q \in \mathcal{D}} \langle |f| \rangle_Q \mathbb{1}_Q(x).$$

2C. $A_p(\mathbb{R}^n)$ *weights.* Let w be a weight on $\mathbb{R}^n$; i.e., w is an almost everywhere positive, locally integrable function. For $1 < p < \infty$, let $L^p(w) := L^p(\mathbb{R}^n; w(x)\,dx)$. For a cube Q in $\mathbb{R}^n$, we let

$$w(Q) := \int_Q w(x)\,dx \quad \text{and} \quad \langle w \rangle_Q := \frac{w(Q)}{|Q|}.$$

We say that w belongs to the Muckenhoupt $A_p(\mathbb{R}^n)$ class provided that

$$[w]_{A_p} := \sup_Q \langle w \rangle_Q \langle w^{1-p'} \rangle_Q^{p-1} < \infty,$$

where p' denotes the Hölder conjugate of p and the supremum above is over all cubes Q in $\mathbb{R}^n$ with sides parallel to the axes. The weight $w' := w^{1-p'}$ is sometimes called the weight "conjugate" to w, because $w \in A_p$ if and only if $w' \in A_{p'}$.

We recall the classical inequalities for the maximal and square functions

$$\|Mf\|_{L^p(w)} \lesssim \|f\|_{L^p(w)} \quad \text{and} \quad \|f\|_{L^p(w)} \simeq \|S_{\mathcal{D}} f\|_{L^p(w)}$$

for all $w \in A_p(\mathbb{R}^n)$, $1 < p < \infty$, where throughout this paper "$A \lesssim B$" denotes $A \le cB$ for some constant c which may depend on the dimensions and the weight w. In dealing with dyadic shifts, we will also need to consider the following shifted dyadic square function: given nonnegative integers i and j, define

$$S_{\mathcal{D}}^{i,j} f(x) := \left[\sum_{R \in \mathcal{D}} \left(\sum_{P \in (R)_i} |\hat{f}(P^\epsilon)| \right)^2 \left(\sum_{Q \in (R)_j} \frac{\mathbb{1}_Q(x)}{|Q|} \right) \right]^{1/2}.$$

It was shown in [Holmes et al. 2017] that

$$\|S_{\mathcal{D}}^{i,j} : L^p(w) \to L^p(w)\| \lesssim 2^{(n/2)(i+j)} \tag{2-4}$$

for all $w \in A_p(\mathbb{R}^n)$, $1 < p < \infty$.

A *martingale transform* on $\mathbb{R}^n$ is an operator of the form

$$f \mapsto f_\tau := \sum_{P \in \mathcal{D}} \tau_P^\epsilon \hat{f}(P^\epsilon) h_P^\epsilon,$$

where each τ_P^ϵ is either $+1$ or -1. Obviously $S_{\mathcal{D}} f = S_{\mathcal{D}} f_\tau$, so one can work with f_τ instead when convenient, without increasing the $L^p(w)$-norm of f.

2D. *The Haar system on* $\mathbb{R}^{\vec{n}}$. In $\mathbb{R}^{\vec{n}} := \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, we work with dyadic rectangles

$$\boldsymbol{\mathcal{D}} := \mathcal{D}_1 \times \mathcal{D}_2 = \{R = Q_1 \times Q_2 : Q_i \in \mathcal{D}_i\},$$

where each $\mathcal{D}_i$ is a dyadic grid on $\mathbb{R}^{n_i}$. While we unfortunately lose the nice nestedness and partitioning properties of one-parameter dyadic grids, we do have the tensor product Haar wavelet orthonormal basis for $L^2(\mathbb{R}^{\vec{n}})$, defined by

$$h_R^{\vec{\epsilon}}(x_1, x_2) := h_{Q_1}^{\epsilon_1}(x_1) \otimes h_{Q_2}^{\epsilon_2}(x_2)$$

for all $R = Q_1 \times Q_2 \in \boldsymbol{\mathcal{D}}$ and $\vec{\epsilon} = (\epsilon_1, \epsilon_2)$. We often write

$$f = \sum_{Q_1 \times Q_2} \hat{f}(Q_1^{\epsilon_1} \times Q_2^{\epsilon_2}) h_{Q_1}^{\epsilon_1} \otimes h_{Q_2}^{\epsilon_2},$$

short for summing over $Q_1 \in \mathcal{D}_1$ and $Q_2 \in \mathcal{D}_2$, and of course over all signatures, where

$$\hat{f}(Q_1^{\epsilon_1} \times Q_2^{\epsilon_2}) := \langle f, h_{Q_1}^{\epsilon_1} \otimes h_{Q_2}^{\epsilon_2} \rangle = \int_{\mathbb{R}^{\vec{n}}} f(x_1, x_2) h_{Q_1}^{\epsilon_1}(x_1) h_{Q_2}^{\epsilon_2}(x_2)\, dx_1\, dx_2.$$

While the averaging formula (2-1) has a straightforward biparameter analogue

$$\langle f \rangle_{Q_1 \times Q_2} = \sum_{\substack{P_1 \supsetneq Q_1 \\ P_2 \supsetneq Q_2}} \hat{f}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) h_{P_1}^{\epsilon_1}(Q_1) h_{P_2}^{\epsilon_2}(Q_2),$$

the expression in (2-3) takes a slightly messier form in two parameters: for any $R = Q_1 \times Q_2$

$$\begin{aligned}
&\mathbb{1}_R(f - \langle f \rangle_R) \\
&= \sum_{\substack{P_1 \subset Q_1 \\ P_2 \subset Q_2}} \hat{f}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) h_{P_1}^{\epsilon_1} \otimes h_{P_2}^{\epsilon_2} + \sum_{P_2 \subset Q_2} \left\langle f, \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes h_{P_2}^{\epsilon_2} \right\rangle \mathbb{1}_{Q_1} \otimes h_{P_2}^{\epsilon_2} + \sum_{P_1 \subset Q_1} \left\langle f, h_{P_1}^{\epsilon_1} \otimes \frac{\mathbb{1}_{Q_2}}{|Q_2|} \right\rangle h_{P_1}^{\epsilon_1} \otimes \mathbb{1}_{Q_2} \\
&= \sum_{\substack{P_1 \subset Q_1 \\ P_2 \subset Q_2}} \hat{f}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) h_{P_1}^{\epsilon_1} \otimes h_{P_2}^{\epsilon_2} + \mathbb{1}_R[m_{Q_1} f(x_2) - \langle f \rangle_R] + \mathbb{1}_R[m_{Q_2} f(x_1) - \langle f \rangle_R], \qquad (2\text{-}5)
\end{aligned}$$

where for any cubes $Q_i \in \mathcal{D}_i$,

$$m_{Q_1} f(x_2) := \frac{1}{|Q_1|} \int_{Q_1} f(x_1, x_2)\, dx_1 \quad \text{and} \quad m_{Q_2} f(x_1) := \frac{1}{|Q_2|} \int_{Q_2} f(x_1, x_2)\, dx_2. \qquad (2\text{-}6)$$

As we shall see later, this particular expression will be quite relevant for biparameter BMO spaces.

2E. $A_p(\mathbb{R}^{\vec{n}})$ *weights*. A weight $w(x_1, x_2)$ on $\mathbb{R}^{\vec{n}}$ belongs to the class $A_p(\mathbb{R}^{\vec{n}})$ for some $1 < p < \infty$, provided that

$$[w]_{A_p} := \sup_R \langle w \rangle_R \langle w^{1-p'} \rangle_R^{p-1} < \infty,$$

where the supremum is over all *rectangles* R. These are the weights which characterize $L^p(w)$ boundedness of the *strong* maximal function

$$M_S f(x_1, x_2) := \sup_R \langle |f| \rangle_R \mathbb{1}_R(x_1, x_2),$$

where the supremum is again over all rectangles. As is well known, the usual weak (1, 1) inequality fails for the strong maximal function, where it is replaced by an Orlicz norm expression. In the weighted case, we have [Bagby and Kurtz 1985] for all $w \in A_p(\mathbb{R}^{\vec{n}})$,

$$w\{x \in \mathbb{R}^{\vec{n}} : M_S f(x) > \lambda\} \lesssim \int_{\mathbb{R}^{\vec{n}}} \left(\frac{|f(x)|}{\lambda}\right)^p \left(1 + \log^+ \frac{|f(x)|}{\lambda}\right)^{k-1} dw(x). \qquad (2\text{-}7)$$

Moreover, w belongs to $A_p(\mathbb{R}^{\vec{n}})$ if and only if w belongs to the *one-parameter* classes $A_p(\mathbb{R}^{n_i})$ in each variable separately and uniformly:

$$[w]_{A_p(\mathbb{R}^{\vec{n}})} \simeq \max\Big\{\operatorname*{ess\,sup}_{x_1\in\mathbb{R}^{n_1}}[w(x_1,\cdot)]_{A_p(\mathbb{R}^{n_2})},\ \operatorname*{ess\,sup}_{x_2\in\mathbb{R}^{n_2}}[w(\cdot,x_2)]_{A_p(\mathbb{R}^{n_1})}\Big\}.$$

It also follows, as in the one-parameter case, that $w \in A_p(\mathbb{R}^{\vec{n}})$ if and only if $w' := w^{1-p'} \in A_{p'}(\mathbb{R}^{\vec{n}})$ and $L^p(w)^* \simeq L^{p'}(w')$, in the sense that

$$\|f\|_{L^p(w)} = \sup\{|\langle f, g\rangle| : g \in L^{p'}(w'),\ \|g\|_{L^{p'}(w')} \le 1\}. \tag{2-8}$$

We may also define weights $m_{Q_1}w$ and $m_{Q_2}w$ on $\mathbb{R}^{n_2}$ and $\mathbb{R}^{n_1}$, respectively, as in (2-6). As shown below, these are then also uniformly in their respective one-parameter A_p classes:

Proposition 2.1. *If $w \in A_p(\mathbb{R}^{\vec{n}})$, $1 < p < \infty$, then $m_{Q_1}w \in A_p(\mathbb{R}^{n_2})$ and $m_{Q_2}w \in A_p(\mathbb{R}^{n_1})$ for any cubes $Q_i \subset \mathbb{R}^{n_i}$, with uniformly bounded A_p constants*:

$$[m_{Q_i}w]_{A_p(\mathbb{R}^{n_j})} \le [w]_{A_p(\mathbb{R}^{\vec{n}})}$$

for all $Q_i \subset \mathbb{R}^{n_i}$, $i \in \{1,2\}$, $i \ne j$.

Proof. Fix a cube $Q_1 \subset \mathbb{R}^{n_1}$. Then for every $x_2 \in \mathbb{R}^{n_2}$,

$$|Q_1| = \int_{Q_1} 1\,dx_1 \le \bigg(\int_{Q_1} w(x_1,x_2)\,dx_1\bigg)^{1/p}\bigg(\int_{Q_1} w'(x_1,x_2)\,dx_1\bigg)^{1/p'},$$

and so

$$(m_{Q_1}w)'(x_2) := (m_{Q_1}w)^{1-p'}(x_2) \le m_{Q_1}w'(x_2).$$

Then for all cubes $Q_2 \subset \mathbb{R}^{n_2}$,

$$\langle m_{Q_1}w\rangle_{Q_2}\langle (m_{Q_1}w)'\rangle_{Q_2}^{p-1} \le \langle w\rangle_{Q_1\times Q_2}\langle w'\rangle_{Q_1\times Q_2}^{p-1} \le [w]_{A^p(\mathbb{R}^{\vec{n}})},$$

proving the result for $m_{Q_1}w$. The other case follows symmetrically. □

Finally, we will later use a reverse Hölder property of biparameter A_p weights. This is well known to experts, but we include a proof here for completeness.

Proposition 2.2. *If $w \in A_p(\mathbb{R}^{\vec{n}})$, then there exist positive constants $C, \epsilon, \delta > 0$ (depending only on $\vec{n}$, p, and $[w]_{A_p(\mathbb{R}^{\vec{n}})}$) such that*:

(i) *For all rectangles $R \subset \mathbb{R}^{\vec{n}}$,*

$$\bigg(\frac{1}{|R|}\int_R w(x)^{1+\epsilon}\,dx\bigg)^{1/(1+\epsilon)} \le \frac{C}{|R|}\int_R w(x)\,dx.$$

(ii) *For all rectangles $R \subset \mathbb{R}^{\vec{n}}$ and all measurable subsets $E \subset R$,*

$$\frac{w(E)}{w(R)} \le C\bigg(\frac{|E|}{|R|}\bigg)^{\delta}.$$

Proof. Note first that (ii) follows easily from (i) by applying the Hölder inequality with exponents $1+\epsilon$ and $(1+\epsilon)/\epsilon$ in $w(E)=\int_E w(x)\,dx$. This gives (ii) with $\delta=\epsilon/(1+\epsilon)$.

In order to prove (i) we first recall a more general statement of the one-parameter reverse Hölder property of A_p weights (see Remark 9.2.3 in [Grafakos 2004]):

For any $1<p<\infty$ and $B>1$, there exist positive constants

$$D=D(n,p,B) \quad \text{and} \quad \beta=\beta(n,p,B) \tag{2-9}$$

such that for all $v\in A_p(\mathbb{R}^n)$ with $[v]_{A_p(\mathbb{R}^{\vec{n}})}\le B$, the reverse Hölder condition

$$\left(\frac{1}{|Q|}\int_Q v(t)^{1+\beta}\,dt\right)^{1/(1+\beta)} \le \frac{D}{|Q|}\int_Q v(t)\,dt \tag{2-10}$$

holds for all cubes $Q\subset\mathbb{R}^n$.

It is easy to see that if a weight v satisfies the reverse Hölder condition (2-10) with constants D,β, then it also satisfies it with any constants C,ϵ with $C\ge D$ and $\epsilon\le\beta$.

Now let $w\in A_p(\mathbb{R}^{\vec{n}})$, set $B:=[w]_{A_p(\mathbb{R}^{\vec{n}})}$, and for $i\in\{1,2\}$ let $D_i:=D(n_i,p,B)$ and $\beta_i:=\beta(n_i,p,B)$ be as in (2-9). Fix a rectangle $R=Q_1\times Q_2$, a measurable subset $E\subset R$, and set

$$C^2:=\max(D_1,D_2) \quad \text{and} \quad \epsilon:=\min(\beta_1,\beta_2).$$

For almost all $x_1\in\mathbb{R}^{n_1}$, we have $w(x_1,\cdot)\in A_p(\mathbb{R}^{n_2})$ with $[w(x_1,\cdot)]_{A_p(\mathbb{R}^{n_2})}\le B$, so $w(x_1,\cdot)$ satisfies reverse Hölder with constants D_2,β_2 — and therefore also with constants $\sqrt{C},\epsilon$. So

$$\begin{aligned}\frac{1}{|R|}\int_R w(x)^{1+\epsilon}\,dx &= \frac{1}{|Q_1|}\int_{Q_1}\left(\frac{1}{|Q_2|}w(x_1,x_2)^{1+\epsilon}\,dx_2\right)dx_1\\ &\le \frac{1}{|Q_1|}\int_{Q_1}\left(\frac{\sqrt{C}}{|Q_2|}\int_{Q_2}w(x_1,x_2)\,dx_2\right)^{1+\epsilon}dx_1\\ &= \frac{C^{(1+\epsilon)/2}}{|Q_1|}\int_{Q_1}(m_{Q_2}w(x_1))^{1+\epsilon}\,dx_1.\end{aligned}$$

By Proposition 2.1, we have $m_{Q_2}w\in A_p(\mathbb{R}^{n_1})$ with $[m_{Q_2}w]_{A_p(\mathbb{R}^{n_1})}\le B$, so this weight satisfies reverse Hölder with constants D_1,β_1 — and therefore also with constants $\sqrt{C},\epsilon$. Then the last inequality above gives

$$\left(\frac{1}{|R|}\int_R w(x)^{1+\epsilon}\,dx\right)^{1/(1+\epsilon)} \le \frac{C}{|Q_1|}\int_{Q_1}m_{Q_2}w(x_1)\,dx_1=\frac{C}{|R|}\int_R w(x)\,dx. \qquad \square$$

3. Biparameter dyadic square functions

Throughout this section, fix dyadic rectangles $\boldsymbol{\mathcal{D}}:=\mathcal{D}_1\times\mathcal{D}_2$ on $\mathbb{R}^{\vec{n}}$. The dyadic square function associated with $\boldsymbol{\mathcal{D}}$ is then defined in the obvious way:

$$S_{\boldsymbol{\mathcal{D}}}f(x_1,x_2):=\left(\sum_{R\in\boldsymbol{\mathcal{D}}}|\hat{f}(R^{\vec{\epsilon}})|^2\frac{\mathbb{1}_R(x_1,x_2)}{|R|}\right)^{1/2}.$$

We also want to look at the dyadic square functions *in each variable*, namely

$$S_{\mathcal{D}_1} f(x_1,x_2) := \Big(\sum_{Q_1\in\mathcal{D}_1} |H^{\epsilon_1}_{Q_1} f(x_2)|^2 \frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\Big)^{1/2}, \quad S_{\mathcal{D}_2} f(x_1,x_2) := \Big(\sum_{Q_2\in\mathcal{D}_2} |H^{\epsilon_2}_{Q_2}(x_1)|^2 \frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\Big)^{2},$$

where for every $Q_i \in \mathcal{D}_i$ and signatures ϵ_i, we define

$$H^{\epsilon_1}_{Q_1} f(x_2) := \int_{\mathbb{R}^{n_1}} f(x_1,x_2) h^{\epsilon_1}_{Q_1}(x_1)\,dx_1, \quad H^{\epsilon_2}_{Q_2} f(x_1) := \int_{\mathbb{R}^{n_2}} f(x_1,x_2) h^{\epsilon_2}_{Q_2}(x_2)\,dx_2.$$

Then for any $w \in A_p(\mathbb{R}^{\vec{n}})$,

$$\|f\|_{L^p(w)} \simeq \|S_{\boldsymbol{\mathcal{D}}} f\|_{L^p(w)} \simeq \|S_{\mathcal{D}_1} f\|_{L^p(w)} \simeq \|S_{\mathcal{D}_2} f\|_{L^p(w)}.$$

More generally, define the shifted biparameter square function, for pairs $\vec{i} = (i_1, i_2)$ and $\vec{j} = (j_1, j_2)$ of nonnegative integers, by

$$S^{\vec{i},\vec{j}}_{\boldsymbol{\mathcal{D}}} f := \Bigg[\sum_{\substack{R_1\in\mathcal{D}_1\\ R_2\in\mathcal{D}_2}} \Big(\sum_{\substack{P_1\in(R_1)_{i_1}\\ P_2\in(R_2)_{i_2}}} |\hat{f}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})|\Big)^2 \Big(\sum_{\substack{Q_1\in(R_1)_{j_1}\\ Q_2\in(R_2)_{j_2}}} \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\Big)\Bigg]^{1/2}. \tag{3-1}$$

We claim that

$$\|S^{\vec{i},\vec{j}}_{\boldsymbol{\mathcal{D}}} : L^p(w) \to L^p(w)\| \lesssim 2^{(n_1/2)(i_1+j_1)} 2^{(n_2/2)(i_2+j_2)} \tag{3-2}$$

for all $w \in A_p(\mathbb{R}^{\vec{n}})$, $1 < p < \infty$. This follows by iteration of the one-parameter result in (2-4), through the following vector-valued version of the extrapolation theorem (see Corollary 9.5.7 in [Grafakos 2004]):

Proposition 3.1. *Suppose that an operator T satisfies $\|T : L^2(w) \to L^2(w)\| \le AC_n[w]_{A_2}$ for all $w \in A_2(\mathbb{R}^n)$, for some constants A and C_n, where the latter only depends on the dimension. Then*

$$\Big\|\Big(\sum_j |Tf_j|^2\Big)^{1/2}\Big\|_{L^p(w)} \le AC'_n[w]_{A_p}^{\max(1,1/(p-1))} \Big\|\Big(\sum_j |f_j|^2\Big)^{1/2}\Big\|_{L^p(w)}$$

for all $w \in A_p(\mathbb{R}^n)$, $1 < p < \infty$, and all sequences $\{f_j\} \subset L^p(w)$, where C'_n is a dimensional constant.

Proof of (3-2). Note that $(S^{\vec{i},\vec{j}}_{\boldsymbol{\mathcal{D}}} f)^2 = \sum_{R_1\in\mathcal{D}_1} (S^{i_2,j_2}_{\mathcal{D}_2} F_{R_1})^2$, where

$$F_{R_1}(x_1,x_2) := \sum_{P_2\in\mathcal{D}_2} \Big(\sum_{P_1\in(R_1)_{i_1}} |\hat{f}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})|\Big)\Big(\sum_{Q_1\in(R_1)_{j_1}} \frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\Big)^{1/2} h^{\epsilon_2}_{P_2}(x_2).$$

Then

$$\|S^{\vec{i},\vec{j}}_{\boldsymbol{\mathcal{D}}} f\|^p_{L^p(w)} = \int_{\mathbb{R}^{n_1}}\int_{\mathbb{R}^{n_2}} \Big(\sum_{R_1\in\mathcal{D}_1} (S^{i_2,j_2}_{\mathcal{D}_2} F_{R_1}(x_1,x_2))^2\Big)^{p/2} w(x_1,x_2)\,dx_2\,dx_1.$$

For almost all fixed $x_1 \in \mathbb{R}^{n_1}$, we know $w(x_1,\cdot\,)$ is in $A_p(\mathbb{R}^{n_2})$ uniformly, so we may apply Proposition 3.1 and (2-4) to the inner integral and obtain

$$\|S^{\vec{i},\vec{j}}_{\boldsymbol{\mathcal{D}}} f\|^p_{L^p(w)} \lesssim 2^{(pn_2/2)(i_2+j_2)} \int_{\mathbb{R}^{n_1}}\int_{\mathbb{R}^{n_2}} \Big(\sum_{R_1\in\mathcal{D}_1} |F_{R_1}(x_1,x_2)|^2\Big)^{p/2} w(x_1,x_2)\,dx_2\,dx_1.$$

Now, we can express the integral above as

$$\int_{\mathbb{R}^{n_2}}\int_{\mathbb{R}^{n_1}} (S^{i_1,j_1}_{\mathcal{D}_1} f_\tau(x_1,x_2))^p w(x_1,x_2)\,dx_1\,dx_2 \lesssim 2^{(pn_1/2)(i_1+j_1)}\|f_\tau\|^p,$$

where

$$f_\tau = \sum_{P_1\times P_2} |\hat{f}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})| h_{P_1}^{\epsilon_1}\otimes h_{P_2}^{\epsilon_2}$$

is just a biparameter martingale transform applied to f, and therefore $\|f\|_{L^p(w)} \simeq \|f_\tau\|_{L^p(w)}$ by passing to the square function. □

3A. ***Mixed square and maximal functions.*** We will later encounter mixed operators such as

$$[SM]f(x_1,x_2) := \bigg(\sum_{Q_1\in\mathcal{D}_1} (M_{\mathcal{D}_2}(H^{\epsilon_1}_{Q_1} f)(x_2))^2 \frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\bigg)^{1/2},$$

$$[MS]f(x_1,x_2) := \bigg(\sum_{Q_2\in\mathcal{D}_2} (M_{\mathcal{D}_1}(H^{\epsilon_2}_{Q_2} f)(x_1))^2 \frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\bigg)^{1/2}.$$

Next we show that these operators are bounded $L^p(w)\to L^p(w)$ for all $w\in A_p(\mathbb{R}^{\vec{n}})$. The proof only relies on the fact that the one-parameter maximal function satisfies a weighted bound. So we state the result in a slightly more general form below, replacing $M_{\mathcal{D}_2}$ and $M_{\mathcal{D}_1}$ by any one-parameter operator that satisfies a weighted bound.

Proposition 3.2. *Let T denote a (one-parameter) operator acting on functions on $\mathbb{R}^n$ that satisfies $\|T : L^2(v)\to L^2(v)\| \le C$ for all $v\in A_2(\mathbb{R}^n)$. Define the following operators on $\mathbb{R}^{\vec{n}}$:*

$$[ST]f(x_1,x_2) := \bigg(\sum_{Q_1\in\mathcal{D}_1} (T(H^{\epsilon_1}_{Q_1} f)(x_2))^2 \frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\bigg)^{1/2},$$

$$[TS]f(x_1,x_2) := \bigg(\sum_{Q_2\in\mathcal{D}_2} (T(H^{\epsilon_2}_{Q_2} f)(x_1))^2 \frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\bigg)^{1/2},$$

where T acts on $\mathbb{R}^{n_2}$ in the first operator, and on $\mathbb{R}^{n_1}$ in the second. Then $[ST]$ and $[TS]$ are bounded $L^p(w)\to L^p(w)$ for all $w\in A_p(\mathbb{R}^{\vec{n}})$.

Proof. We have

$$\begin{aligned}\|[ST]f\|^p_{L^p(w)} &= \int_{\mathbb{R}^{n_1}}\int_{\mathbb{R}^{n_2}}\bigg(\sum_{Q_1\in\mathcal{D}_1}\bigg(T(H^{\epsilon_1}_{Q_1})(x_2)\frac{\mathbb{1}_{Q_1}(x_1)}{\sqrt{|Q_1|}}\bigg)^2\bigg)^{p/2} w(x_1,x_2)\,dx_2\,dx_1\\ &\lesssim \int_{\mathbb{R}^{n_1}}\int_{\mathbb{R}^{n_2}}\bigg(\sum_{Q_1\in\mathcal{D}_1}(H^{\epsilon_1}_{Q_1})^2(x_2)\frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\bigg)^{p/2} w(x_1,x_2)\,dx_2\,dx_1\\ &= \|S_{\mathcal{D}_1} f\|^p_{L^p(w)} \lesssim \|f\|^p_{L^p(w)},\end{aligned}$$

where the first inequality follows as before from Proposition 3.1. The proof for $[TS]$ is symmetrical. □

More generally, define *shifted* versions of these mixed operators:

$$[ST]^{i_1,j_1} f(x_1,x_2) := \Bigg(\sum_{R_1\in\mathcal{D}_1}\bigg(\sum_{P_1\in(R_1)_{i_1}} T(H^{\epsilon_1}_{P_1} f)(x_2)\bigg)^2 \sum_{Q_1\in(R_1)_{j_1}} \frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\Bigg)^{1/2},$$

$$[TS]^{i_2,j_2} f(x_1,x_2) := \Bigg(\sum_{R_2\in\mathcal{D}_2}\bigg(\sum_{P_2\in(R_2)_{i_2}} T(H^{\epsilon_2}_{P_2} f)(x_1)\bigg)^2 \sum_{Q_2\in(R_2)_{j_2}} \frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\Bigg)^{1/2}.$$

Under the same assumptions on T, it is easy to see that

$$\|[ST]^{i_1,j_1} : L^p(w)\to L^p(w)\| \lesssim 2^{(n_1/2)(i_1+j_1)} \quad\text{and}\quad \|[TS]^{i_2,j_2} : L^p(w)\to L^p(w)\| \lesssim 2^{(n_2/2)(i_2+j_2)} \tag{3-3}$$

for all $w\in A_p(\mathbb{R}^{\vec{n}})$. Specifically,

$$\|[ST]^{i_1,j_1} f\|^p_{L^p(w)} = \int |S^{i_1,j_1}_{\mathcal{D}_1} F(x_1,x_2)|^p\,dw, \quad \text{where } F(x_1,x_2) := \sum_{P_1\in\mathcal{D}_1} T(H^{\epsilon_1}_{P_1} f)(x_2) h^{\epsilon_1}_{P_1}(x_1),$$

so $\|[ST]^{i_1,j_1} f\|_{L^p(w)} \lesssim 2^{(n_1/2)(i_1+j_1)}\|F\|_{L^p(w)}$. Now,

$$\|F\|_{L^p(w)} \simeq \|S_{\mathcal{D}_1} F\|_{L^p(w)} = \|[ST]f\|_{L^p(w)} \lesssim \|f\|_{L^p(w)}.$$

4. Biparameter weighted BMO spaces

Given a weight w on $\mathbb{R}^n$, a locally integrable function b is said to be in the weighted $\mathrm{BMO}(w)$ space if

$$\|b\|_{\mathrm{BMO}(w)} := \sup_Q \frac{1}{w(Q)} \int_Q |b(x) - \langle b\rangle_Q|\,dx < \infty,$$

where the supremum is over all cubes Q in $\mathbb{R}^n$. If $w=1$, we obtain the unweighted $\mathrm{BMO}(\mathbb{R}^n)$ space. The dyadic version $\mathrm{BMO}_{\mathcal{D}}(w)$ is obtained by only taking the supremum over $Q\in\mathcal{D}$ for some given dyadic grid $\mathcal{D}$ on $\mathbb{R}^n$. If $w\in A_p(\mathbb{R}^n)$ for some $1<p<\infty$, Muckenhoupt and Wheeden [1976] showed that

$$\|b\|_{\mathrm{BMO}(w)} \simeq \|b\|_{\mathrm{BMO}(w';p')} := \sup_Q \bigg(\frac{1}{w(Q)} \int_Q |b-\langle b\rangle_Q|^{p'}\,dw'\bigg)^{1/p'}, \tag{4-1}$$

where w' is the conjugate weight to w. Moreover, if $w\in A_2(\mathbb{R}^n)$, the argument in [Wu 1992] shows that $\mathrm{BMO}_{\mathcal{D}}(w) \simeq H^1_{\mathcal{D}}(w)^*$, where the dyadic Hardy space $H^1_{\mathcal{D}}(w)$ is defined by the norm

$$\|\phi\|_{H^1_{\mathcal{D}}(w)} := \|S_{\mathcal{D}}\phi\|_{L^1(w)}.$$

Then

$$|\langle b,\phi\rangle| \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}(w)} \|S_{\mathcal{D}}\phi\|_{L^1(w)} \quad \text{for all } w\in A_2(\mathbb{R}^n). \tag{4-2}$$

Now suppose μ and λ are $A_p(\mathbb{R}^n)$ weights for some $1<p<\infty$, and define the Bloom weight $\nu := \mu^{1/p}\lambda^{-1/p}$. As shown in [Holmes et al. 2017], we have $\nu\in A_2(\mathbb{R}^n)$, which means we may use (4-2) with ν. A two-weight John–Nirenberg theorem for the Bloom BMO space $\mathrm{BMO}(\nu)$ is also proved in that paper, namely

$$\|b\|_{\mathrm{BMO}(\nu)} \simeq \|b\|_{\mathrm{BMO}(\mu,\lambda,p)} \simeq \|b\|_{\mathrm{BMO}(\lambda',\mu',p')},$$

where

$$\|b\|_{\mathrm{BMO}(\mu,\lambda,p)} := \sup_Q \left(\frac{1}{\mu(Q)} \int_Q |b - \langle b \rangle_Q|^p \, d\lambda \right)^{1/p},$$

$$\|b\|_{\mathrm{BMO}(\lambda',\mu',p')} := \sup_Q \left(\frac{1}{\lambda'(Q)} \int_Q |b - \langle b \rangle_Q|^{p'} \, d\mu' \right)^{1/p'}.$$

We now look at weighted BMO spaces in the product setting $\mathbb{R}^{\vec{n}} = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$. Suppose $w(x_1, x_2)$ is a weight on $\mathbb{R}^{\vec{n}}$. Then we have three BMO spaces:

- *Weighted little* $\mathrm{bmo}(w)$ is the space of all locally integrable functions b on $\mathbb{R}^{\vec{n}}$ such that

$$\|b\|_{\mathrm{bmo}(w)} := \sup_R \frac{1}{w(R)} \int_R |b - \langle b \rangle_R| \, dx < \infty,$$

where the supremum is over all *rectangles* $R = Q_1 \times Q_2$ in $\mathbb{R}^{\vec{n}}$. Given a choice of dyadic rectangles $\boldsymbol{\mathcal{D}} = \mathcal{D}_1 \times \mathcal{D}_2$, we define the dyadic weighted little $\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(w)$ by taking supremum over $R \in \boldsymbol{\mathcal{D}}$.

- *Weighted product* $\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)$ is the space of all locally integrable functions b on $\mathbb{R}^{\vec{n}}$ such that

$$\|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)} := \sup_\Omega \left(\frac{1}{w(\Omega)} \sum_{R \subset \Omega, R \in \boldsymbol{\mathcal{D}}} |\hat{b}(R)|^2 \frac{1}{\langle w \rangle_R} \right)^{1/2} < \infty,$$

where the supremum is over all *open sets* $\Omega \subset \mathbb{R}^{\vec{n}}$ with $w(\Omega) < \infty$.

- *Weighted rectangular* $\mathrm{BMO}_{\boldsymbol{\mathcal{D}},\mathrm{Rec}}(w)$ is defined in a similar fashion to the unweighted case — just like product BMO, but taking the supremum over rectangles instead of over open sets:

$$\|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}},\mathrm{Rec}}(w)} := \sup_R \left(\frac{1}{w(R)} \sum_{T \subset R} |\hat{b}(T^\epsilon)|^2 \frac{1}{\langle w \rangle_T} \right)^{1/2},$$

where the supremum is over all rectangles R, and the summation is over all subrectangles $T \in \boldsymbol{\mathcal{D}},\ T \subset R$.

We have the inclusions

$$\mathrm{bmo}_{\mathcal{D}}(w) \subsetneq \mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w) \subsetneq \mathrm{BMO}_{\boldsymbol{\mathcal{D}},\mathrm{Rec}}(w).$$

Let us look more closely at some of these spaces.

4A. *Weighted product* $\mathbf{BMO}_{\boldsymbol{\mathcal{D}}}(w)$. As in the one-parameter case, we define the dyadic weighted Hardy space $\mathcal{H}^1_{\boldsymbol{\mathcal{D}}}(w)$ to be the space of all $\phi \in L^1(w)$ such that $S_{\boldsymbol{\mathcal{D}}}\phi \in L^1(w)$, a Banach space under the norm $\|\phi\|_{\mathcal{H}^1_{\boldsymbol{\mathcal{D}}}(w)} := \|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(w)}$. The following result exists in the literature in various forms, but we include a proof here for completeness.

Proposition 4.1. *With the notation above*, $\mathcal{H}^1_{\boldsymbol{\mathcal{D}}}(w)^* \equiv \mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)$. *Specifically*, *every* $b \in \mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)$ *determines a continuous linear functional on* $\mathcal{H}^1_{\boldsymbol{\mathcal{D}}}(w)$ *by* $\phi \mapsto \langle b, \phi \rangle$,

$$|\langle b, \phi \rangle| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)} \|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(w)}, \tag{4-3}$$

and, *conversely*, *every* $L \in \mathcal{H}^1_{\boldsymbol{\mathcal{D}}}(w)^*$ *may be realized as* $L\phi = \langle b, \phi \rangle$ *for some* $b \in \mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)$.

Proof. To prove the first statement, let $b \in \mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(w)$ and $\phi \in \mathcal{H}^1_{\boldsymbol{\mathcal{D}}}(w)$. For every $j \in \mathbb{Z}$, define the set $U_j := \{x \in \mathbb{R}^{\vec{n}} : S_{\boldsymbol{\mathcal{D}}}\phi(x) > 2^j\}$, and the collection of rectangles $\mathcal{R}_j := \bigl\{R \in \boldsymbol{\mathcal{D}} : w(R \cap U_j) > \frac{1}{2}w(R)\bigr\}$. Clearly $U_{j+1} \subset U_j$ and $\mathcal{R}_{j+1} \subset \mathcal{R}_j$. Moreover,

$$\sum_{j\in\mathbb{Z}} 2^j w(U_j) \simeq \|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(w)}, \tag{4-4}$$

which comes from the measure-theoretical fact that for any integrable function f on a measure space $(\mathcal{X}, \mu)$, we have $\|f\|_{L^1(\mu)} \simeq \sum_{j\in\mathbb{Z}} 2^j \mu\{x \in \mathcal{X} : |f(x)| > 2^j\}$.

As shown in Proposition 2.2, there exist $C, \delta > 0$ such that $w(E)/w(R) \le C(|E|/|R|)^\delta$ for all rectangles R and measurable subsets $E \subset R$. Define then for every $j \in \mathbb{Z}$ the (open) set

$$V_j := \{x \in \mathbb{R}^{\vec{n}} : M_S \mathbb{1}_{U_j}(x) > \theta\}, \quad \text{where } \theta := \left(\frac{1}{2C}\right)^{1/\delta}.$$

First note that if $R \in \mathcal{R}_j$, then

$$\frac{1}{2} < \frac{w(R \cap U_j)}{w(R)} \le C\left(\frac{|R \cap U_j|}{|R|}\right)^\delta$$

so

$$\theta < \langle \mathbb{1}_{U_j}\rangle_R \le M_S \mathbb{1}_{U_j}(x) \quad \text{for all } x \in R.$$

Therefore

$$\bigcup_{R\in\mathcal{R}_j} R \subset V_j. \tag{4-5}$$

Using (2-7), we have

$$w(V_j) \lesssim \int_{U_j} \frac{1}{\theta^p}\left(1 + \log^+ \frac{1}{\theta}\right)^{k-1} dw \simeq w(U_j). \tag{4-6}$$

Now suppose $R \in \boldsymbol{\mathcal{D}}$ but $R \notin \bigcup_{j\in\mathbb{Z}} \mathcal{R}_j$. Then $w(R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi \le 2^j\}) \ge \frac{1}{2}w(R)$ for all $j \in \mathbb{Z}$, and so

$$w(R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi = 0\}) = w\left(\bigcap_{j=1}^\infty R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi \le 2^{-j}\}\right) \ge \tfrac{1}{2}w(R).$$

Then $|\{S_{\boldsymbol{\mathcal{D}}}\phi = 0\}| \ge |R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi = 0\}| \ge \theta|R| > 0$, and we may write

$$|\hat{\phi}(R)|^2 = \int_{\{S_{\boldsymbol{\mathcal{D}}}\phi=0\}} |\hat{\phi}(R)|^2 \frac{\mathbb{1}_R}{|R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi = 0\}|}\, dx \le \frac{1}{\theta}\int_{\{S_{\boldsymbol{\mathcal{D}}}\phi=0\}} (S_{\boldsymbol{\mathcal{D}}}\phi)^2\, dx = 0.$$

So

$$\hat{\phi}(R) = 0 \quad \text{for all } R \in \boldsymbol{\mathcal{D}},\ R \notin \bigcup_{j\in\mathbb{Z}} \mathcal{R}_j. \tag{4-7}$$

Finally, if $R \in \bigcap_{j\in\mathbb{Z}} \mathcal{R}_j$, then

$$0 = w(R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi = \infty\}) = \lim_{j\to\infty} w(R \cap \{S_{\boldsymbol{\mathcal{D}}}\phi > 2^j\}) \ge \tfrac{1}{2}w(R),$$

a contradiction. In light of this and (4-7),

$$\sum_{R\in\mathcal{D}} |\hat{b}(R)|\,|\hat{\phi}(R)| = \sum_{j\in\mathbb{Z}} \sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{b}(R)|\,|\hat{\phi}(R)|$$
$$\le \sum_{j\in\mathbb{Z}} \biggl(\sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{b}(R)|^2 \frac{1}{\langle w\rangle_R}\biggr)^{1/2} \biggl(\sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{\phi}(R)|^2 \langle w\rangle_R\biggr)^{1/2}.$$

To estimate the first term, we simply note that

$$\sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{b}(R)|^2 \frac{1}{\langle w\rangle_R} \le \sum_{R\in\mathcal{R}_j} |\hat{b}(R)|^2 \frac{1}{\langle w\rangle_R} \le \sum_{R\subset V_j,\, R\in\mathcal{D}} |\hat{b}(R)|^2 \frac{1}{\langle w\rangle_R} \le \|b\|^2_{\mathrm{BMO}_{\mathcal{D}}(w)} w(V_j),$$

where the second inequality follows from (4-5). For the second term, note that any $R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}$ satisfies $R \subset V_j$ and $w(R \setminus U_{j+1}) \ge \frac{1}{2} w(R)$. Then

$$\sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{\phi}(R)|^2 \langle w\rangle_R \le 2 \sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{\phi}(R)|^2 \frac{w(R\setminus U_{j+1})}{|R|}$$
$$= 2\int_{V_j\setminus U_{j+1}} \sum_{R\in\mathcal{R}_j\setminus\mathcal{R}_{j+1}} |\hat{\phi}(R)|^2 \frac{\mathbb{1}_R}{|R|}\, dw$$
$$\le 2\int_{V_j\setminus U_{j+1}} (S_{\mathcal{D}}\phi)^2\, dw \lesssim 2^{2j} w(V_j),$$

since $S_{\mathcal{D}}\phi \le 2^{j+1}$ off U_{j+1}. Finally, we have by (4-6),

$$\sum_{R\in\mathcal{D}} |\hat{b}(R)|\,|\hat{\phi}(R)| \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}(w)} \sum_{j\in\mathbb{Z}} 2^j w(V_j) \simeq \|b\|_{\mathrm{BMO}_{\mathcal{D}}(w)} \sum_{j\in\mathbb{Z}} 2^j w(U_j).$$

Combining this with (4-4), we obtain (4-3).

To see the converse, let $L \in \mathcal{H}^1_{\mathcal{D}}(w)$. Then L is given by $L\phi = \langle b, \phi\rangle$ for some function b. Fix an open set Ω with $w(\Omega) < \infty$. Then

$$\biggl(\sum_{R\subset\Omega,\, R\in\mathcal{D}} |\hat{b}(R)|^2 \frac{1}{\langle w\rangle_R}\biggr)^{1/2} \le \sup_{\|\phi\|_{l^2(\Omega,w)}\le 1} \biggl|\sum_{R\subset\Omega,\, R\in\mathcal{D}} \hat{b}(R)\hat{\phi}(R)\biggr|,$$

where $\|\phi\|^2_{l^2(\Omega,w)} := \sum_{R\subset\Omega,\, R\in\mathcal{D}} |\hat{\phi}(R)|^2 \langle w\rangle_R$. By a simple application of Hölder's inequality,

$$\biggl|\sum_{R\subset\Omega,\, R\in\mathcal{D}} \hat{b}(R)\hat{\phi}(R)\biggr| \lesssim \|L\|_\star \|\phi\|_{\mathcal{H}^1_{\mathcal{D}}(w)} \le \|L\|_\star (w(\Omega))^{1/2} \|\phi\|_{l^2(\Omega,w)},$$

so $\|b\|_{\mathrm{BMO}_{\mathcal{D}}(w)} \lesssim \|L\|_\star$. □

4B. *Weighted little* $\mathbf{bmo}_{\mathcal{D}}(w)$. In this case, we also want to look at each variable separately. Specifically, we look at the space $\mathrm{BMO}(w_1, x_2)$: for each $x_2 \in \mathbb{R}^{n_2}$, this is the weighted BMO space over $\mathbb{R}^{n_1}$, with respect to the weight $w(\cdot, x_2)$:

$$\mathrm{BMO}(w_1, x_2) := \mathrm{BMO}(w(\cdot, x_2);\ \mathbb{R}^{n_1}) \quad \text{for each } x_2 \in \mathbb{R}^{n_2}.$$

The norm in this space is given by

$$\|b(\cdot, x_2)\|_{\mathrm{BMO}(w_1,x_2)} := \sup_{Q_1} \frac{1}{w(Q_1, x_2)} \int_{Q_1} |b(x_1, x_2) - m_{Q_1} b(x_2)|\, dx_1,$$

where

$$w(Q_1, x_2) := \int_{Q_1} w(x_1, x_2)\, dx_1 \quad \text{and} \quad m_{Q_1} b(x_2) := \frac{1}{|Q_1|} \int_{Q_1} b(x_1, x_2)\, dx_1.$$

The space $\mathrm{BMO}(w_2, x_1)$ and the quantities $w(Q_2, x_1)$ and $m_{Q_2} b(x_1)$ are defined symmetrically.

Proposition 4.2. *Let $w(x_1, x_2)$ be a weight on $\mathbb{R}^{\vec{n}} = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$. Then $b \in L^1_{\mathrm{loc}}(\mathbb{R}^{\vec{n}})$ is in* $\mathrm{bmo}(w)$ *if and only if b is in the one-parameter weighted BMO spaces* $\mathrm{BMO}(w_i, x_j)$ *separately in each variable, uniformly*:

$$\|b\|_{\mathrm{bmo}(w)} \simeq \max\Big\{ \operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{n_1}} \|b(x_1, \cdot)\|_{\mathrm{BMO}(w_2,x_1)},\ \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{n_2}} \|b(\cdot, x_2)\|_{\mathrm{BMO}(w_1,x_2)} \Big\}.$$

Remark 4.3. In the *unweighted* case $\mathrm{bmo}(\mathbb{R}^{\vec{n}})$, if we fixed $x_2 \in \mathbb{R}^{n_2}$, we would look at $b(\cdot, x_2)$ in the space $\mathrm{BMO}(\mathbb{R}^{n_1})$ — the *same* one-parameter BMO space for all x_2. In the weighted case however, the one-parameter space for $b(\cdot, x_2)$ *changes with* x_2, because the weight $w(\cdot, x_2)$ changes with x_2.

Proof. Suppose first that $b \in \mathrm{bmo}(w)$. Then for all cubes Q_1, Q_2,

$$\begin{aligned} \|b\|_{\mathrm{bmo}(w)} &\geq \frac{1}{w(Q_1 \times Q_2)} \int_{Q_1} \int_{Q_2} |b(x_1, x_2) - \langle b \rangle_{Q_1 \times Q_2}|\, dx_2\, dx_1 \\ &\geq \frac{1}{w(Q_1 \times Q_2)} \int_{Q_1} \left| \int_{Q_2} b(x_1, x_2) - \langle b \rangle_{Q_1 \times Q_2}\, dx_2 \right| dx_1, \end{aligned}$$

so

$$\int_{Q_1} |m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2}|\, dx_1 \leq \frac{w(Q_1 \times Q_2)}{|Q_2|} \|b\|_{\mathrm{bmo}(w)}. \tag{4-8}$$

Now fix a cube Q_2 in $\mathbb{R}^{n_2}$ and let $f_{Q_2}(x_1) := \int_{Q_2} |b(x_1, x_2) - m_{Q_2} b(x_1)|\, dx_2$. Then for any Q_1,

$$\begin{aligned} \langle f_{Q_2} \rangle_{Q_1} &\leq \frac{1}{|Q_1|} \int_{Q_1} \int_{Q_2} |b(x_1, x_2) - \langle b \rangle_{Q_1 \times Q_2}|\, dx + \frac{1}{|Q_1|} \int_{Q_1} \int_{Q_2} |m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2}|\, dx \\ &\leq \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{\mathrm{bmo}(w)} + \frac{|Q_2|}{|Q_1|} \int_{Q_1} |m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2}|\, dx_1 \\ &\leq 2 \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{\mathrm{bmo}(w)} = 2 \langle w(Q_2, \cdot) \rangle_{Q_1} \|b\|_{\mathrm{bmo}(w)}, \end{aligned}$$

where the last inequality follows from (4-8). By the Lebesgue differentiation theorem,

$$f_{Q_2}(x_1) = \lim_{Q_1 \to x_1} \langle f_{Q_2} \rangle_{Q_1} \leq 2 \|b\|_{\mathrm{bmo}(w)} \lim_{Q_1 \to x_1} \langle w(Q_2, \cdot) \rangle_{Q_1} = 2 \|b\|_{\mathrm{bmo}(w)} w(Q_2, x_1)$$

for almost all $x_1 \in \mathbb{R}^{n_1}$, where $Q_1 \to x_1$ denotes a sequence of cubes containing x_1 with side length tending to 0.

We would like to say at this point that $\|b(x_1,\cdot)\|_{\mathrm{BMO}(w_2,x_1)} = \sup_{Q_2} 1/(w(Q_2,x_1)) f_{Q_2}(x_1)$ is uniformly (a.a. x_1) bounded. However, we must be a little careful and note that at this point we really have that for every cube Q_2 in $\mathbb{R}^{n_2}$, there is a *null set* $N(Q_2) \subset \mathbb{R}^{n_1}$ such that

$$f_{Q_2}(x_1) \le 2\|b\|_{\mathrm{bmo}(w)} w(Q_2, x_1) \quad \text{for all } x_1 \in \mathbb{R}^{n_1} \setminus N(Q_2).$$

In order to obtain the inequality we want, holding for a.a. x_1, let $N := \bigcup N(\widetilde{Q}_2)$ where $\widetilde{Q}_2$ are the cubes in $\mathbb{R}^{n_2}$ with rational side length and centers with rational coordinates. Then N is a null set and $f_{\widetilde{Q}_2}(x_1) \le 2\|b\|_{\mathrm{bmo}(w)} w(\widetilde{Q}_2, x_1)$ for all $x_1 \in \mathbb{R}^{n_1} \setminus N$. By density, this statement then holds for *all* cubes Q_2 and $x_1 \notin N$, so

$$\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{n_1}} \|b(x_1,\cdot)\|_{\mathrm{BMO}(w_2,x_1)} \le 2\|b\|_{\mathrm{bmo}(w)}.$$

The result for the other variable follows symmetrically.

Conversely, suppose

$$\|b(x_1,\cdot)\|_{\mathrm{BMO}(w_2,x_1)} \le C_1 \quad \text{for a.a. } x_1, \qquad \|b(\cdot,x_2)\|_{\mathrm{BMO}(w_1,x_2)} \le C_2 \quad \text{for a.a. } x_2.$$

Then for any $R = Q_1 \times Q_2$,

$$\begin{aligned}
\int_R |b - \langle b\rangle_R|\,dx &\le \int_{Q_1}\int_{Q_2} |b(x_1,x_2) - m_{Q_2}(x_1)|\,dx + \int_{Q_1} |Q_2|\,|m_{Q_2}b(x_1) - \langle b\rangle_{Q_1\times Q_2}|\,dx_1 \\
&\le \int_{Q_1} C_2 w(Q_2,x_1)\,dx_1 + \int_{Q_1}\int_{Q_2} |b(x_1,x_2) - m_{Q_1}b(x_2)|\,dx_2\,dx_1 \\
&\le C_2 w(R) + \int_{Q_2} C_1 w(Q_1,x_2)\,dx_2 \\
&= (C_1 + C_2) w(R),
\end{aligned}$$

so

$$\|b\|_{\mathrm{bmo}(w)} \le 2\max\Bigl\{\operatorname*{ess\,sup}_{x_1\in\mathbb{R}^{n_1}} \|b(x_1,\cdot)\|_{\mathrm{BMO}(w_2,x_1)},\ \operatorname*{ess\,sup}_{x_2\in\mathbb{R}^{n_2}} \|b(\cdot,x_2)\|_{\mathrm{BMO}(w_1,x_2)}\Bigr\}. \qquad \square$$

Corollary 4.4. *Let $w \in A_2(\mathbb{R}^{\vec{n}})$ and $b \in \mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(w)$. Then*

$$|\langle b, \phi\rangle| \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(w)} \|S_{\mathcal{D}_i}\phi\|_{L^1(w)}$$

for all $i \in \{1, 2\}$.

Proof. This follows immediately from the one-parameter result in (4-2) and the proposition above:

$$\begin{aligned}
|\langle b,\phi\rangle| &\le \int_{\mathbb{R}^{n_1}} |\langle b(x_1,\cdot), \phi(x_1,\cdot)\rangle_{\mathbb{R}^{n_2}}|\,dx_1 \\
&\lesssim \int_{\mathbb{R}^{n_1}} \|b(x_1,\cdot)\|_{\mathrm{BMO}_{\mathcal{D}_2}(w(x_1,\cdot))} \|S_{\mathcal{D}_2}\phi(x_1,\cdot)\|_{L^1(w(x_1,\cdot))}\,dx_1 \\
&\lesssim \|b\|_{\mathrm{bmo}(w)} \|S_{\mathcal{D}_2}\phi\|_{L^1(w)},
\end{aligned}$$

and similarly for $S_{\mathcal{D}_1}$. $\square$

We now look at the little bmo version of (4-1).

Proposition 4.5. *If* $w \in A_p(\mathbb{R}^{\vec{n}})$ *for some* $1 < p < \infty$, *then*

$$\|b\|_{\mathrm{bmo}(w)} \simeq \|b\|_{\mathrm{bmo}(w;\,p')} := \sup_R \biggl(\frac{1}{w(R)} \int_R |b - \langle b\rangle_R|^{p'}\, dw'\biggr)^{1/p'}.$$

Proof. By Proposition 4.2 and (4-1),

$$\|b\|_{\mathrm{bmo}(w)} \simeq \max\Bigl\{\operatorname*{ess\,sup}_{x_1\in\mathbb{R}^{n_1}} \|b(x_1,\cdot\,)\|_{\mathrm{BMO}(w(x_1,\cdot);\,p')},\ \operatorname*{ess\,sup}_{x_2\in\mathbb{R}^{n_2}} \|b(\,\cdot\,,x_2)\|_{\mathrm{BMO}(w(\,\cdot\,,x_2);\,p')}\Bigr\}.$$

Suppose first that $b \in \mathrm{bmo}(w;\,p')$. Note that for some function g on $\mathbb{R}^{\vec{n}}$ and a cube Q_2 in $\mathbb{R}^{n_2}$, we have

$$\int_{Q_2} |g(x_1,x_2)|^{p'} w'(x_1,x_2)\, dx_2 \ge \frac{1}{w(Q_2,x_1)^{p'-1}} \biggl|\int_{Q_2} g(x_1,x_2)\, dx_2\biggr|^{p'}.$$

Then

$$\begin{aligned}
\|b\|^{p'}_{\mathrm{bmo}(w;\,p')} &\ge \frac{1}{w(R)} \int_{Q_1} \frac{1}{w(Q_2,x_1)^{p'-1}} \biggl|\int_{Q_2} b(x_1,x_2) - \langle b\rangle_{Q_1\times Q_2}\, dx_2\biggr|^{p'} dx_1\\
&= \frac{1}{w(R)} \int_{Q_1} |m_{Q_2}b(x_1) - \langle b\rangle_{Q_1\times Q_2}|^{p'} \frac{|Q_2|^{p'}}{w(Q_2,x_1)^{p'-1}}\, dx_1\\
&\ge \frac{1}{w(R)} \int_{Q_1} |m_{Q_2}b(x_1) - \langle b\rangle_{Q_1\times Q_2}|^{p'} w'(Q_2,x_1)\, dx_1,
\end{aligned}$$

where the last inequality follows from

$$\frac{|Q_2|^{p'}}{w(Q_2,x_1)^{p'-1}} = |Q_2| \frac{1}{\langle w(x_1,\cdot\,)\rangle_{Q_2}^{p'-1}} \ge |Q_2| \frac{\langle w'(x_1,\cdot)\rangle_{Q_2}}{[w(x_1,\cdot\,)]_{A_p}^{p'-1}} \simeq w'(Q_2,x_1).$$

Now fix Q_2 and consider $f_{Q_2}(x_1) := \int_{Q_2} |b(x_1,x_2) - m_{Q_2}b(x_1)|^{p'} w'(x_1,x_2)\, dx_2$. Then

$$\begin{aligned}
\langle f_{Q_2}\rangle_{Q_1} &\lesssim \frac{1}{|Q_1|} \int_{Q_1}\int_{Q_2} \bigl(|b(x_1,x_2) - \langle b\rangle_{Q_1\times Q_2}|^{p'} + |m_{Q_2}b(x_1) - \langle b\rangle_{Q_1\times Q_2}|^{p'}\bigr) w'(x_1,x_2)\, dx_2\, dx_1\\
&\lesssim \frac{w(Q_1\times Q_2)}{|Q_1|} \|b\|^{p'}_{\mathrm{bmo}(w;\,p')} + \frac{1}{|Q_1|} \int_{Q_1} |m_{Q_2}b(x_1) - \langle b\rangle_{Q_1\times Q_2}|^{p'} w'(Q_2,x_1)\, dx_1\\
&\lesssim \frac{w(Q_1\times Q_2)}{|Q_1|} \|b\|^{p'}_{\mathrm{bmo}(w;\,p')}.
\end{aligned}$$

Then for almost all x_1,

$$f_{Q_2}(x_1) = \lim_{Q_1\to x_1} \langle f_{Q_2}\rangle_{Q_1} \lesssim \lim_{Q_1\to x_1} \frac{w(Q_1\times Q_2)}{|Q_1|} \|b\|^{p'}_{\mathrm{bmo}(w;\,p')} = w(Q_2,x_1) \|b\|^{p'}_{\mathrm{bmo}(w;\,p')}.$$

Taking again rational cubes, we obtain

$$\|b(x_1,\cdot\,)\|_{\mathrm{BMO}(w(x_1,\cdot);\,p')} = \sup_{Q_2} \biggl(\frac{1}{w(Q_2,x_1)} f_{Q_2}(x_1)\biggr)^{1/p'} \lesssim \|b\|_{\mathrm{bmo}(w;\,p')}$$

for almost all x_1.

Conversely, if $b \in \mathrm{bmo}(w)$, then there exist C_1 and C_2 such that

$$\|b(x_1, \cdot\,)\|_{\mathrm{BMO}(w(x_1,\cdot\,);\,p')} \leq C_1 \quad \text{for a.a. } x_1, \qquad \text{and} \qquad \|b(\,\cdot\,, x_2)\|_{\mathrm{BMO}(w(\,\cdot\,,x_2);\,p')} \leq C_2 \quad \text{for a.a. } x_2.$$

Then

$$\int_R |b - \langle b\rangle_R|^{p'}\, dw' \lesssim \int_{Q_1}\int_{Q_2} |b(x_1, x_2) - m_{Q_2} b(x_1)|^{p'} w'(x_1, x_2)\, dx_2\, dx_1 \\ + \int_{Q_1}\int_{Q_2} |m_{Q_2} b(x_1) - \langle b\rangle_{Q_1\times Q_2}|^{p'} w'(x_1, x_2)\, dx_2\, dx_1.$$

The first integral is easily seen to be bounded by

$$\int_{Q_1} \|b(x_1, \cdot\,)\|^{p'}_{\mathrm{BMO}(w(x_1,\cdot\,))} w(Q_2, x_1)\, dx_1 \leq C_1^{p'} w(Q_1 \times Q_2).$$

The second integral is equal to

$$\begin{aligned}\int_{Q_1} &|m_{Q_2} b(x_1) - \langle b\rangle_{Q_1\times Q_2}|^{p'} w'(Q_2, x_1)\, dx_1 \\ &\leq \int_{Q_1} \frac{w'(Q_2, x_1)}{|Q_2|^{p'}} \biggl(\int_{Q_2} |b(x_1, x_2) - m_{Q_1} b(x_2)|\, dx_2\biggr)^{p'} dx_1 \\ &\leq \int_{Q_1} \frac{w'(Q_2, x_1) w(Q_2, x_1)^{p'-1}}{|Q_2|^{p'}} \int_{Q_2} |b(x_1, x_2) - m_{Q_1} b(x_2)|^{p'} w'(x_1, x_2)\, dx_2\, dx_1.\end{aligned}$$

We may express the first term as $\langle w'(x_1, \cdot\,)\rangle_{Q_2} \langle w(x_1, \cdot\,)\rangle_{Q_2}^{p'-1} \lesssim [w]_{A_p}^{p'-1}$ for almost all x_1. Then, the integral is further bounded by

$$\int_{Q_2} w(Q_1, x_2) \|b(\,\cdot\,, x_2)\|_{\mathrm{BMO}(w(\,\cdot\,,x_2);\,p')}\, dx_2 \lesssim C_2^{p'} w(Q_1 \times Q_2).$$

Finally, this gives

$$\|b\|_{\mathrm{bmo}(w;\,p')} \lesssim (C_1^{p'} + C_2^{p'})^{1/p'} \lesssim \max(C_1, C_2) \simeq \|b\|_{\mathrm{bmo}(w)}. \qquad \square$$

We also have a two-weight John–Nirenberg for Bloom little bmo, which follows very similarly to the proof above.

Proposition 4.6. *Let $\mu, \lambda \in A_p(\mathbb{R}^{\vec{n}})$ for $1 < p < \infty$, and $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|b\|_{\mathrm{bmo}(\nu)} \simeq \|b\|_{\mathrm{bmo}(\mu,\lambda,p)} \simeq \|b\|_{\mathrm{bmo}(\lambda',\mu',p')},$$

where

$$\|b\|_{\mathrm{bmo}(\mu,\lambda,p)} := \sup_R \biggl(\frac{1}{\mu(R)} \int_R |b - \langle b\rangle_R|^p\, d\lambda\biggr)^{1/p},$$

$$\|b\|_{\mathrm{bmo}(\lambda',\mu',p')} := \sup_R \biggl(\frac{1}{\lambda'(R)} \int_R |b - \langle b\rangle_R|^{p'}\, d\mu'\biggr)^{1/p'}.$$

Note that it also easily follows that $\nu \in A_2(\mathbb{R}^{\vec{n}})$.

5. Proof of the lower bound

Proof of Theorem 1.3. To see the lower bound, we adapt the argument of Coifman, Rochberg and Weiss [Coifman et al. 1976]. Let $\{X_k(x)\}$ and $\{Y_l(y)\}$ both be orthonormal bases for the space of spherical harmonics of degree n in $\mathbb{R}^n$. Then $\sum_k |X_k(x)|^2 = c_n |x|^{2n}$ and thus

$$1 = \frac{1}{c_n} \sum_k \frac{X_k(x-x')}{|x-x'|^{2n}} X_k(x-x')$$

and similarly for Y_l.

Furthermore $X_k(x-x') = \sum_{|\alpha|+|\beta|=n} \boldsymbol{x}^{(k)}_{\alpha\beta} x^\alpha x'^\beta$ and similarly for Y_l. Remember that

$$b(x,y) \in \mathrm{bmo}(\nu) \iff \|b\|_{\mathrm{bmo}(\nu)} = \sup_Q \frac{1}{\nu(Q)} \int_Q |b(x,y) - \langle b \rangle_Q| \, dx \, dy < \infty.$$

Here, $Q = I \times J$ and I and J are cubes in $\mathbb{R}^n$. Let us define the function

$$\Gamma_Q(x,y) = \mathrm{sign}(b(x,y) - \langle b \rangle_Q) \mathbb{1}_Q(x,y).$$

So

$$\begin{aligned}
&|b(x,y) - \langle b \rangle_Q| |Q| \mathbb{1}_Q(x,y) \\
&\quad = (b(x,y) - \langle b \rangle_Q) |Q| \Gamma_Q(x,y) \\
&\quad = \int_Q (b(x,y) - b(x',y')) \Gamma_Q(x,y) \, dx' \, dy' \\
&\quad \sim \sum_{k,l} \int_Q (b(x,y) - b(x',y')) \frac{X_k(x-x')}{|x-x'|^{2n}} X_k(x-x') \frac{Y_l(y-y')}{|y-y'|^{2n}} Y_l(y-y') \Gamma_Q(x,y) \, dx' \, dy' \\
&\quad = \sum_{k,l} \int_{\mathbb{R}^{2n}} \frac{b(x,y) - b(x',y')}{|x-x'|^{2n} |y-y'|^{2n}} X_k(x-x') Y_l(y-y') \cdot \\
&\qquad\qquad \cdot \sum_{|\alpha|+|\beta|=n} \boldsymbol{x}^{(k)}_{\alpha\beta} x^\alpha x'^\beta \sum_{|\gamma|+|\delta|=n} \boldsymbol{y}^{(l)}_{\gamma\delta} y^\gamma y'^\delta \Gamma_Q(x,y) \mathbb{1}_Q(x',y') \, dx' \, dy'.
\end{aligned}$$

Note that

$$\int_{\mathbb{R}^{2n}} \frac{b(x,y) - b(x',y')}{|x-x'|^{2n} |y-y'|^{2n}} X_k(x-x') Y_l(y-y') \cdot x'^\beta y'^\delta \mathbb{1}_Q(x',y') \, dx' \, dy' = [b, T_k T_l](x'^\beta y'^\delta \mathbb{1}_Q(x',y')).$$

Here T_k and T_l are the Calderón–Zygmund operators that correspond to the kernels

$$\frac{X_k(x)}{|x|^{2n}} \quad \text{and} \quad \frac{Y_l(y)}{|y|^{2n}}.$$

Observe that these have the correct homogeneity due to the homogeneity of the X_k and Y_l. With this notation, the above becomes

$$\begin{aligned}
&|b(x,y) - \langle b \rangle_Q| |Q| \mathbb{1}_Q(x,y) \\
&\qquad = \sum_{k,l} \sum_{|\alpha|+|\beta|=n} \sum_{|\gamma|+|\delta|=n} \boldsymbol{x}^{(k)}_{\alpha\beta} x^\alpha \boldsymbol{y}^{(l)}_{\gamma\delta} y^\gamma \Gamma_Q(x,y) [b, T_k T_l](x'^\beta y'^\delta \mathbb{1}_Q(x',y'))(x,y).
\end{aligned}$$

Now, we integrate with respect to (x, y) and the measure λ. Let us assume for a moment that both I and J are centered at 0 and thus Q is centered at 0. In this case, since Γ_Q and $\mathbb{1}_Q$ are supported in Q, there is only contribution for x, x', y, y' in Q:

$$
\begin{aligned}
|Q|\biggl(\int_Q |b(x,y)-\langle b\rangle_Q|^p\,d\lambda(x,y)\biggr)^{1/p}
&\leqslant \sum_{k,l}\sum_{|\alpha|+|\beta|=n}\sum_{|\gamma|+|\delta|=n} \bigl\|\boldsymbol{x}^{(k)}_{\alpha\beta}x^\alpha \boldsymbol{y}^{(l)}_{\gamma\delta}y^\gamma\Gamma_Q(x,y)[b,T_kT_l](x'^\beta y'^\delta\mathbb{1}_Q(x',y'))(x,y)\bigr\|_{L^p(\lambda)}\\
&\lesssim \sum_{k,l}\sum_{|\alpha|+|\beta|=n}\sum_{|\gamma|+|\delta|=n} \mathfrak{l}(I)^{|\alpha|}\mathfrak{l}(J)^{|\gamma|}\bigl\|[b,T_kT_l](x'^\beta y'^\delta\mathbb{1}_Q(x',y'))\bigr\|_{L^p(\lambda)}\\
&\lesssim \sum_{k,l}\sum_{|\alpha|+|\beta|=n}\sum_{|\gamma|+|\delta|=n} \mathfrak{l}(I)^{|\alpha|}\mathfrak{l}(J)^{|\gamma|}\|[b,T_kT_l]\|_{L^p(\mu)\to L^p(\lambda)}\|x'^\beta y'^\delta\mathbb{1}_Q(x',y')\|_{L^p(\mu)}\\
&\lesssim \sum_{k,l}\sum_{|\alpha|+|\beta|=n}\sum_{|\gamma|+|\delta|=n} \mathfrak{l}(I)^{|\alpha|}\mathfrak{l}(J)^{|\gamma|}\mathfrak{l}(I)^{|\beta|}\mathfrak{l}(J)^{|\delta|}\|[b,T_kT_l]\|_{L^p(\mu)\to L^p(\lambda)}\mu(Q)^{1/p}.
\end{aligned}
$$

We disregarded the coefficients of X and Y at the cost of a constant.

Notice that the T_k and T_l are homogeneous polynomials in Riesz transforms. Therefore the commutator $[b, T_kT_l]$ can be written as a linear combination of terms such as $M[b, R^1_i R^2_j]N$, where M and N are compositions of Riesz transforms: First write $[b, T_kT_l]$ as linear combination of terms of the form $[b, R^k_{(n)}R^l_{(n)}]$, where

$$R^k_{(n)} = \prod_s R^1_{i^{(k)}_s}$$

is a composition of n Riesz transforms acting in the first variable with the choice $i^{(k)} = (i^{(k)}_s)^n_{s=1} \in \{1,\dots,n\}^n$ for each k and similarly for $R^l_{(n)}$ acting in the second variable Then, for each term, apply $[AB, b] = A[B, b] + [A, b]B$ successively as follows. Use $A = R^1_{i_1}R^2_{j_1}$ and B of the form $R^k_{(n-1)}R^l_{(n-1)}$ and repeat. It then follows that for each k, l the commutator $[b, T_kT_l]$ can be written as a linear combination of terms such as $M[b, R^1_i R^2_j]N$, where M and N are compositions of Riesz transforms. Thus T_k and T_l are homogeneous polynomials in Riesz transforms of the same degree. We require that all commutators of the form $[b, R^1_i R^2_j]$ are bounded, and we have shown the bmo estimate for b for rectangles Q whose sides are centered at 0. We now translate b in the two directions separately and obtain what we need, by Proposition 4.6:

$$\|b\|_{\mathrm{bmo}(\nu)} \simeq \|b\|_{\mathrm{bmo}(\mu,\lambda,p)} := \sup_R\biggl(\frac{1}{\mu(R)}\int_R |b-\langle b\rangle_R|^p\,d\lambda\biggr)^{1/p} \lesssim \sup_{1\leqslant k,l\leqslant n}\|[b, R^1_k R^2_l]\|_{L^p(\mu)\to L^p(\lambda)}. \qquad \square$$

6. Biparameter paraproducts

Decomposing two functions b and f on $\mathbb{R}^n$ into their Haar series adapted to some dyadic grid $\mathcal{D}$ and analyzing the different inclusion properties of the dyadic cubes, one may express their product as

$$bf = \Pi_b f + \Pi^*_b f + \Gamma_b f + \Pi_f b,$$

where

$$\Pi_b f := \sum_{Q\in\mathcal{D}} \hat{b}(Q^\epsilon)\langle f\rangle_Q h_Q^\epsilon, \quad \Pi_b^* f := \sum_{Q\in\mathcal{D}} \hat{b}(Q^\epsilon)\hat{f}(Q^\epsilon)\frac{\mathbb{1}_Q}{|Q|}, \quad \Gamma_b f := \sum_{Q\in\mathcal{D}} \hat{b}(Q^\epsilon)\hat{f}(Q^\delta)\frac{1}{\sqrt{|Q|}} h_Q^{\epsilon+\delta}.$$

In [Holmes et al. 2017], it was shown that, when $b \in \mathrm{BMO}(\nu)$, the operators Π_b, Π_b^*, and Γ_b are bounded $L^p(\mu) \to L^p(\lambda)$.

6A. *Product BMO paraproducts.* In the biparameter setting $\boldsymbol{\mathcal{D}} = \mathcal{D}_1 \times \mathcal{D}_2$, we have *fifteen* paraproducts. We treat them beginning with the nine paraproducts associated with product BMO. First, we have the three "pure" paraproducts, direct adaptations of the one-parameter paraproducts:

$$\begin{aligned}
\Pi_b f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\langle f\rangle_{Q_1\times Q_2} h_{Q_1}^{\epsilon_1}\otimes h_{Q_2}^{\epsilon_2},\\
\Pi_b^* f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\hat{f}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|},\\
\Gamma_b f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\hat{f}(Q_1^{\delta_1}\times Q_2^{\delta_2})\frac{1}{\sqrt{|Q_1|}}\frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{\epsilon_1+\delta_1}\otimes h_{Q_2}^{\epsilon_2+\delta_2} = \Gamma_b^* f.
\end{aligned}$$

Next, we have the "mixed" paraproducts. We index these based on the types of Haar functions acting on f, since the action on b is the same for all of them, namely $\hat{b}(Q_1 \times Q_2)$ — this is the property which associates these paraproducts with product $\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}$: in a proof using duality, one would separate out the b function and be left with the biparameter square function $S_{\boldsymbol{\mathcal{D}}}$. They are

$$\begin{aligned}
\Pi_{b;(0,1)} f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\left\langle f, h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2},\\
\Pi_{b;(1,0)} f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\left\langle f, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2}\right\rangle h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|} = \Pi_{b;(0,1)}^*,\\
\Gamma_{b;(0,1)} f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\left\langle f, h_{Q_1}^{\delta_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle\frac{1}{\sqrt{|Q_1|}} h_{Q_1}^{\epsilon_1+\delta_1}\otimes h_{Q_2}^{\epsilon_2},\\
\Gamma_{b;(0,1)}^* f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\hat{f}(Q_1^{\delta_1}\times Q_2^{\epsilon_2})\frac{1}{\sqrt{|Q_1|}} h_{Q_1}^{\epsilon_1+\delta_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|},\\
\Gamma_{b;(1,0)} f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\left\langle f, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\delta_2}\right\rangle\frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{\epsilon_1}\otimes h_{Q_2}^{\epsilon_2+\delta_2},\\
\Gamma_{b;(1,0)}^* f &:= \sum_{Q_1\times Q_2} \hat{b}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\hat{f}(Q_1^{\epsilon_1}\times Q_2^{\delta_2})\frac{1}{\sqrt{|Q_2|}}\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2+\delta_2}.
\end{aligned}$$

Proposition 6.1. *If $\nu := \mu^{1/p}\lambda^{-1/p}$ for $A_p(\mathbb{R}^{\vec{n}})$ weights μ and λ, and $\mathsf{P_b}$ denotes any one of the nine paraproducts defined above, then*

$$\|\mathsf{P_b} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)}, \tag{6-1}$$

where $\|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)}$ denotes the norm of b in the dyadic weighted product $\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)$ space on $\mathbb{R}^{\vec{n}}$.

Proof. We first outline the general strategy we use to prove (6-1). From (2-8), it suffices to take $f \in L^p(\mu)$ and $g \in L^{p'}(\lambda')$ and show that

$$|\langle \mathsf{P}_{\mathsf{b}} f, g\rangle| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')}.$$

(1) Write $\langle \mathsf{P}_{\mathsf{b}} f, g\rangle = \langle b, \phi\rangle$, where ϕ depends on f and g. By (4-3), $|\langle \mathsf{P}_{\mathsf{b}} f, g\rangle| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(\nu)}$.

(2) Show that $S_{\boldsymbol{\mathcal{D}}}\phi \lesssim (\mathcal{O}_1 f)(\mathcal{O}_2 g)$, where $\mathcal{O}_1$ and $\mathcal{O}_2$ are operators satisfying a *one-weight bound* $L^p(w) \to L^p(w)$ for all $w \in A_p(\mathbb{R}^{\vec{n}})$ — these operators will usually be a combination of maximal and square functions.

(3) Then the $L^1(\nu)$-norm of $S_{\boldsymbol{\mathcal{D}}}\phi$ can be separated into the $L^p(\mu)$ and $L^{p'}(\lambda')$ norms of these operators $\mathcal{O}_i$ by a simple application of Hölder's inequality,

$$\|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(\nu)} \lesssim \|\mathcal{O}_1 f\|_{L^p(\mu)} \|\mathcal{O}_2 g\|_{L^{p'}(\lambda')} \lesssim \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')},$$

and the result follows.

Remark also that we will not have to treat the adjoints $\mathsf{P}^*_{\mathsf{b}}$ separately: interchanging the roles of f and g in the proof strategy above will show that P_{b} is also bounded $L^{p'}(\lambda') \to L^{p'}(\mu')$, which means that $\mathsf{P}^*_{\mathsf{b}}$ is bounded $L^p(\mu) \to L^p(\lambda)$.

Let us begin with $\Pi_b f$. We write

$$\langle \Pi_b f, g\rangle = \langle b, \phi\rangle, \quad \text{where } \phi := \sum_{Q_1 \times Q_2} \langle f\rangle_{Q_1 \times Q_2} \hat{g}(Q_1^{\epsilon_1} \times Q_2^{\epsilon_2}) h_{Q_1}^{\epsilon_1} \otimes h_{Q_2}^{\epsilon_2}.$$

Then

$$(S_{\boldsymbol{\mathcal{D}}}\phi)^2 \le \sum_{Q_1 \times Q_2} \langle |f|\rangle^2_{Q_1 \times Q_2} |\hat{g}(Q_1^{\epsilon_1} \times Q_2^{\epsilon_2})|^2 \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes \frac{\mathbb{1}_{Q_2}}{|Q_2|} \le (M_S f)^2 \cdot (S_{\boldsymbol{\mathcal{D}}} g)^2,$$

so

$$|\langle \Pi_b f, g\rangle| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|M_S f\|_{L^p(\mu)} \|S_{\boldsymbol{\mathcal{D}}} g\|_{L^{p'}(\lambda')} \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')}.$$

Note that if we take instead $f \in L^{p'}(\lambda')$ and $g \in L^p(\mu)$, we have

$$|\langle \Pi_b f, g\rangle| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|M_S f\|_{L^{p'}(\lambda')} \|S_{\boldsymbol{\mathcal{D}}} g\|_{L^p(\mu)} \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|f\|_{L^{p'}(\lambda')} \|g\|_{L^p(\mu)},$$

proving that $\|\Pi_b : L^{p'}(\lambda') \to L^{p'}(\mu')\| = \|\Pi_b^* : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)}$. For Γ_b,

$$\langle \Gamma_b f, g\rangle = \langle b, \phi\rangle, \quad \text{where } \phi := \sum_{Q_1 \times Q_2} \hat{f}(Q_1^{\epsilon_1} \times Q_2^{\epsilon_2}) \hat{g}(Q_1^{\delta_1} \times Q_2^{\delta_2}) \frac{1}{\sqrt{|Q_1|}} \frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{\epsilon_1+\delta_1} \otimes h_{Q_2}^{\epsilon_2+\delta_2},$$

from which it easily follows that $S_{\boldsymbol{\mathcal{D}}}\phi \lesssim S_{\boldsymbol{\mathcal{D}}} f \cdot S_{\boldsymbol{\mathcal{D}}} g$.

Let us now look at $\Pi_{b;(0,1)}$. In this case,

$$\phi := \sum_{Q_1 \times Q_2} \left\langle f, h_{Q_1}^{\epsilon_1} \otimes \frac{\mathbb{1}_{Q_2}}{|Q_2|} \right\rangle \left\langle g, \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes h_{Q_2}^{\epsilon_2} \right\rangle h_{Q_1}^{\epsilon_1} \otimes h_{Q_2}^{\epsilon_2}.$$

Then

$$\begin{aligned}(S_{\mathcal{D}}\phi)^2 &= \sum_{Q_1\times Q_2}\left\langle f, h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle^2\left\langle g, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2}\right\rangle^2\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\\ &= \sum_{Q_1\times Q_2}\langle H_{Q_1}^{\epsilon_1}f\rangle_{Q_2}^2\langle H_{Q_2}^{\epsilon_2}g\rangle_{Q_1}^2\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\\ &\le\left(\sum_{Q_1}(M_{\mathcal{D}_2}H_{Q_1}^{\epsilon_1}f)^2(x_2)\frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\right)\left(\sum_{Q_2}(M_{\mathcal{D}_1}H_{Q_2}^{\epsilon_2}g)^2(x_1)\frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\right)=[SM]^2f\cdot[MS]^2g,\end{aligned}$$

where $[SM]$ and $[MS]$ are the mixed square-maximal operators in Section 3A. Boundedness of $\Pi_{b;(0,1)}$ then follows from Proposition 3.2. By the usual duality trick, the same holds for $\Pi_{b;(1,0)}$. Finally, for $\Gamma_{b;(0,1)}$,

$$\phi=\sum_{Q_1\times Q_2}\langle H_{Q_1}^{\delta_1}f\rangle_{Q_2}\frac{1}{\sqrt{|Q_1|}}\hat{g}(Q_1^{\epsilon_1+\delta_1}\times Q_2^{\epsilon_2})h_{Q_1}^{\epsilon_1}\otimes h_{Q_2}^{\epsilon_2},$$

so $S_{\mathcal{D}}\phi\lesssim[SM]f\cdot S_{\mathcal{D}}g$. Note that $\Gamma_{b;(1,0)}$ works the same way, except we bound $S_{\mathcal{D}}\phi$ by $[MS]f\cdot S_{\mathcal{D}}g$, and the remaining two paraproducts follow by duality. □

6B. ***Little bmo paraproducts.*** Next, we have the six paraproducts associated with little bmo. We denote these by the small Greek letters corresponding to the previous paraproducts, and index them based on the Haar functions acting on b — in this case, separating out the b function will yield one of the square functions $S_{\mathcal{D}_i}$ in one of the variables:

$$\begin{aligned}\pi_{b;(0,1)}f &:= \sum_{Q_1\times Q_2}\left\langle b, h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle\left\langle f, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2}\right\rangle h_{Q_1}^{\epsilon_1}\otimes h_{Q_2}^{\epsilon_2},\\ \pi^*_{b;(0,1)}f &:= \sum_{Q_1\times Q_2}\left\langle b, h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle\hat{f}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2},\\ \pi_{b;(1,0)}f &:= \sum_{Q_1\times Q_2}\left\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2}\right\rangle\left\langle f, h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle h_{Q_1}^{\epsilon_1}\otimes h_{Q_2}^{\epsilon_2},\\ \pi^*_{b;(1,0)}f &:= \sum_{Q_1\times Q_2}\left\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\epsilon_2}\right\rangle\hat{f}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})h_{Q_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|},\\ \gamma_{b;(0,1)}f &:= \sum_{Q_1\times Q_2}\left\langle b, h_{Q_1}^{\delta_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\right\rangle\hat{f}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\frac{1}{\sqrt{|Q_1|}}h_{Q_1}^{\epsilon_1+\delta_1}\otimes h_{Q_2}^{\epsilon_2}=\gamma^*_{b;(0,1)}f,\\ \gamma_{b;(1,0)}f &:= \sum_{Q_1\times Q_2}\left\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{Q_2}^{\delta_2}\right\rangle\hat{f}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\frac{1}{\sqrt{|Q_2|}}h_{Q_1}^{\epsilon_1}\otimes h_{Q_2}^{\epsilon_2+\delta_2}=\gamma^*_{b;(1,0)}f.\end{aligned}$$

Proposition 6.2. *If $\nu:=\mu^{1/p}\lambda^{-1/p}$ for $A_p(\mathbb{R}^{\vec{n}})$ weights μ and λ, and $\mathsf{p_b}$ denotes any one of the six paraproducts defined above, then*

$$\|\mathsf{p_b}:L^p(\mu)\to L^p(\lambda)\|\lesssim\|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)},$$

where $\|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}$ denotes the norm of b in the dyadic weighted little $\mathrm{bmo}_{\mathcal{D}}(\nu)$ space on $\mathbb{R}^{\vec{n}}$.

Proof. The proof strategy is the same as that of the product BMO paraproducts, with the modification that we use one of the $S_{\mathcal{D}_i}$ square functions and Corollary 4.4. For instance, in the case of $\pi_{b;(0,1)}$ we write

$$\langle \pi_{b;(0,1)} f, g\rangle = \langle b, \phi\rangle, \quad \text{where } \phi := \sum_{Q_1\times Q_2} \langle H^{\epsilon_2}_{Q_2} f\rangle_{Q_1} \hat{g}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2}) h^{\epsilon_1}_{Q_1} \otimes \frac{\mathbb{1}_{Q_2}}{|Q_2|}.$$

Then

$$\begin{aligned}(S_{\mathcal{D}_1}\phi)^2 &\le \sum_{Q_1}\Bigg(\sum_{Q_2} \langle |H^{\epsilon_2}_{Q_2} f|\rangle^2_{Q_1}\mathbb{1}_{Q_1}(x_1)\frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\Bigg)\Bigg(\sum_{Q_2}|\hat{g}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})|^2\frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\Bigg)\frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\\ &\le \Bigg(\sum_{Q_2} M^2_{\mathcal{D}_1}(H^{\epsilon_2}_{Q_2} f)(x_1)\frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\Bigg)\Bigg(\sum_{Q_1}\sum_{Q_2}|\hat{g}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})|^2\frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}(x_2)}{|Q_2|}\Bigg)\\ &= [MS]^2 f\cdot S^2_{\boldsymbol{\mathcal{D}}} g,\end{aligned}$$

and so

$$|\langle \pi_{b;(0,1)} f, g\rangle| \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}\|S_{\mathcal{D}_1}\phi\|_{L^1(\nu)} \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}\|f\|_{L^p(\mu)}\|g\|_{L^{p'}(\lambda')}.$$

The proof for $\pi_{b;(1,0)}$ is symmetrical — we take $S_{\mathcal{D}_2}\phi$, which will be bounded by $[SM]f\cdot S_{\boldsymbol{\mathcal{D}}} g$. The adjoint paraproducts $\pi^*_{b;(0,1)}$ and $\pi^*_{b;(1,0)}$ follow again by duality. Finally, for $\gamma_{b;(0,1)}$,

$$\phi := \sum_{Q_1\times Q_2} \hat{f}(Q_1^{\epsilon_1}\times Q_2^{\epsilon_2})\frac{1}{\sqrt{|Q_1|}}\hat{g}(Q_1^{\epsilon_1+\delta_1}\times Q_2^{\epsilon_2}) h^{\epsilon_1}_{Q_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|},$$

from which it easily follows that $S_{\mathcal{D}_1}\phi \le S_{\boldsymbol{\mathcal{D}}} f\cdot S_{\boldsymbol{\mathcal{D}}} g$. The proof for $\gamma_{b;(1,0)}$ is symmetrical. □

7. Commutators with Journé operators

7A. ***Definition of Journé operators.*** We begin with the definition of biparameter Calderón–Zygmund operators, or Journé operators, on $\mathbb{R}^{\vec{n}} := \mathbb{R}^{n_1}\otimes\mathbb{R}^{n_2}$, as outlined in [Martikainen 2012]. As shown later in [Grau de la Herrán 2016], these conditions are equivalent to the original definition of [Journé 1985].

I. Structural assumptions: Given $f = f_1\otimes f_2$ and $g = g_1\otimes g_2$, where $f_i, g_i : \mathbb{R}^{n_i}\to\mathbb{C}$ satisfy $\mathrm{spt}(f_i)\cap\mathrm{spt}(g_i) = \varnothing$ for $i = 1, 2$, we assume the kernel representation

$$\langle Tf, g\rangle = \int_{\mathbb{R}^{\vec{n}}}\int_{\mathbb{R}^{\vec{n}}} K(x, y) f(y) g(x)\, dx\, dy.$$

The kernel $K : \mathbb{R}^{\vec{n}}\times\mathbb{R}^{\vec{n}}\setminus\{(x, y)\in\mathbb{R}^{\vec{n}}\times\mathbb{R}^{\vec{n}} : x_1 = y_1 \text{ or } x_2 = y_2\}\to\mathbb{C}$ is assumed to satisfy:

(1) Size condition:

$$|K(x, y)| \le C\frac{1}{|x_1-y_1|^{n_1}}\frac{1}{|x_2-y_2|^{n_2}}.$$

(2) Hölder conditions:

(a) If $|y_1-y_1'|\le\frac{1}{2}|x_1-y_1|$ and $|y_2-y_2'|\le\frac{1}{2}|x_2-y_2|$, then

$$\big|K(x, y) - K(x, (y_1, y_2')) - K(x, (y_1', y_2)) + K(x, y')\big| \le C\frac{|y_1-y_1'|^\delta}{|x_1-y_1|^{n_1+\delta}}\frac{|y_2-y_2'|^\delta}{|x_2-y_2|^{n_2+\delta}}.$$

(b) If $|x_1 - x_1'| \le \frac{1}{2}|x_1 - y_1|$ and $|x_2 - x_2'| \le \frac{1}{2}|x_2 - y_2|$, then

$$\left|K(x, y) - K((x_1, x_2'), y) - K((x_1', x_2), y) + K(x', y)\right| \le C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{n_2+\delta}}.$$

(c) If $|y_1 - y_1'| \le \frac{1}{2}|x_1 - y_1|$ and $|x_2 - x_2'| \le \frac{1}{2}|x_2 - y_2|$, then

$$\left|K(x, y) - K((x_1, x_2'), y) - K(x, (y_1', y_2)) + K((x_1, x_2'), (y_1', y_2))\right| \le C \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{n_2+\delta}}.$$

(d) If $|x_1 - x_1'| \le \frac{1}{2}|x_1 - y_1|$ and $|y_2 - y_2'| \le \frac{1}{2}|x_2 - y_2|$, then

$$\left|K(x, y) - K(x, (y_1, y_2')) - K((x_1', x_2), y) + K((x_1', x_2), (y_1, y_2'))\right| \le C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{n_2+\delta}}.$$

(3) Mixed size and Hölder conditions:

(a) If $|x_1 - x_1'| \le \frac{1}{2}|x_1 - y_1|$, then

$$|K(x, y) - K((x_1', x_2), y)| \le C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}} \frac{1}{|x_2 - y_2|^{n_2}}.$$

(b) If $|y_1 - y_1'| \le \frac{1}{2}|x_1 - y_1|$, then

$$|K(x, y) - K(x, (y_1', y_2))| \le C \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}} \frac{1}{|x_2 - y_2|^{n_2}}.$$

(c) If $|x_2 - x_2'| \le \frac{1}{2}|x_2 - y_2|$, then

$$|K(x, y) - K((x_1, x_2'), y)| \le C \frac{1}{|x_1 - y_1|^{n_1}} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{n_2+\delta}}.$$

(d) If $|y_2 - y_2'| \le \frac{1}{2}|x_2 - y_2|$, then

$$|K(x, y) - K(x, (y_1, y_2'))| \le C \frac{1}{|x_1 - y_1|^{n_1}} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{n_2+\delta}}.$$

(4) Calderón–Zygmund structure in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ separately: If $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ with $\operatorname{spt}(f_1) \cap \operatorname{spt}(g_1) = \varnothing$, we assume the kernel representation

$$\langle Tf, g\rangle = \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_1}} K_{f_2,g_2}(x_1, y_1) f_1(y_1) g_1(x_1)\, dx_1\, dy_1,$$

where the kernel $K_{f_2,g_2} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_1} \setminus \{(x_1, y_1) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_1} : x_1 = y_1\}$ satisfies the size condition

$$|K_{f_2,g_2}(x_1, y_1)| \le C(f_2, g_2) \frac{1}{|x_1 - y_1|^{n_1}},$$

and Hölder conditions:

(a) If $|x_1 - x_1'| \le \frac{1}{2}|x_1 - y_1|$, then

$$|K_{f_2,g_2}(x_1, y_1) - K_{f_2,g_2}(x_1', y_1)| \le C(f_2, g_2) \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}}.$$

(b) If $|y_1 - y_1'| \le \frac{1}{2}|x_1 - y_1|$, then

$$|K_{f_2,g_2}(x_1, y_1) - K_{f_2,g_2}(x_1, y_1')| \le C(f_2, g_2)\frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}}.$$

We only assume the above representation and a certain control over $C(f_2, g_2)$ in the diagonal; that is,

$$C(\mathbb{1}_{Q_2}, \mathbb{1}_{Q_2}) + C(\mathbb{1}_{Q_2}, u_{Q_2}) + C(u_{Q_2}, \mathbb{1}_{Q_2}) \le C|Q_2|$$

for all cubes $Q_2 \subset \mathbb{R}^{n_2}$ and all "Q_2-adapted zero-mean" functions u_{Q_2} — that is, $\operatorname{spt}(u_{Q_2}) \subset Q_2$, $|u_{Q_2}| \le 1$, and $\int u_{Q_2} = 0$. We assume the symmetrical representation with kernel K_{f_1,g_1} in the case $\operatorname{spt}(f_2) \cap \operatorname{spt}(g_2) = \varnothing$.

II. Boundedness and cancellation assumptions:

(1) Assume $T1$, T^*1, $T_1(1)$ and $T_1^*(1)$ are in product $\mathrm{BMO}(\mathbb{R}^{\vec{n}})$, where T_1 is the partial adjoint of T, defined by

$$\langle T_1(f_1 \otimes f_2), g_1 \otimes g_2\rangle = \langle T(g_1 \otimes f_2), f_1 \otimes g_2\rangle.$$

(2) Assume

$$|\langle T(\mathbb{1}_{Q_1} \otimes \mathbb{1}_{Q_2}), \mathbb{1}_{Q_1} \otimes \mathbb{1}_{Q_2}\rangle| \le C|Q_1||Q_2|$$

for all cubes $Q_i \subset \mathbb{R}^{n_i}$ (weak boundedness).

(3) Diagonal BMO conditions: for all cubes $Q_i \subset \mathbb{R}^{n_i}$ and all zero-mean functions a_{Q_1} and b_{Q_2} that are Q_1- and Q_2-adapted, respectively, assume:

$$|\langle T(a_{Q_1} \otimes \mathbb{1}_{Q_2}), \mathbb{1}_{Q_1} \otimes \mathbb{1}_{Q_2}\rangle| \le C|Q_1||Q_2|, \quad |\langle T(\mathbb{1}_{Q_1} \otimes \mathbb{1}_{Q_2}), a_{Q_1} \otimes \mathbb{1}_{Q_2}\rangle| \le C|Q_1||Q_2|,$$
$$|\langle T(\mathbb{1}_{Q_1} \otimes b_{Q_2}), \mathbb{1}_{Q_1} \otimes \mathbb{1}_{Q_2}\rangle| \le C|Q_1||Q_2|, \quad |\langle T(\mathbb{1}_{Q_1} \otimes \mathbb{1}_{Q_2}), \mathbb{1}_{Q_1} \otimes b_{Q_2}\rangle| \le C|Q_1||Q_2|.$$

7B. ***Biparameter dyadic shifts and Martikainen's representation theorem.*** Given dyadic rectangles $\boldsymbol{\mathcal{D}} = \mathcal{D}_1 \times \mathcal{D}_2$ and pairs of nonnegative integers $\vec{i} = (i_1, i_2)$ and $\vec{j} = (j_1, j_2)$, a (cancellative) biparameter dyadic shift is an operator of the form

$$\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f := \sum_{\substack{R_1 \in \mathcal{D}_1 \\ R_2 \in \mathcal{D}_2}} \sum_{\substack{P_1 \in (R_1)_{i_1} \\ P_2 \in (R_2)_{i_2}}} \sum_{\substack{Q_1 \in (R_1)_{j_1} \\ Q_2 \in (R_2)_{j_2}}} a_{P_1 Q_1 R_1 P_2 Q_2 R_2}\, \hat{f}(P_1^{\epsilon_1} \times P_2^{\epsilon_2})\, h_{Q_1}^{\delta_1} \otimes h_{Q_2}^{\delta_2}, \tag{7-1}$$

where

$$|a_{P_1 Q_1 R_1 P_2 Q_2 R_2}| \le \frac{\sqrt{|P_1||Q_1|}}{|R_1|}\frac{\sqrt{|P_2||Q_2|}}{|R_2|} = 2^{(-n_1/2)(i_1+j_1)} 2^{(-n_2/2)(i_2+j_2)}.$$

We suppress for now the signatures of the Haar functions, and assume summation over them is understood. We use the simplified notation

$$\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f := \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}}\, \hat{f}(P_1 \times P_2)\, h_{Q_1} \otimes h_{Q_2}$$

for the summation above.

First note that

$$S_{\mathcal{D}}^2(\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f) = \sum_{R_1\times R_2} \sum_{\substack{Q_1\in(R_1)_{j_1}\\ Q_2\in(R_2)_{j_2}}} \biggl(\sum_{\substack{P_1\in(R_1)_{i_1}\\ P_2\in(R_2)_{i_2}}} a_{P_1Q_1R_1P_2Q_2R_2}\, \hat{f}(P_1\times P_2)\biggr)^2 \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}$$
$$\lesssim 2^{-n_1(i_1+j_1)}2^{-n_2(i_2+j_2)}(S_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f)^2,$$

where $S_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}$ is the shifted biparameter square function in (3-1). Then, by (3-2),

$$\|\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f\|_{L^p(w)} \lesssim 2^{(-n_1/2)(i_1+j_1)}2^{(-n_2/2)(i_2+j_2)}\|S_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f\|_{L^p(w)} \lesssim \|f\|_{L^p(w)} \tag{7-2}$$

for all $w\in A_p(\mathbb{R}^{\vec{n}})$.

Next, we state Martikainen's representation theorem [2012]:

Theorem 7.1 (Martikainen). *For a biparameter singular integral operator T as defined in Section 7A, for some biparameter shifts $\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}$ it holds that*

$$\langle Tf, g\rangle = C_T\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2}\sum_{\vec{i},\vec{j}\in\mathbb{Z}_+^2} 2^{-\max(i_1,j_1)\delta/2}2^{-\max(i_2,j_2)\delta/2}\langle \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f, g\rangle,$$

where noncancellative shifts may only appear if $(i_1, j_1)=(0,0)$ or $(i_2, j_2)=(0,0)$.

In light of this theorem, in order to prove Theorem 1.1, it suffices to prove the two-weight bound for commutators $[b,\mathbb{S}_{\boldsymbol{\mathcal{D}}}]$ with the dyadic shifts, with the requirements that the bounds be *independent* of the choice of $\boldsymbol{\mathcal{D}}$ and that they depend on $\vec{i}$ and $\vec{j}$ at most *polynomially*. We first look at the case of cancellative shifts, and then treat the noncancellative case in Section 7D.

7C. *Cancellative case.*

Theorem 7.2. *Let $\boldsymbol{\mathcal{D}}=\mathcal{D}_1\times\mathcal{D}_2$ be dyadic rectangles in $\mathbb{R}^{\vec{n}}=\mathbb{R}^{n_1}\otimes\mathbb{R}^{n_2}$ and $\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}$ be a cancellative dyadic shift as defined in (7-1). If $\mu,\lambda\in A_p(\mathbb{R}^{\vec{n}})$, $1<p<\infty$, and $\nu=\mu^{1/p}\lambda^{-1/p}$, then*

$$\|[b,\mathbb{S}_{\mathcal{D}}^{\vec{i},\vec{j}}]: L^p(\mu)\to L^p(\lambda)\| \lesssim \bigl((1+\max(i_1,j_1))(1+\max(i_2,j_2))\bigr)\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)},$$

where $\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}$ denotes the norm of b in the dyadic weighted little $\mathrm{bmo}(\nu)$ *space on $\mathbb{R}^{\vec{n}}$.*

Proof. We may express the product of two functions b and f on $\mathbb{R}^{\vec{n}}$ as

$$bf = \sum \mathsf{P}_{\mathsf{b}} f + \sum \mathsf{p}_{\mathsf{b}} f + \Pi_f b,$$

where P_{b} runs through the nine paraproducts associated with $\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)$ in Section 6A, and p_{b} runs through the six paraproducts associated with $\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)$ in Section 6B. Then

$$[b,\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}]f = \sum[\mathsf{P}_{\mathsf{b}},\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}]f + \sum[\mathsf{p}_{\mathsf{b}},\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}]f + \mathcal{R}_{\vec{i},\vec{j}} f,$$

where

$$\mathcal{R}_{\vec{i},\vec{j}} f := \Pi_{\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}} f} b - \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}\Pi_f b.$$

From the two-weight inequalities for the paraproducts in Propositions 6.1 and 6.2, and the one-weight inequality for the shifts in (7-2),

$$\left\| \sum [\mathsf{P}_{\mathsf{b}}, \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}] + \sum [\mathsf{p}_{\mathsf{b}}, \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{\vec{i},\vec{j}}] : L^p(\mu) \to L^p(\lambda) \right\| \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)},$$

so we are left with bounding the remainder term $\mathcal{R}_{\vec{i},\vec{j}}$. We claim that

$$\|\mathcal{R}_{\vec{i},\vec{j}} : L^p(\mu) \to L^p(\lambda)\| \lesssim \big((1+\max(i_1, j_1))(1+\max(i_2, j_2))\big)\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)},$$

from which the result follows.

A straightforward calculation shows that

$$\mathcal{R}_{\vec{i},\vec{j}} f = \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}} \hat{f}(P_1 \times P_2)(\langle b\rangle_{Q_1\times Q_2} - \langle b\rangle_{P_1\times P_2}) h_{Q_1} \otimes h_{Q_2}.$$

We write this as a sum $\mathcal{R}_{\vec{i},\vec{j}} f = \mathcal{R}^1_{\vec{i},\vec{j}} f + \mathcal{R}^2_{\vec{i},\vec{j}} f$ by splitting the term in parentheses as

$$\langle b\rangle_{Q_1\times Q_2} - \langle b\rangle_{P_1\times P_2} = (\langle b\rangle_{Q_1\times Q_2} - \langle b\rangle_{R_1\times R_2}) + (\langle b\rangle_{R_1\times R_2} - \langle b\rangle_{P_1\times P_2}).$$

For the first term, we may apply the biparameter version of (2-2), where we keep in mind that $R_1 = Q_1^{(j_1)}$ and $R_2 = Q_2^{(j_2)}$:

$$\begin{aligned}\langle b\rangle_{Q_1\times Q_2} - \langle b\rangle_{R_1\times R_2} = &\sum_{\substack{1\le k_1\le j_1\\ 1\le k_2\le j_2}} \hat{b}(Q_1^{(k_1)} \times Q_2^{(k_2)}) h_{Q_1^{(k_1)}}(Q_1) h_{Q_2^{(k_2)}}(Q_2)\\ &+ \sum_{1\le k_1\le j_1} \left\langle b, h_{Q_1^{(k_1)}} \otimes \frac{\mathbb{1}_{R_2}}{|R_2|}\right\rangle h_{Q_1^{(k_1)}}(Q_1) + \sum_{1\le k_2\le j_2} \left\langle b, \frac{\mathbb{1}_{R_1}}{|R_1|} \otimes h_{Q_2^{(k_2)}}\right\rangle h_{Q_2^{(k_2)}}(Q_2).\end{aligned}$$

Then, we may write the operator $\mathcal{R}^1_{\vec{i},\vec{j}}$ as

$$\mathcal{R}^1_{\vec{i},\vec{j}} f = \sum_{\substack{1\le k_1\le j_1\\ 1\le k_2\le j_2}} A_{k_1,k_2} f + \sum_{1\le k_1\le j_1} B_{k_1}^{(0,1)} f + \sum_{1\le k_2\le j_2} B_{k_2}^{(1,0)} f, \tag{7-3}$$

where

$$A_{k_1,k_2} f := \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}} \hat{f}(P_1 \times P_2) \hat{b}(Q_1^{(k_1)} \times Q_2^{(k_2)}) h_{Q_1^{(k_1)}}(Q_1) h_{Q_2^{(k_2)}}(Q_2) h_{Q_1} \otimes h_{Q_2},$$

$$B_{k_1}^{(0,1)} f := \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}} \hat{f}(P_1 \times P_2) \left\langle b, h_{Q_1^{(k_1)}} \otimes \frac{\mathbb{1}_{R_2}}{|R_2|}\right\rangle h_{Q_1^{(k_1)}}(Q_1) h_{Q_1} \otimes h_{Q_2},$$

$$B_{k_2}^{(1,0)} f := \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}} \hat{f}(P_1 \times P_2) \left\langle b, \frac{\mathbb{1}_{R_1}}{|R_1|} \otimes h_{Q_2^{(k_2)}}\right\rangle h_{Q_2^{(k_2)}}(Q_2) h_{Q_1} \otimes h_{Q_2}.$$

We show that these operators satisfy

$$\begin{aligned}
&\|A_{k_1,k_2} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}(\nu)} && \text{for all } k_1, k_2,\\
&\|B^{(0,1)}_{k_1} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)} && \text{for all } k_1,\\
&\|B^{(1,0)}_{k_2} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)} && \text{for all } k_2.
\end{aligned}$$

Going back to the decomposition in (7-3), these inequalities will give

$$\|\mathcal{R}^1_{\vec{i},\vec{j}} : L^p(\mu) \to L^p(\lambda)\| \lesssim (j_1 j_2 + j_1 + j_2)\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

A symmetrical proof for the term $\mathcal{R}^2_{\vec{i},\vec{j}}$ coming from $(\langle b\rangle_{R_1\times R_2} - \langle b\rangle_{P_1\times P_2})$ will show that

$$\|\mathcal{R}^2_{\vec{i},\vec{j}} : L^p(\mu) \to L^p(\lambda)\| \lesssim (i_1 i_2 + i_1 + i_2)\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

Putting these estimates together, we obtain the desired result

$$\begin{aligned}
&\|\mathcal{R}_{\vec{i},\vec{j}} : L^p(\mu) \to L^p(\lambda)\|\\
&\qquad\lesssim (i_1 + i_2 + i_1 i_2 + j_1 + j_2 + j_1 j_2)\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)} \lesssim (1+\max(i_1, j_1))(1+\max(i_2, j_2))\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.
\end{aligned}$$

Note that we are allowed to have one of the situations $(i_1, i_2) = (0, 0)$ or $(j_1, j_2) = (0, 0)$ — but not both — and then either the term $\mathcal{R}^2_{\vec{i},\vec{j}} f$ or $\mathcal{R}^1_{\vec{i},\vec{j}} f$, respectively, will vanish.

Let us now look at the estimate for A_{k_1,k_2}. Taking again $f \in L^p(\mu)$ and $g \in L^{p'}(\lambda')$, we write $\langle A_{k_1,k_2} f, g\rangle = \langle b, \phi\rangle$, where

$$\begin{aligned}
\phi &= \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}} \hat{f}(P_1\times P_2) h_{Q_1^{(k_1)}}(Q_1) h_{Q_2^{(k_2)}}(Q_2) \hat{g}(Q_1\times Q_2) h_{Q_1^{(k_1)}} \otimes h_{Q_2^{(k_2)}}\\
&= \sum_{R_1\times R_2} \sum_{\substack{P_1\in(R_1)_{i_1}\\ P_2\in(R_2)_{i_2}}} \sum_{\substack{N_1\in(R_1)_{j_1-k_1}\\ N_2\in(R_2)_{j_2-k_2}}} \hat{f}(P_1\times P_2)\Bigg(\sum_{\substack{Q_1\in(N_1)_{k_1}\\ Q_2\in(N_2)_{k_2}}} a_{\boldsymbol{PQR}} \hat{g}(Q_1\times Q_2) h_{N_1}(Q_1) h_{N_2}(Q_2)\Bigg) h_{N_1}\otimes h_{N_2}.
\end{aligned}$$

Then

$$\begin{aligned}
S^2_{\boldsymbol{\mathcal{D}}}\phi &\lesssim \sum_{N_1\times N_2}\Bigg(\sum_{\substack{P_1\in(N_1^{(j_1-k_1)})_{i_1}\\ P_2\in(N_2^{(j_2-k_2)})_{i_2}}} |\hat{f}(P_1\times P_2)| \sum_{\substack{Q_1\in(N_1)_{k_1}\\ Q_2\in(N_2)_{k_2}}} |a_{\boldsymbol{PQR}}||\hat{g}(Q_1\times Q_2)| \frac{1}{\sqrt{|N_1|}}\frac{1}{\sqrt{|N_2|}}\Bigg)^2 \frac{\mathbb{1}_{N_1}\otimes\mathbb{1}_{N_2}}{|N_1||N_2|}\\
&\lesssim 2^{-n_1(i_1+j_1)} 2^{-n_2(i_2+j_2)} \sum_{N_1\times N_2}\Bigg(\sum_{\substack{P_1\in(N_1^{(j_1-k_1)})_{i_1}\\ P_2\in(N_2^{(j_2-k_2)})_{i_2}}} |\hat{f}(P_1\times P_2)| 2^{n_1 k_1/2} 2^{n_2 k_2/2} \langle |g|\rangle_{N_1\times N_2}\Bigg)^2 \frac{\mathbb{1}_{N_1}\otimes\mathbb{1}_{N_2}}{|N_1||N_2|}\\
&\lesssim 2^{-n_1(i_1+j_1-k_1)} 2^{-n_2(i_2+j_2-k_2)} (M_S g)^2 \sum_{R_1\times R_2}\Bigg(\sum_{\substack{P_1\in(R_1)_{i_1}\\ P_2\in(R_2)_{i_2}}} |\hat{f}(P_1\times P_2)|\Bigg)^2 \sum_{\substack{N_1\in(R_1)_{j_1-k_1}\\ N_2\in(R_2)_{j_2-k_2}}} \frac{\mathbb{1}_{N_1}\otimes\mathbb{1}_{N_2}}{|N_1||N_2|}\\
&= 2^{-n_1(i_1+j_1-k_1)} 2^{-n_2(i_2+j_2-k_2)} (M_S g)^2 (S^{(i_1,i_2),(j_1-k_1,j_2-k_2)}_{\boldsymbol{\mathcal{D}}} f)^2,
\end{aligned}$$

where the last operator is the shifted square function in (3-1). Then, from (3-2),

$$\begin{aligned}\|A_{k_1,k_2} : L^p(\mu) \to L^p(\lambda)\| &\lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(\nu)} \\ &\lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} 2^{(-n_1/2)(i_1+j_1-k_1)} 2^{(-n_2/2)(i_2+j_2-k_2)} \|M_S g\|_{L^{p'}(\lambda')} \|S_{\boldsymbol{\mathcal{D}}}^{(i_1,i_2),(j_1-k_1,j_2-k_2)} f\|_{L^p(\mu)} \\ &\lesssim \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|g\|_{L^{p'}(\lambda')} \|f\|_{L^p(\mu)}.\end{aligned}$$

Finally, we look at $B_{k_1}^{(0,1)}$, with the proof for $B_{k_2}^{(1,0)}$ being symmetrical. We write again $\langle B_{k_1}^{(0,1)} f, g\rangle = \langle b, \phi\rangle$, where

$$\phi = \sum_{\boldsymbol{R},\boldsymbol{P},\boldsymbol{Q}}^{\vec{i},\vec{j}} a_{\boldsymbol{PQR}} \hat{f}(P_1 \times P_2) h_{Q_1^{(k_1)}}(Q_1) \hat{g}(Q_1 \times Q_2) h_{Q_1^{(k_1)}} \otimes \frac{\mathbb{1}_{R_2}}{|R_2|}.$$

Then

$$S_{\mathcal{D}_1}^2 f \lesssim 2^{-n_1(i_1+j_1)} 2^{-n_2(i_2+j_2)} \sum_{\substack{R_1\in\mathcal{D}_1\\ N_1\in(R_1)_{j_1-k_1}}} \frac{\mathbb{1}_{N_1}}{|N_1|} \Bigg(\sum_{R_2\in\mathcal{D}_2} \sum_{\substack{P_1\in(R_1)_{i_1}\\ P_2\in(R_2)_{i_2}}} |\hat{f}(P_1\times P_2)| \sum_{Q_2\in(R_2)_{j_2}} \langle |H_{Q_2} g|\rangle_{N_1} 2^{n_1k_1/2} \frac{\mathbb{1}_{R_2}}{|R_2|} \Bigg)^2,$$

and the summation above is bounded by

$$\Bigg(\sum_{\substack{R_1\in\mathcal{D}_1\\ N_1\in(R_1)_{j_1-k_1}}} \frac{\mathbb{1}_{N_1}}{|N_1|} \sum_{R_2\in\mathcal{D}_2} \Bigg(\sum_{\substack{P_1\in(R_1)_{i_1}\\ P_2\in(R_2)_{i_2}}} |\hat{f}(P_1 \times P_2)| \Bigg)^2 \frac{\mathbb{1}_{R_2}}{|R_2|} \Bigg) \Bigg(\sum_{R_2\in\mathcal{D}_2} \Bigg(\sum_{Q_2\in(R_2)_{j_2}} M_{\mathcal{D}_1}(H_{Q_2} g) \Bigg)^2 \frac{\mathbb{1}_{R_2}}{|R_2|} \Bigg),$$

which is exactly

$$(S_{\boldsymbol{\mathcal{D}}}^{(i_1,i_2),(j_1-k_1,0)} f)^2 ([MS]^{j_2,0} g)^2.$$

From (3-2) and (3-3), we obtain exactly $\|S_{\mathcal{D}_1}\phi\|_{L^1(\nu)} \lesssim \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')}$, and the proof is complete. □

7D. ***The noncancellative case.*** Following the proof in [Martikainen 2012], we are left with three types of terms to consider, all of paraproduct type,

- the full standard paraproduct, Π_a and Π_a^*,
- the full mixed paraproducts, $\Pi_{a;(0,1)}$ and $\Pi_{a;(1,0)}$,

where, in each case, a is some fixed function in *unweighted* product $\mathrm{BMO}(\mathbb{R}^{\vec{n}})$, with $\|a\|_{\mathrm{BMO}(\mathbb{R}^{\vec{n}})} \leq 1$, and

- the *partial* paraproducts, defined for every $i_1, j_1 \geq 0$ as

$$\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{i_1,j_1} f := \sum_{\substack{R_1\in\mathcal{D}_1\\ R_2\in\mathcal{D}_2}} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_1}}} \hat{a}_{P_1Q_1R_1}(R_2^{\delta_2}) \hat{f}(P_1^{\epsilon_1} \times R_2^{\epsilon_2}) h_{Q_1}^{\delta_1} \times \frac{\mathbb{1}_{R_2}}{|R_2|},$$

  where, for every fixed P_1, Q_1, R_1, we have $a_{P_1Q_1R_1}(x_2)$ is a $\mathrm{BMO}(\mathbb{R}^{n_2})$ function with

$$\|a_{P_1Q_1R_1}\|_{\mathrm{BMO}(\mathbb{R}^{n_2})} \leq \frac{\sqrt{|P_1|}\sqrt{|Q_1|}}{|R_1|} = 2^{(-n_1/2)(i_1+j_1)},$$

and

$$\hat{a}_{P_1Q_1R_1}(R_2^{\delta_2}) := \langle a_{P_1Q_1R_1}, h_{R_2}^{\delta_2}\rangle_{\mathbb{R}^{n_2}} := \int_{\mathbb{R}^{n_2}} a_{P_1Q_1R_1}(x_2) h_{R_2}^{\delta_2}(x_2)\,dx_2.$$

The symmetrical partial paraproduct $\mathbb{S}_{\mathcal{D}}^{i_2,j_2}$ is defined analogously.

We treat each case separately.

7D1. *The full standard paraproduct.* In this case, we are looking at the commutator $[b, \Pi_a]$, where

$$\Pi_a f := \sum_{R\in\mathcal{D}} \hat{a}(R)\langle f\rangle_R h_R,$$

and $a \in \mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{\vec{n}})$ with $\|a\|_{\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{\vec{n}})} \le 1$. We prove that:

Theorem 7.3. *Let $\mu, \lambda \in A_p(\mathbb{R}^{\vec{n}})$, $1 < p < \infty$ and $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|[b, \Pi_a] : L^p(\mu) \to L^p(\lambda)\| \lesssim \|a\|_{\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{\vec{n}})} \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}.$$

Proof. We remark first that

$$\Pi_a(bf) = \sum_{R\in\mathcal{D}} \hat{a}(R)\langle bf\rangle_R h_R \quad\text{and}\quad \Pi_{\Pi_a f} b = \sum_{R\in\mathcal{D}} \hat{a}(R)\langle b\rangle_R\langle f\rangle_R h_R,$$

so

$$\begin{aligned}\Pi_a(bf) - \Pi_{\Pi_a f} b &= \sum_{R\in\mathcal{D}} \hat{a}(R)\big(\langle bf\rangle_R - \langle b\rangle_R\langle f\rangle_R\big)h_R \\ &= \Pi_a\Big(\sum \mathsf{P}_\mathsf{b} f + \sum \mathsf{p}_\mathsf{b} f + \Pi_f b\Big) - \Pi_{\Pi_a f} b,\end{aligned}$$

where the last equality was obtained by simply expanding bf into paraproducts. Then

$$\Pi_{\Pi_a f} b - \Pi_a \Pi_f b = \sum \Pi_a \mathsf{P}_\mathsf{b} f + \sum \Pi_a \mathsf{p}_\mathsf{b} f - \sum_{R\in\mathcal{D}} \hat{a}(R)\big(\langle bf\rangle_R - \langle b\rangle_R\langle f\rangle_R\big)h_R.$$

Noting that

$$[b, \Pi_a] f = \sum \mathsf{P}_\mathsf{b}\Pi_a f + \sum \mathsf{p}_\mathsf{b}\Pi_a f - \sum \Pi_a \mathsf{P}_\mathsf{b} f - \sum \Pi_a \mathsf{p}_\mathsf{b} f + \Pi_{\Pi_a f} b - \Pi_a \Pi_f b,$$

we obtain

$$[b, \Pi_a] f = \sum \mathsf{P}_\mathsf{b}\Pi_a f + \sum \mathsf{p}_\mathsf{b}\Pi_a f - \sum_{R\in\mathcal{D}} \hat{a}(R)\big(\langle bf\rangle_R - \langle b\rangle_R\langle f\rangle_R\big)h_R.$$

The first terms are easily handled:

$$\|\mathsf{P}_\mathsf{b}\Pi_a f\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}(\nu)}\|\Pi_a f\|_{L^p(\mu)} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}(\nu)}\|a\|_{\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{\vec{n}})}\|f\|_{L^p(\mu)},$$

$$\|\mathsf{p}_\mathsf{b}\Pi_a f\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}\|\Pi_a f\|_{L^p(\mu)} \lesssim \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}\|a\|_{\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{\vec{n}})}\|f\|_{L^p(\mu)}.$$

So we are left with the third term.

Now, for any dyadic rectangle R,

$$\langle bf\rangle_R - \langle b\rangle_R\langle f\rangle_R = \frac{1}{|R|}\int_R f(x)\mathbb{1}_R(x)(b(x) - \langle b\rangle_R)\,dx.$$

Expressing $\mathbb{1}_R(b-\langle b\rangle_R)$ as in (2-5), we obtain

$$\langle bf\rangle_R-\langle b\rangle_R\langle f\rangle_R=\frac{1}{|R|}\sum_{\substack{P_1\subset Q_1\\ P_2\subset Q_2}}\hat{b}(P_1\times P_2)\hat{f}(P_1\times P_2)\\ +\frac{1}{|R|}\sum_{P_1\subset Q_1}\Big\langle b,h_{P_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\Big\rangle\langle f,h_{P_1}\otimes\mathbb{1}_{Q_2}\rangle+\frac{1}{|R|}\sum_{P_2\subset Q_2}\Big\langle b,\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}\Big\rangle\langle f,\mathbb{1}_{Q_1}\otimes h_{P_2}\rangle.$$

Therefore

$$\sum_{R\in\mathcal{D}}\hat{a}(R)\big(\langle bf\rangle_R-\langle b\rangle_R\langle f\rangle_R\big)h_R=\Lambda_{a,b}f+\lambda_{a,b}^{(0,1)}f+\lambda_{a,b}^{(1,0)}f,$$

where

$$\Lambda_{a,b}f:=\sum_{Q_1\times Q_2}\hat{a}(Q_1\times Q_2)\frac{1}{|Q_1||Q_2|}\bigg(\sum_{\substack{P_1\subset Q_1\\ P_2\subset Q_2}}\hat{b}(P_1\times P_2)\hat{f}(P_1\times P_2)\bigg)h_{Q_1}\otimes h_{Q_2},$$

$$\lambda_{a,b}^{(0,1)}f:=\sum_{Q_1\times Q_2}\hat{a}(Q_1\times Q_2)\frac{1}{|Q_1||Q_2|}\bigg(\sum_{P_1\subset Q_1}\Big\langle b,h_{P_1}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|}\Big\rangle\langle f,h_{P_1}\otimes\mathbb{1}_{Q_2}\rangle\bigg)h_{Q_1}\otimes h_{Q_2},$$

$$\lambda_{a,b}^{(1,0)}f:=\sum_{Q_1\times Q_2}\hat{a}(Q_1\times Q_2)\frac{1}{|Q_1||Q_2|}\bigg(\sum_{P_2\subset Q_2}\Big\langle b,\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}\Big\rangle\langle f,\mathbb{1}_{Q_1}\otimes h_{P_2}\rangle\bigg)h_{Q_1}\otimes h_{Q_2}.$$

To analyze the term $\Lambda_{a,b}$, we write $\langle\Lambda_{a,b}f,g\rangle=\langle b,\phi\rangle$, where

$$\begin{aligned}\phi&=\sum_{P_1\times P_2}\hat{f}(P_1\times P_2)\bigg(\sum_{\substack{Q_1\supset P_1\\ Q_2\supset P_2}}\hat{a}(Q_1\times Q_2)\hat{g}(Q_1\times Q_2)\frac{1}{|Q_1||Q_2|}\bigg)h_{P_1}\otimes h_{P_2}\\ &=\sum_{R\in\mathcal{D}}\hat{f}(R)\bigg(\sum_{T\in\mathcal{D},T\supset R}\hat{a}(T)\hat{g}(T)\frac{1}{|T|}\bigg)h_R.\end{aligned}$$

So $|\langle\Lambda_{a,b}f,g\rangle|\lesssim\|b\|_{\mathrm{BMO}_{\mathcal{D}}(\nu)}\|S_{\mathcal{D}}\phi\|_{L^1(\nu)}$, and

$$S_{\mathcal{D}}^2\phi=\sum_{R\in\mathcal{D}}|\hat{f}(R)|^2\bigg(\sum_{T\in\mathcal{D},T\supset R}\hat{a}(T)\hat{g}(T)\frac{1}{|T|}\bigg)^2\frac{\mathbb{1}_R}{|R|}\leq\sum_{R\in\mathcal{D}}|\hat{f}(R)|^2\bigg(\sum_{T\in\mathcal{D},T\supset R}\hat{a}_\tau(T)\hat{g}_\tau(T)\frac{1}{|T|}\bigg)^2\frac{\mathbb{1}_R}{|R|},$$

where $a_\tau:=\sum_{R\in\mathcal{D}}|\hat{a}(R)|h_R$ and $g_\tau:=\sum_{R\in\mathcal{D}}|\hat{g}(R)|h_R$ are martingale transforms which do not increase either the BMO norm of a, or the $L^{p'}(\lambda')$ norm of g. Now note that

$$\langle\Pi_{a_\tau}^*g_\tau\rangle_R=\sum_{T\subsetneq R}\hat{a}_\tau(T)\hat{g}_\tau(T)\frac{1}{|R|}+\sum_{T\supset R}\hat{a}_\tau(T)\hat{g}_\tau(T)\frac{1}{|T|},$$

and since all the Haar coefficients of a_τ and g_τ are nonnegative, we may write

$$\sum_{T\supset R}\hat{a}_\tau(T)\hat{g}_\tau(T)\frac{1}{|T|}\leq\langle\Pi_{a_\tau}^*g_\tau\rangle_R.$$

Then

$$S^2_{\boldsymbol{\mathcal{D}}}\phi \le \sum_{R\in\boldsymbol{\mathcal{D}}} |\hat{f}(R)|^2 \langle \Pi^*_{a_\tau} g_\tau\rangle_R^2 \frac{\mathbb{1}_R}{|R|} \le (M_S \Pi^*_{a_\tau} g_\tau)^2 S^2_{\boldsymbol{\mathcal{D}}} f,$$

and

$$\begin{aligned}\|S_{\boldsymbol{\mathcal{D}}}\phi\|_{L^1(\nu)} &\le \|M_S \Pi^*_{a_\tau} g_\tau\|_{L^{p'}(\lambda')} \|S_{\boldsymbol{\mathcal{D}}} f\|_{L^p(\mu)}\\ &\lesssim \|\Pi^*_{a_\tau} g_\tau\|_{L^{p'}(\lambda')} \|f\|_{L^p(\mu)}\\ &\lesssim \|a_\tau\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})} \|g_\tau\|_{L^{p'}(\lambda')} \|f\|_{L^p(\mu)},\end{aligned}$$

which gives us the desired estimate

$$\|\Lambda_{a,b} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|a\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})} \|b\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

Finally, we analyze the term $\lambda^{(0,1)}_{a,b}$, with the last term being symmetrical. We have $\langle \lambda^{(0,1)}_{a,b} f, g\rangle = \langle b, \phi\rangle$ with

$$\phi = \sum_{P_1} \Bigg(\sum_{P_2} \langle f, h_{P_1} \otimes \mathbb{1}_{P_2}\rangle \frac{1}{|P_2|} \sum_{Q_1\supset P_1} \hat{a}(Q_1\times P_2)\hat{g}(Q_1\times P_2) \frac{1}{|Q_1|}\frac{\mathbb{1}_{P_2}}{|P_2|}\Bigg) h_{P_1},$$

and $|\langle \lambda^{(0,1)}_{a,b} f, g\rangle| \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)} \|S_{\mathcal{D}_1}\phi\|_{L^1(\nu)}$. Now

$$S^2_{\mathcal{D}_1}\phi \le \sum_{P_1} \Bigg(\sum_{P_2} \langle |H_{P_1} f|\rangle_{P_2} \Bigg(\sum_{Q_1\supset P_1} \hat{a}_\tau(Q_1\times P_2)\hat{g}_\tau(Q_2\times P_2)\frac{1}{|Q_1|}\Bigg) \frac{\mathbb{1}_{P_2}}{|P_2|}\Bigg)^2 \frac{\mathbb{1}_{P_1}}{|P_1|},$$

where we are using the same martingale transforms as above. Note that

$$\left\langle \Pi^*_{a_\tau} g_\tau, \frac{\mathbb{1}_{P_1}}{|P_1|}\right\rangle_{\mathbb{R}^{n_1}} (x_2) = \sum_{P_2} \frac{\mathbb{1}_{P_2}(x_2)}{|P_2|} \sum_{Q_1} \hat{a}_\tau(Q_1\times P_2)\hat{g}_\tau(Q_1\times P_2) \frac{|Q_1\cap P_1|}{|Q_1||P_1|},$$

and again since all terms are nonnegative:

$$\begin{aligned} S^2_{\mathcal{D}_1}\phi &\le \sum_{P_1} M^2_{\mathcal{D}_2}(H_{P_1} f)(x_2) \Bigg(\sum_{Q_1\supset P_1}\sum_{P_2} \hat{a}_\tau(Q_1\times P_2)\hat{g}_\tau(Q_1\times P_2)\frac{1}{|Q_1|}\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\Bigg)^2 \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|}\\ &\le \sum_{P_1} M^2_{\mathcal{D}_2}(H_{P_1} f)(x_2) \Bigg(\left\langle \Pi^*_{a_\tau} g_\tau, \frac{\mathbb{1}_{P_1}}{|P_1|}\right\rangle_{\mathbb{R}^{n_1}} (x_2)\Bigg)^2 \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|}\\ &\le (M_{\mathcal{D}_1}(\Pi^*_{a_\tau} g_\tau)(x_1,x_2))^2 \sum_{P_1} M^2_{\mathcal{D}_2}(H_{P_1} f)(x_2)\frac{\mathbb{1}_{P_1}(x_1)}{|P_1|}\\ &= (M_{\mathcal{D}_1}(\Pi^*_{a_\tau} g_\tau)(x_1,x_2))^2 ([SM]f(x_1,x_2))^2.\end{aligned}$$

Then

$$\|S_{\mathcal{D}_1}\phi\|_{L^1(\nu)} \lesssim \|\Pi^*_{a_\tau} g_\tau\|_{L^{p'}(\lambda')} \|[SM]f\|_{L^p(\mu)} \lesssim \|a\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})} \|g\|_{L^{p'}(\lambda')} \|f\|_{L^p(\mu)},$$

and so

$$\|\lambda^{(0,1)}_{a,b} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|a\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\nu)} \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}. \qquad \square$$

7D2. *The full mixed paraproduct.* We are now dealing with $[b, \Pi_{a;(0,1)}]$, where

$$\Pi_{a;(0,1)} f := \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \left\langle f, h_{P_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \frac{\mathbb{1}_{P_1}}{|P_1|} \otimes h_{P_2}.$$

Theorem 7.4. *Let $\mu, \lambda \in A_p(\mathbb{R}^{\vec{n}})$, $1 < p < \infty$ and $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|[b, \Pi_{a;(0,1)}] : L^p(\mu) \to L^p(\lambda)\| \lesssim \|a\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})} \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

Note that the case $[b, \Pi_{a;(1,0)}]$ follows symmetrically.

Proof. By the standard considerations, we only need to bound the remainder term

$$\mathcal{R}^{(0,1)}_{a,b} f := \Pi_{\Pi_{a;(0,1)} f} b - \Pi_{a;(0,1)} \Pi_f b.$$

Explicitly, these terms are

$$\Pi_{\Pi_{a;(0,1)} f} b = \sum_{P_1 \times P_2} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h^{\epsilon_1}_{P_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \left(\sum_{Q_1 \supsetneq P_1} \langle b \rangle_{Q_1 \times P_2} h^{\delta_1}_{Q_1}(P_1) h^{\delta_1}_{Q_1}(x_1) \right) h^{\epsilon_2}_{P_2}(x_2),$$

$$\Pi_{a;(0,1)} \Pi_f b = \sum_{P_1 \times P_2} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left(\sum_{Q_2 \supsetneq P_2} \hat{f}(P_1^{\epsilon_1} \times Q_2^{\delta_2}) \langle b \rangle_{P_1 \times Q_2} h^{\delta_2}_{Q_2}(P_2) \right) \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} \otimes h^{\epsilon_2}_{P_2}(x_2).$$

Consider now a third term

$$T := \sum_{P_1 \times P_2} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \langle b \rangle_{P_1 \times P_2} \left\langle f, h^{\epsilon_1}_{P_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \frac{\mathbb{1}_{P_1}}{|P_1|} \otimes h^{\epsilon_2}_{P_2}.$$

Using the one-parameter formula

$$\frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} = \sum_{Q_1 \supsetneq P_1} h^{\delta_1}_{Q_1}(P_1) h^{\delta_1}_{Q_1}(x_1),$$

we write T as

$$T = \sum_{P_1 \times P_2} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h^{\epsilon_1}_{P_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \left(\sum_{Q_1 \supsetneq P_1} \langle b \rangle_{P_1 \times P_2} h^{\delta_1}_{Q_1}(P_1) h^{\delta_1}_{Q_1}(x_1) \right) h^{\epsilon_2}_{P_2}(x_2),$$

allowing us to combine this term with $\Pi_{\Pi_{a;(0,1)} f} b$:

$$\Pi_{\Pi_{a;(0,1)} f} b - T = \sum_{P_1 \times P_2} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h^{\epsilon_1}_{P_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \left(\sum_{Q_1 \supsetneq P_1} (\langle b \rangle_{Q_1 \times P_2} - \langle b \rangle_{P_1 \times P_2}) h^{\delta_1}_{Q_1}(P_1) h^{\delta_1}_{Q_1}(x_1) \right) h^{\epsilon_2}_{P_2}(x_2).$$

Using (2-2), we have

$$\langle b \rangle_{Q_1 \times P_2} - \langle b \rangle_{P_1 \times P_2} = - \sum_{R_1 : P_1 \subsetneq R_1 \subset Q_1} \left\langle b, h^{\tau_1}_{R_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h^{\tau_1}_{R_1}(P_1),$$

and then the term in parentheses above becomes

$$- \sum_{Q_1 \supsetneq P_1} \left(\sum_{R_1 : P_1 \subsetneq R_1 \subset Q_1} \left\langle b, h^{\tau_1}_{R_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h^{\tau_1}_{R_1}(P_1) \right) h^{\delta_1}_{Q_1}(P_1) h^{\delta_1}_{Q_1}(x_1). \tag{7-4}$$

Next, we analyze this term depending on the relationship between R_1 and Q_1:

Case 1: $R_1 \subsetneq Q_1$. Then we may rewrite the sum as

$$\sum_{R_1 \supsetneq P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \sum_{Q_1 \supsetneq R_1} \underbrace{h_{Q_1}^{\delta_1}(P_1)}_{=h_{Q_1}^{\delta_1}(R_1)} h_{Q_1}^{\delta_1}(x_1) = \sum_{R_1 \supsetneq P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}.$$

This then leads to

$$\begin{aligned}
&\sum_{P_1 \times P_2} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h_{P_1}^{\epsilon_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \left(\sum_{R_1 \supsetneq P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|} \right) h_{P_2}^{\epsilon_2}(x_2) \\
&\qquad = \sum_{R_1 \times P_2} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \left(\sum_{P_1 \subsetneq R_1} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h_{P_1}^{\epsilon_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \right) \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|} \otimes h_{P_2}^{\epsilon_2}(x_2) \\
&\qquad = \sum_{R_1 \times P_2} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \langle \Pi_{a;(0,1)} f, h_{R_1}^{\tau_1} \otimes h_{P_2}^{\epsilon_2} \rangle \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|} \otimes h_{P_2}^{\epsilon_2}(x_2) \\
&\qquad = \pi^*_{b;(0,1)} \Pi_{a;(0,1)} f.
\end{aligned}$$

Case 2a: $R_1 = Q_1$ and $\tau_1 \neq \delta_1$. Then (7-4) becomes

$$-\sum_{Q_1 \supsetneq P_1} \left\langle b, h_{Q_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \frac{1}{\sqrt{|Q_1|}} h_{Q_1}^{\tau_1 + \delta_1}(P_1) h_{Q_1}^{\delta_1}(x_1),$$

which leads to

$$\begin{aligned}
&\sum_{Q_1 \times P_2} \left\langle b, h_{Q_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \frac{1}{\sqrt{|Q_1|}} h_{Q_1}^{\delta_1}(x_1) h_{P_2}^{\epsilon_2}(x_2) \sum_{P_1 \subsetneq Q_1} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h_{P_1}^{\epsilon_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h_{Q_1}^{\tau_1+\delta_1}(P_1) \\
&\qquad = \sum_{Q_1 \times P_2} \left\langle b, h_{Q_1}^{\tau_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \langle \Pi_{a;(0,1)} f, h_{Q_1}^{\tau_1+\delta_1} \otimes h_{P_2}^{\epsilon_2} \rangle \frac{1}{\sqrt{|Q_1|}} h_{Q_1}^{\delta_1}(x_1) \otimes h_{P_2}^{\epsilon_2}(x_2) \\
&\qquad = \gamma_{b;(0,1)} \Pi_{a;(0,1)} f.
\end{aligned}$$

Case 2b: $R_1 = Q_1$ and $\tau_1 = \delta_1$. Then (7-4) becomes

$$\sum_{Q_1 \supsetneq P_1} \left\langle b, h_{Q_1}^{\delta_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle \frac{1}{|Q_1|} h_{Q_1}^{\delta_1},$$

which gives rise to the term

$$T_{a,b}^{(0,1)} f := \sum_{Q_1 \times P_2} \left\langle b, h_{Q_1}^{\delta_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle h_{Q_1}^{\delta_1}(x_1) h_{P_2}^{\epsilon_2}(x_2) \frac{1}{|Q_1|} \sum_{P_1 \subsetneq Q_1} \hat{a}(P_1^{\epsilon_1} \times P_2^{\epsilon_2}) \left\langle f, h_{P_1}^{\epsilon_1} \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right\rangle.$$

We have proved that

$$\Pi_{\Pi_{a;(0,1)} f} b - T = -\pi^*_{b;(0,1)} \Pi_{a;(0,1)} f - \gamma_{b;(0,1)} \Pi_{a;(0,1)} f - T_{a,b}^{(0,1)} f.$$

Expressing T instead as

$$T=\sum_{P_1\times P_2}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})\bigg(\sum_{Q_2\supsetneq P_2}\hat{f}(P_1^{\epsilon_1}\times Q_2^{\delta_2})\langle b\rangle_{P_1\times P_2}h_{Q_2}^{\delta_2}(P_2)\bigg)\frac{\mathbb{1}_{P_1}}{|P_1|}\otimes h_{P_2}^{\epsilon_2},$$

we are able to pair it with $\Pi_{a;(0,1)}\Pi_f b$. Then, a similar analysis yields

$$T-\Pi_{a;(0,1)}\Pi_f b=\Pi_{a;(0,1)}\pi_{b;(1,0)}f+\Pi_{a;(0,1)}\gamma_{b;(1,0)}f+T_{a,b}^{(1,0)}f,$$

where

$$T_{a,b}^{(1,0)}f:=\sum_{P_1\times P_2}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})\frac{\mathbb{1}_{P_1}(x_1)}{|P_1|}\otimes h_{P_2}^{\epsilon_2}(x_2)\bigg(\sum_{Q_2\supsetneq P_2}\Big\langle b,\frac{\mathbb{1}_{P_1}}{|P_1|}\otimes h_{Q_2}^{\delta_2}\Big\rangle\hat{f}(P_1^{\epsilon_1}\times Q_2^{\delta_2})\frac{1}{|Q_2|}\bigg).$$

Then

$$\mathcal{R}_{a,b}^{(0,1)}f=\Pi_{a;(0,1)}\pi_{b;(1,0)}f+\Pi_{a;(0,1)}\gamma_{b;(1,0)}f-\pi_{b;(0,1)}^{*}\Pi_{a;(0,1)}f-\gamma_{b;(0,1)}\Pi_{a;(0,1)}f+T_{a,b}^{(1,0)}f-T_{a,b}^{(0,1)}f.$$

It is now obvious that the first four terms are bounded as desired, and it remains to bound the terms $T_{a,b}$.

We look at $T_{a,b}^{(0,1)}$, for which we can write $\langle T_{a,b}^{(0,1)}f,g\rangle=\langle b,\phi\rangle$, where

$$\phi=\sum_{Q_1\times P_2}\hat{g}(Q_1^{\delta_1}\times P_2^{\epsilon_2})\frac{1}{|Q_1|}\bigg(\sum_{P_1\subsetneq Q_1}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})\Big\langle f,h_{P_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{P_2}}{|P_2|}\Big\rangle\bigg)h_{Q_1}^{\delta_1}\otimes\frac{\mathbb{1}_{P_2}}{|P_2|}.$$

Then $|\langle T_{a,b}^{(0,1)}f,g\rangle|\lesssim\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}\|S_{\mathcal{D}_1}\phi\|_{L^1(\nu)}$, and

$$S_{\mathcal{D}_1}^2\phi=\sum_{Q_1}\bigg(\sum_{P_2}\hat{g}(Q_1^{\delta_1}\times P_2^{\epsilon_2})\bigg(\frac{1}{|Q_1|}\sum_{P_1\subsetneq Q_1}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})\Big\langle f,h_{P_1}^{\epsilon_2}\otimes\frac{\mathbb{1}_{P_2}}{|P_2|}\Big\rangle\bigg)\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\bigg)^2\frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}.$$

Now,

$$\Big\langle\Pi_{a;(0,1)}f,\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}^{\epsilon_2}\Big\rangle=\sum_{P_1}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})\Big\langle f,h_{P_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{P_2}}{|P_2|}\Big\rangle\frac{|P_1\cap Q_1|}{|P_1||Q_1|}.$$

Define the martingale transform $a\mapsto a_\tau=\sum_{P_1\times P_2}\tau_{P_1,P_2}^{\epsilon_1,\epsilon_2}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})$, where

$$\tau_{P_1,P_2}^{\epsilon_1,\epsilon_2}=\begin{cases}l+1 & \text{if } \langle f,h_{P_1}^{\epsilon_1}\otimes\mathbb{1}_{P_2}/|P_2|\rangle\geq 0,\\ -1 & \text{otherwise.}\end{cases}$$

Note that, while this transform does depend on f, in the end it will not matter, as this will be absorbed into the product BMO norm of a_τ. Then we have

$$\frac{1}{|Q_1|}\bigg|\sum_{P_1\subsetneq Q_1}\hat{a}(P_1^{\epsilon_1}\times P_2^{\epsilon_2})\Big\langle f,h_{P_1}^{\epsilon_1}\otimes\frac{\mathbb{1}_{P_2}}{|P_2|}\Big\rangle\bigg|\leq\Big\langle\Pi_{a_\tau;(0,1)}f,\frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}^{\epsilon_2}\Big\rangle.$$

Returning to the square function estimate, we now have

$$\begin{aligned}S_{\mathcal{D}_1}^2\phi&\leq\sum_{Q_1}\bigg(\sum_{P_2}|\hat{g}(Q_1^{\delta_1}\times P_2^{\epsilon_2})|^2\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\bigg)\bigg(\sum_{P_2}\langle|H_{P_2}^{\epsilon_2}\Pi_{a_\tau;(0,1)}f|\rangle_{Q_1}^2\mathbb{1}_{Q_1}(x_1)\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\bigg)\frac{\mathbb{1}_{Q_1}(x_1)}{|Q_1|}\\&\leq S_{\boldsymbol{\mathcal{D}}}^2 g\bigg(\sum_{P_2}M_{\mathcal{D}_1}^2(H_{P_2}^{\epsilon_2}\Pi_{a_\tau;(0,1)}f)(x_1)\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\bigg)=S_{\boldsymbol{\mathcal{D}}}^2 g([MS]\Pi_{a_\tau;(0,1)}f)^2.\end{aligned}$$

Finally,

$$\begin{aligned}\|S_{\mathcal{D}_1}\phi\|_{L^1(\nu)} &\le \|S_{\boldsymbol{\mathcal{D}}}g\|_{L^{p'}(\lambda')}\|[MS]\Pi_{a_\tau;(0,1)}f\|_{L^p(\mu)} \\ &\lesssim \|g\|_{L^{p'}(\lambda')}\underbrace{\|\Pi_{a_\tau;(0,1)}f\|_{L^p(\mu)}}_{\lesssim\|a_\tau\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})}\|f\|_{L^p(\mu)}} \lesssim \|a\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})}\|f\|_{L^p(\mu)}\|g\|_{L^{p'}(\lambda')},\end{aligned}$$

showing that

$$\|T_{a,b}^{(0,1)} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|a\|_{\mathrm{BMO}_{\boldsymbol{\mathcal{D}}}(\mathbb{R}^{\vec{n}})}\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

The estimate for $T_{a,b}^{(1,0)}$ follows similarly. □

7D3. *The partial paraproducts.* We work with

$$\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{i_1,j_1} f := \sum_{R_1\times R_2}\sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_1}}} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2})\hat{f}(P_1^{\epsilon_1}\times R_2^{\epsilon_2})h_{Q_1}^{\delta_1}\otimes\frac{\mathbb{1}_{R_2}}{|R_2|},$$

where i_1, j_1 are nonnegative integers, and for every P_1, Q_1, R_1,

$$a_{P_1Q_1R_1}(x_2) \in \mathrm{BMO}(\mathbb{R}^{n_2}) \quad \text{with } \|a_{P_1Q_1R_1}\|_{\mathrm{BMO}(\mathbb{R}^{n_2})} \le 2^{(-n_1/2)(i_1+j_1)}.$$

Theorem 7.5. *Let $\mu, \lambda \in A_p(\mathbb{R}^{\vec{n}})$, $1 < p < \infty$ and $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|[b, \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{i_1,j_1}] : L^p(\mu) \to L^p(\lambda)\| \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

First we need the one-weight bound for the partial paraproducts:

Proposition 7.6. *For any $w \in A_p(\mathbb{R}^{\vec{n}})$, $1 < p < \infty$,*

$$\|\mathbb{S}_{\boldsymbol{\mathcal{D}}}^{i_1,j_1} : L^p(w) \to L^p(w)\| \lesssim 1. \tag{7-5}$$

Proof. Let $f \in L^p(w)$ and $g \in L^{p'}(w')$, and we will show that $|\langle \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{i_1,j_1} f, g\rangle| \lesssim \|f\|_{L^p(w)}\|g\|_{L^{p'}(w')}$. First,

$$\begin{aligned}\|\langle \mathbb{S}_{\boldsymbol{\mathcal{D}}}^{i_1,j_1} f, g\rangle\| &\le \sum_{R_1}\sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} |\langle a_{P_1Q_1R_1}, \phi_{P_1Q_1R_1}\rangle_{\mathbb{R}^{n_2}}| \\ &\le \sum_{R_1}\sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \|a_{P_1Q_1R_1}\|_{\mathrm{BMO}(\mathbb{R}^{n_2})}\|S_{\mathcal{D}_2}\phi_{P_1Q_1R_1}\|_{L^1(\mathbb{R}^{n_2})} \\ &\le 2^{(-n_1/2)(i_1+j_1)}\sum_{R_1}\sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \|S_{\mathcal{D}_2}\phi_{P_1Q_1R_1}\|_{L^1(\mathbb{R}^{n_2})},\end{aligned}$$

where for every P_1, Q_1, R_1,

$$\phi_{P_1Q_1R_1}(x_2) := \sum_{R_2}\hat{f}(P_1\times R_2)\Big\langle g, h_{Q_1}\otimes\frac{\mathbb{1}_{R_2}}{|R_2|}\Big\rangle h_{R_2}(x_2).$$

Now,

$$S^2_{\mathcal{D}_2}\phi_{P_1Q_1R_1} = \sum_{R_2} |\widehat{H_{P_1} f}(R_2)|^2 \langle |H_{Q_1} g|\rangle^2_{R_2} \frac{\mathbb{1}_{R_2}(x_2)}{|R_2|} \leq (M_{\mathcal{D}_2} H_{Q_1} g)^2(x_2)(S_{\mathcal{D}_2} H_{P_1} f)^2(x_2),$$

so

$$\begin{aligned}
&\sum_{R_1} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \|S_{\mathcal{D}_2}\phi_{P_1Q_1R_1}\|_{L^1(\mathbb{R}^{n_2})} \\
&\quad\leq \sum_{R_1} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \int_{\mathbb{R}^{n_2}} (M_{\mathcal{D}_2} H_{Q_1} g)(x_2)(S_{\mathcal{D}_2} H_{P_1} f)(x_2)\, dx_2 \\
&\quad= \int_{\mathbb{R}^{n_2}} \int_{\mathbb{R}^{n_1}} \sum_{R_1} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} (M_{\mathcal{D}_2} H_{Q_1} g)(x_2)(S_{\mathcal{D}_2} H_{P_1} f)(x_2) \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}\, dx_1\, dx_2 \\
&\quad\leq \int_{\mathbb{R}^{\vec{n}}} \biggl(\sum_{R_1}\biggl(\sum_{P_1\in(R_1)_{i_1}} S_{\mathcal{D}_2} H_{P_1} f(x_2)\biggr)^2 \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}\biggr)^{1/2} \biggl(\sum_{R_1}\biggl(\sum_{Q_1\in(R_1)_{j_1}} M_{\mathcal{D}_2} H_{Q_1} g(x_2)\biggr)^2 \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}\biggr)^{1/2} dx \\
&\quad= \int_{\mathbb{R}^{\vec{n}}} [SS_{\mathcal{D}_2}]^{i_1,0} f \cdot [SM_{\mathcal{D}_2}]^{j_1,0} g w^{1/p} w^{-1/p}\, dx.
\end{aligned}$$

Then, from the estimates in (3-3),

$$\begin{aligned}
\|\langle \mathbb{S}^{i_1,j_1}_{\boldsymbol{\mathcal{D}}} f, g\rangle\| &\leq 2^{(-n_1/2)(i_1+j_1)} \|[SS_{\mathcal{D}_2}]^{i_1,0} f\|_{L^p(w)} \|[SM_{\mathcal{D}_2}]^{j_1,0} g\|_{L^{p'}(w')} \\
&\lesssim 2^{(-n_1/2)(i_1+j_1)} 2^{(n_1 i_1/2)} \|f\|_{L^p(w)} 2^{(n_1 j_1/2)} \|g\|_{L^{p'}(w')},
\end{aligned}$$

and the result follows. □

Proof of Theorem 7.5. In light of (7-5), we only need to bound the remainder term

$$\mathcal{R}^{i_1,j_1} f := \Pi_{\mathbb{S}^{i_1,j_1}_{\boldsymbol{\mathcal{D}}} f} b - \mathbb{S}^{i_1,j_1}_{\boldsymbol{\mathcal{D}}} \Pi_f b.$$

The proof is somewhat similar to that of the full mixed paraproducts, in that we combine each of these terms

$$\Pi_{\mathbb{S}^{i_1,j_1}_{\boldsymbol{\mathcal{D}}} f} b = \sum_{R_1\times R_2} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_1}}} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2}) \hat{f}(P_1^{\epsilon_1}\times R_2^{\epsilon_2}) \biggl(\sum_{Q_2\supsetneq R_2} \langle b\rangle_{Q_1\times Q_2} h^{\delta_2}_{Q_2}(R_2) h^{\delta_2}_{Q_2}(x_2)\biggr) h^{\delta_1}_{Q_1}(x_1),$$

$$\mathbb{S}^{i_1,j_1}_{\boldsymbol{\mathcal{D}}} \Pi_f b = \sum_{R_1\times R_2} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2}) \hat{f}(P_1^{\epsilon_1}\times R_2^{\epsilon_2}) \langle b\rangle_{P_1\times R_2} h^{\delta_1}_{Q_1}(x_1) \otimes \frac{\mathbb{1}_{R_2}(x_2)}{|R_2|},$$

with a third term

$$T := \sum_{R_1\times R_2} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2}) \hat{f}(P_1^{\epsilon_1}\times R_2^{\epsilon_2}) \langle b\rangle_{Q_1\times R_2} h^{\delta_1}_{Q_1} \otimes \frac{\mathbb{1}_{R_2}}{|R_2|}.$$

As before, expanding the indicator function in T into its Haar series, we may combine T with $\Pi_{\mathbb{S}_{\mathcal{D}}^{i_1,j_1}f}b$:

$$\Pi_{\mathbb{S}_{\mathcal{D}}^{i_1,j_1}f}b - T = \sum_{R_1\times R_2}\sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2})\hat{f}(P_1^{\epsilon_1}\times R_2^{\epsilon_2})T_b(x_2)h_{Q_1}^{\delta_1}(x_1),$$

where

$$\begin{aligned}T_b(x_2) &= \sum_{Q_2\supsetneq R_2}(\langle b\rangle_{Q_1\times Q_2} - \langle b\rangle_{Q_1\times P_2})h_{Q_2}^{\delta_2}(R_2)h_{Q_2}^{\delta_2}(x_2)\\ &= \sum_{Q_2\supsetneq R_2}\Bigg(\sum_{P_2:R_2\subsetneq P_2\subset Q_2}\Big\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}^{\tau_2}\Big\rangle h_{P_2}^{\tau_2}(R_2)\Bigg)h_{Q_2}^{\delta_2}(R_2)h_{Q_2}^{\delta_2}(x_2).\end{aligned}$$

We analyze this term depending on the relationship of P_2 with Q_2.

<u>Case 1</u>: $P_2\subsetneq Q_2$. Then

$$T_b(x_2) = \sum_{P_2\supsetneq R_2}\Big\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}^{\tau_2}\Big\rangle h_{P_2}^{\tau_2}(R_2)\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|},$$

which gives the operator

$$\begin{aligned}&\sum_{Q_1\times P_2}\Big\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}^{\tau_2}\Big\rangle h_{Q_1}^{\tau_1}(x_1)\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\Bigg(\sum_{P_1\in(Q_1^{(j_1)})_{i_1}}\sum_{R_2\subsetneq P_2}\hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2})\widehat{H_{P_1}^{\epsilon_1}f}(R_2^{\epsilon_2})h_{P_2}^{\tau_2}(R_2)\Bigg)\\ &\qquad= \sum_{Q_1\times P_2}\Big\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h_{P_2}^{\tau_2}\Big\rangle h_{Q_1}^{\tau_1}(x_1)\frac{\mathbb{1}_{P_2}(x_2)}{|P_2|}\Bigg(\sum_{P_1\in(Q_1^{(j_1)})_{i_1}}\langle\Pi^*_{a_{P_1Q_1R_1}}(H_{P_1}^{\epsilon_1}f), h_{P_2}^{\tau_2}\rangle_{\mathbb{R}^{n_2}}\Bigg)\\ &\qquad= \pi^*_{b;(1,0)}F,\end{aligned}$$

where

$$F := \sum_{Q_1}\Bigg(\sum_{P_1\in(Q_1^{(j_1)})_{i_1}}\Pi^*_{a_{P_1Q_1R_1}}(H_{P_1}^{\epsilon_1}f)(x_2)\Bigg)h_{Q_1}^{\delta_1}(x_1).$$

Now

$$\|\pi^*_{b;(1,0)}F\|_{L^p(\lambda)}\lesssim\|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}\|F\|_{L^p(\mu)},$$

so we are done if we can show that

$$\|F\|_{L^p(\mu)}\lesssim\|f\|_{L^p(\mu)}. \tag{7-6}$$

Take $g\in L^{p'}(\mu')$. Then

$$|\langle F, g\rangle|\le\sum_{Q_1}\sum_{P_1\in(Q_1^{(j_1)})_{i_1}}|\langle\Pi^*_{a_{P_1Q_1R_1}}(H_{P_1}^{\epsilon_1}f), H_{Q_1}^{\delta_1}g\rangle_{\mathbb{R}^{n_2}}|.$$

Notice that we may write

$$\langle\Pi^*_{a_{P_1Q_1R_1}}(H_{P_1}^{\epsilon_1}f), H_{Q_1}^{\delta_1}g\rangle_{\mathbb{R}^{n_2}} = \langle a_{P_1Q_1R_1}, \phi_{P_1Q_1R_1}\rangle_{\mathbb{R}^{n_2}},$$

where

$$\phi_{P_1Q_1R_1}(x_2) = \sum_{R_2}\widehat{H_{P_1}^{\epsilon_1}f}(R_2^{\delta_2})\langle H_{Q_1}^{\delta_1}g\rangle_{R_2}h_{R_2}^{\delta_2}(x_2).$$

Then

$$|\langle F, g\rangle| \le \sum_{Q_1} \sum_{P_1\in(Q_1^{(j_1)})_{i_1}} \|a_{P_1Q_1R_1}\|_{\mathrm{BMO}(\mathbb{R}^{n_2})} \|S_{\mathcal{D}_2}\phi_{P_1Q_1R_1}\|_{L^1(\mathbb{R}^{n_2})}$$

$$\le 2^{(-n_1/2)(i_1+j_1)} \sum_{R_1} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} \int_{\mathbb{R}^{n_2}} \biggl(\sum_{R_2} |\widehat{H_{P_1}^{\epsilon_1} f}(R_2^{\delta_2})|^2 \langle |H_{Q_1}^{\delta_1} g|\rangle_{R_2}^2 \frac{\mathbb{1}_{R_2}(x_2)}{|R_2|}\biggr)^{1/2} dx_2$$

$$\le 2^{(-n_1/2)(i_1+j_1)} \int_{\mathbb{R}^{\vec{n}}} \sum_{R_1} \sum_{\substack{P_1\in(R_1)_{i_1}\\ Q_1\in(R_1)_{j_2}}} (M_{\mathcal{D}_2} H_{Q_1}^{\delta_1} g)(x_2)(S_{\mathcal{D}_2} H_{P_1}^{\epsilon_1} f)(x_2) \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}\, dx.$$

The integral above is bounded by

$$\int_{\mathbb{R}^{\vec{n}}} \biggl(\sum_{R_1}\biggl(\sum_{P_1\in(R_1)_{i_1}} (S_{\mathcal{D}_2} H_{P_1}^{\epsilon_1} f)(x_2)\biggr)^2 \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}\biggr)^{1/2} \biggl(\sum_{R_1}\biggl(\sum_{P_1\in(R_1)_{i_1}} (S_{\mathcal{D}_2} H_{P_1}^{\epsilon_1} f)(x_2)\biggr)^2 \frac{\mathbb{1}_{R_1}(x_1)}{|R_1|}\biggr)^{1/2} dx$$

$$= \int_{\mathbb{R}^{\vec{n}}} ([SS_{\mathcal{D}_2}]^{i_1,0} f)([SM_{\mathcal{D}_2}]^{j_1,0} g)\, dx \le \|[SS_{\mathcal{D}_2}]^{i_1,0} f\|_{L^p(\mu)} \|[SM_{\mathcal{D}_2}]^{j_1,0} g\|_{L^{p'}(\mu')}$$

$$\lesssim 2^{(n_1/2)(i_1+j_1)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu')} \quad \text{by (3-3).}$$

The desired estimate in (7-6) is now proved.

Case 2a: $P_2 = Q_2$ and $\tau_2 \neq \delta_2$. Then

$$T_b(x_2) = \sum_{Q_2\supsetneq R_2} \biggl\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes h_{Q_2}^{\tau_2}\biggr\rangle \frac{1}{\sqrt{|Q_2|}} h_{Q_2}^{\tau_2+\delta_2}(R_2) h_{Q_2}^{\delta_2}(x_2),$$

giving rise to the operator

$$\sum_{Q_1\times Q_2} \biggl\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes h_{Q_2}^{\tau_2}\biggr\rangle \biggl(\sum_{P_1\in(Q_1^{(j_1)})_{i_1}} \langle \Pi^*_{a_{P_1Q_1R_1}}(H_{P_1}^{\epsilon_1} f), h_{Q_2}^{\tau_2+\delta_2}\rangle_{\mathbb{R}^{n_2}}\biggr) \frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{\delta_1} \otimes h_{Q_2}^{\delta_2} = \gamma_{b;(1,0)} F,$$

which is handled as in the previous case.

Case 2b: $P_2 = Q_2$ and $\tau_2 = \delta_2$. In this case, $T_b(x_2)$ gives rise to the operator

$$T' := \sum_{Q_1\times Q_2} \biggl\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes h_{Q_2}^{\delta_2}\biggr\rangle h_{Q_1}^{\delta_1} \otimes h_{Q_2}^{\delta_2} \sum_{P_1\in(Q_1^{(j_1)})_{i_1}} \frac{1}{|Q_2|} \sum_{R_2\subsetneq Q_2} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2}) \widehat{H_{P_1}^{\epsilon_1} f}(R_2^{\epsilon_2}).$$

Now define

$$F_\tau := \sum_{Q_1} \biggl(\sum_{P_1\in(Q_1^{(j_1)})_{i_1}} \Pi^*_{a^\tau_{P_1Q_1R_1}}(H_{P_1}^{\epsilon_1} f)(x_2)\biggr) h_{Q_1}^{\delta_1}(x_1),$$

just as we defined F before, except now to every function $a_{P_1Q_1R_1}$ we apply the martingale transform

$$a_{P_1Q_1R_1} \mapsto a^\tau_{P_1Q_1R_1} = \sum_{R_2} \tau_{R_2}^{\epsilon_2} \hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2}) h_{R_2}^{\epsilon_2}, \quad \text{where } \tau_{R_2}^{\epsilon_2} := \begin{cases} +1 & \text{if } \widehat{H_{P_1}^{\epsilon_1} f}(R_2^{\epsilon_2}) \ge 0,\\ -1 & \text{otherwise.}\end{cases}$$

Since this does not increase the BMO($\mathbb{R}^{n_2}$) norms of the $a_{P_1Q_1R_1}$ functions, the estimate (7-6) still holds: $\|F_\tau\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)}$.

Moreover, note that

$$\langle \Pi^*_{a^\tau_{P_1Q_1R_1}}(H^{\epsilon_1}_{P_1} f)\rangle_{Q_2} = \sum_{R_2} \underbrace{\hat{a}^\tau_{P_1Q_1R_1}(R_2^{\epsilon_2})\widehat{H^{\epsilon_1}_{P_1} f}(R_2^{\epsilon_2})}_{\geq 0} \frac{|R_2\cap Q_2|}{|R_2||Q_2|}$$

and that

$$\pi_{b;(1,0)}F_\tau = \sum_{Q_1\times Q_2}\left\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h^{\delta_2}_{Q_2}\right\rangle \sum_{P_1\in (Q_1^{(j_1)})_{i_1}} \langle \Pi^*_{a^\tau_{P_1Q_1R_1}}(H^{\epsilon_1}_{P_1} f)\rangle_{Q_2} h^{\delta_1}_{Q_1}\otimes h^{\delta_2}_{Q_2}.$$

Then

$$\begin{aligned} S^2_{\mathcal{D}} T' &\leq \sum_{Q_1\times Q_2}\left|\left\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h^{\delta_2}_{Q_2}\right\rangle\right|^2 \left(\sum_{P_1\in (Q_1^{(j_1)})_{i_1}} \frac{1}{|Q_2|}\sum_{R_2\subsetneq Q_2} |\hat{a}_{P_1Q_1R_1}(R_2^{\epsilon_2})\widehat{H^{\epsilon_1}_{P_1} f}(R_2^{\epsilon_2})|\right)^2 \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|} \\ &\leq \sum_{Q_1\times Q_2}\left|\left\langle b, \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes h^{\delta_2}_{Q_2}\right\rangle\right|^2 \left(\sum_{P_1\in (Q_1^{(j_1)})_{i_1}} \langle \Pi^*_{a^\tau_{P_1Q_1R_1}}(H^{\epsilon_1}_{P_1} f)\rangle_{Q_2}\right)^2 \frac{\mathbb{1}_{Q_1}}{|Q_1|}\otimes\frac{\mathbb{1}_{Q_2}}{|Q_2|} \\ &= S^2_{\mathcal{D}}(\pi_{b;(1,0)}F_\tau). \end{aligned}$$

Finally, this gives us

$$\begin{aligned} \|T'\|_{L^p(\lambda)} &\simeq \|S_{\mathcal{D}}T'\|_{L^p(\lambda)} \leq \|S_{\mathcal{D}}\pi_{b;(1,0)}F_\tau\|_{L^p(\lambda)} \simeq \|\pi_{b;(1,0)}F_\tau\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}\|F_\tau\|_{L^p(\mu)} \\ &\lesssim \|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)}\|f\|_{L^p(\mu)}. \end{aligned}$$

This proves that $\Pi_{\mathbb{S}^{i_1,j_1}_{\mathcal{D}} f} b - T$ obeys the desired bound, and the case $T - \mathbb{S}^{i_1,j_1}_{\mathcal{D}}\Pi_f b$ is handled similarly. □

7E. *Proof of Theorem 1.4.* Having now proved all the one-weight inequalities for dyadic shifts, we may conclude that

$$\|\mathbb{S}^{\vec{i},\vec{j}}_{\mathcal{D}} : L^p(w)\to L^p(w)\| \lesssim 1$$

for all $w\in A_p(\mathbb{R}^{\vec{n}})$. For the cancellative shifts, this was proved in (7-2). For the noncancellative shifts, the first two types are simply paraproducts with symbol $\|a\|_{\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{\vec{n}})}\leq 1$, while the third type, a partial paraproduct, was proved to be bounded on $L^p(w)$ in Proposition 7.6.

Theorem 1.4 now follows trivially from Martikainen's representation theorem, Theorem 7.1: Take $f\in L^p(w)$ and $g\in L^{p'}(w')$. Then

$$\begin{aligned} |\langle Tf, g\rangle| &\leq C_T\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2}\sum_{\vec{i},\vec{j}\in\mathbb{Z}^2_+} 2^{-\max(i_1,j_1)\delta/2}2^{-\max(i_2,j_2)\delta/2}|\langle \mathbb{S}^{\vec{i},\vec{j}}_{\mathcal{D}} f, g\rangle| \\ &\lesssim \|f\|_{L^p(w)}\|g\|_{L^{p'}(w')}\sum_{\vec{i},\vec{j}\in\mathbb{Z}^2_+} 2^{-\max(i_1,j_1)\delta/2}2^{-\max(i_2,j_2)\delta/2} \\ &\simeq \|f\|_{L^p(w)}\|g\|_{L^{p'}(w')}. \end{aligned}$$

□

8. The unweighted case of higher-order Journé commutators

Here is the definition of the BMO spaces which are in between little BMO and product BMO.

Let $b : \mathbb{R}^{\vec{d}} \to \mathbb{C}$ with $\vec{d} = (d_1, \dots, d_t)$. Take a partition $\mathcal{I} = \{I_s : 1 \le s \le l\}$ of $\{1, 2, \dots, t\}$ so that $\dot{\bigcup}_{1\le s\le l} I_s = \{1, 2, \dots, t\}$. We say that $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ if for any choice $\boldsymbol{v} = (v_s)$, $v_s \in I_s$, we have b is uniformly in product BMO in the variables indexed by v_s. We call a BMO space of this type a "little product BMO". If for any $\vec{x} = (x_1, \dots, x_t) \in \mathbb{R}^{\vec{d}}$ we define $\vec{x}_{\hat{\boldsymbol{v}}}$ by removing those variables indexed by v_s, the little product BMO norm becomes

$$\|b\|_{\mathrm{BMO}_{\mathcal{I}}} = \max_{\boldsymbol{v}} \bigl\{ \sup_{\vec{x}_{\hat{\boldsymbol{v}}}} \|b(\vec{x}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}} \bigr\},$$

where the BMO norm is product BMO in the variables indexed by v_s.

In [Ou et al. 2016] it was proved that commutators involving tensor products of Riesz transforms in L^p are a testing class for these BMO spaces:

Theorem 8.1 (Ou, Petermichl and Strouse). *Let $\vec{j} = (j_1, \dots, j_t)$ with $1 \le j_k \le d_k$ and let for each $1 \le s \le l$, $\vec{j}^{(s)} = (j_k)_{k\in I_s}$ be associated a tensor product of Riesz transforms $\vec{R}_{s,\vec{j}^{(s)}} = \bigotimes_{k\in I_s} R_{k,j_k}$; here R_{k,j_k} are j_k-th Riesz transforms acting on functions defined on the k-th variable. We have the two-sided estimate*

$$\|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[\vec{R}_{1,\vec{j}^{(1)}}, \dots, [\vec{R}_{t,\vec{j}^{(t)}}, b], \dots]\|_{L^p(\mathbb{R}^{\vec{d}}) \to L^p(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

It was also proved that the estimate self-improves to paraproduct-free Journé commutators in L^2, in the sense T is paraproduct free $T(1 \otimes \cdot\,) = T(\,\cdot \otimes 1) = T^*(1 \otimes \cdot\,) = T^*(\,\cdot \otimes 1) = 0$.

Theorem 8.2 (Ou, Petermichl and Strouse). *Let us consider $\mathbb{R}^{\vec{d}}$, $\vec{d} = (d_1, \dots, d_t)$, with a partition $\mathcal{I} = (I_s)_{1\le s\le l}$ of $\{1, \dots, t\}$ as discussed before. Let $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ and let T_s denote a multiparameter paraproduct-free Journé operator acting on function defined on $\bigotimes_{k\in I_s} \mathbb{R}^{d_k}$. Then we have the estimate*

$$\|[T_1, \dots, [T_l, b], \dots]\|_{L^2(\mathbb{R}^{\vec{d}}) \to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

This estimate was generalized somewhat in [Ou and Petermichl 2018] in that the paraproduct-free condition was slightly weakened; the considerations in the present text in combination with arguments from [Dalenc and Ou 2016; Ou et al. 2016] to pass to the iterated case, readily give us the following full result, for all Journé operators and all p:

Theorem 8.3. *Let us consider $\mathbb{R}^{\vec{d}}$, $\vec{d} = (d_1, \dots, d_t)$, with a partition $\mathcal{I} = (I_s)_{1\le s\le l}$ of $\{1, \dots, t\}$ as discussed before. Let $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ and let T_s denote a multiparameter Journé operator acting on functions defined on $\bigotimes_{k\in I_s} \mathbb{R}^{d_k}$. Then we have the estimate*

$$\|[T_1, \dots, [T_l, b], \dots]\|_{L^p(\mathbb{R}^{\vec{d}}) \to L^p(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Acknowledgements

We are grateful to Jill Pipher and Yumeng Ou for helpful suggestions, especially in pointing out to us the connection with [Fefferman 1987; 1988]. Holmes and Petermichl would like to thank MSRI for their support during January 2017, as well as Michigan State University's Department of Mathematics for their hospitality in May 2017. Finally, we are grateful to the referee for the careful reading of our paper and the very valuable suggestions to improve the presentation.

References

[Bagby and Kurtz 1985] R. J. Bagby and D. S. Kurtz, "$L(\log L)$ spaces and weights for the strong maximal function", *J. Analyse Math.* **44** (1985), 21–31. MR Zbl

[Bloom 1985] S. Bloom, "A commutator theorem and weighted BMO", *Trans. Amer. Math. Soc.* **292**:1 (1985), 103–122. MR Zbl

[Coifman et al. 1976] R. R. Coifman, R. Rochberg, and G. Weiss, "Factorization theorems for Hardy spaces in several variables", *Ann. of Math.* (2) **103**:3 (1976), 611–635. MR Zbl

[Cotlar and Sadosky 1996] M. Cotlar and C. Sadosky, "Two distinguished subspaces of product BMO and Nehari-AAK theory for Hankel operators on the torus", *Integral Equations Operator Theory* **26**:3 (1996), 273–304. MR Zbl

[Dalenc and Ou 2016] L. Dalenc and Y. Ou, "Upper bound for multi-parameter iterated commutators", *Publ. Mat.* **60**:1 (2016), 191–220. MR Zbl

[Fefferman 1987] R. Fefferman, "Harmonic analysis on product spaces", *Ann. of Math.* (2) **126**:1 (1987), 109–130. MR Zbl

[Fefferman 1988] R. Fefferman, "A^p weights and singular integrals", *Amer. J. Math.* **110**:5 (1988), 975–987. MR Zbl

[Fefferman and Stein 1982] R. Fefferman and E. M. Stein, "Singular integrals on product spaces", *Adv. in Math.* **45**:2 (1982), 117–143. MR Zbl

[Grafakos 2004] L. Grafakos, *Classical and modern Fourier analysis*, Pearson Education, Upper Saddle River, NJ, 2004. MR Zbl

[Grau de la Herrán 2016] A. Grau de la Herrán, "Comparison of $T1$ conditions for multi-parameter operators", *Proc. Amer. Math. Soc.* **144**:6 (2016), 2437–2443. MR Zbl

[Holmes et al. 2017] I. Holmes, M. T. Lacey, and B. D. Wick, "Commutators in the two-weight setting", *Math. Ann.* **367**:1-2 (2017), 51–80. MR Zbl

[Hytönen 2012] T. P. Hytönen, "The sharp weighted bound for general Calderón–Zygmund operators", *Ann. of Math.* (2) **175**:3 (2012), 1473–1506. MR Zbl

[Journé 1985] J.-L. Journé, "Calderón–Zygmund operators on product spaces", *Rev. Mat. Iberoamericana* **1**:3 (1985), 55–91. MR Zbl

[Martikainen 2012] H. Martikainen, "Representation of bi-parameter singular integrals by dyadic operators", *Adv. Math.* **229**:3 (2012), 1734–1761. MR Zbl

[Martikainen and Orponen 2016] H. Martikainen and T. Orponen, "Some obstacles in characterising the boundedness of bi-parameter singular integrals", *Math. Z.* **282**:1-2 (2016), 535–545. MR Zbl

[Muckenhoupt and Wheeden 1976] B. Muckenhoupt and R. L. Wheeden, "Weighted bounded mean oscillation and the Hilbert transform", *Studia Math.* **54**:3 (1976), 221–237. MR Zbl

[Ou and Petermichl 2018] Y. Ou and S. Petermichl, "Little bmo and Journé commutators", pp. 207–219 in *Harmonic analysis, function theory, operator theory and applications* (Bordeaux, 2015), edited by P. Jaming et al., Theta Foundation International Book Series of Mathematical Texts **22**, Theta Foundation, Bucharest, 2018.

[Ou et al. 2016] Y. Ou, S. Petermichl, and E. Strouse, "Higher order Journé commutators and characterizations of multi-parameter BMO", *Adv. Math.* **291** (2016), 24–58. MR Zbl

[Petermichl 2000] S. Petermichl, "Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol", *C. R. Acad. Sci. Paris Sér. I Math.* **330**:6 (2000), 455–460. MR Zbl

[Wu 1992] S. J. Wu, "A wavelet characterization for weighted Hardy spaces", *Rev. Mat. Iberoamericana* **8**:3 (1992), 329–349. MR Zbl

Received 14 Jun 2017. Revised 10 Sep 2017. Accepted 24 Oct 2017.

IRINA HOLMES: holmesir@math.msu.edu
Department of Mathematics, Michigan State University, East Lansing, MI, United States

STEFANIE PETERMICHL: stefanie.petermichl@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse, France

BRETT D. WICK: wick@math.wustl.edu
Department of Mathematics, Washington University - St. Louis, St. Louis, MO, United States

Analysis & PDE

msp.org/apde

See inside back cover or msp.org/apde for submission instructions.

The subscription price for 2018 is US $275/year for the electronic version, and $480/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscriber address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

ANALYSIS & PDE

Volume 11 No. 7 2018