WEIGHTED LITTLE BMO AND TWO-WEIGHT INEQUALITIES FOR JOURNÉ COMMUTATORS

IRINA HOLMES†, STEFANIE PETERMICHL§, AND BRETT D. WICK‡

Abstract. 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 $\mu$ and $\lambda$ 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 $\|b\|_{\text{bmo}(\nu)} \gtrsim \sup \|[b, R_1^k R_2^l] : L^p(\mu) \to L^p(\lambda)\|$, where $R_1^k$ and $R_2^l$ 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.

Contents

1. Introduction and Statement of Main Results
2. Background and Notation
3. Biparameter Dyadic Square Functions
4. Biparameter Weighted BMO Spaces
5. Proof of the Lower Bound
6. Biparameter Paraproducts
7. Commutators with Journé Operators
8. The unweighted case of higher order Journé commutators

References

1. Introduction and Statement of Main Results

In 1985, Bloom [2] proved a two-weight version of the celebrated commutator theorem of Coifman, Rochberg and Weiss [3]. Specifically, [2] 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:

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

where $\mu, \lambda$ 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 [11], using the modern dyadic methods started by [19] and continued in [12]. 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 [14].

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 [4]. The main results are:

2000 Mathematics Subject Classification. Primary: 42, 42A, 42B, 42B20, 42B25, 42A50, 42A40.
Key words and phrases. Commutators, Calderón–Zygmund Operators, Bounded Mean Oscillation, Weights, Journé Operators, Multiparameter.

† This material is based upon work supported by the National Science Foundation under Award No. 1606270.
§ Research supported in part by ERC project CRHSHarMa no. DLV-682402.
‡ Research supported in part by a National Science Foundation DMS grant #1560955.
Theorem 1.1 (Upper Bound). Let $T$ be a biparameter Journé operator on $\mathbb{R}^d = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, as defined in Section 7.1. Let $\mu$ and $\lambda$ be $A_p(\mathbb{R}^d)$ weights, $1 < p < \infty$, and define $\nu := \mu^{1/p} \lambda^{-1/p}$. Then

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

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

We make a few remarks about the proof of this result. At its core, the strategy is the same as in [11], 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 [11], $T$ was a Calderón-Zygmund operator on $\mathbb{R}^n$, and the representation theorem was that of Hytönen [12]. In the present paper, $T$ is a biparameter Journé operator on $\mathbb{R}^d = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ (see Section 7.1) and we use Martikainen’s representation theorem [14] to reduce the problem to commutators $[b, S_D]$, where $S_D$ is now a biparameter dyadic shift. These can be cancellative, i.e. all Haar functions have mean zero, (defined in Section 7.3), or non-cancellative (defined in Section 7.4). 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 $S_{D_1}$ and $S_{D_2}$ 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 $a_{p_1 Q_1 R_1 p_2 Q_2 R_2}$ in the biparameter shifts to “separate” into a product $a_{p_1 Q_1 R_1} \cdot a_{p_2 Q_2 R_2}$, 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 non-cancellative shifts. As outlined in Section 7.4, 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.

These methods were considered previously in [17] and [18] for the unweighted, $p = 2$ case. In [17] it was shown that

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

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 non-cancellative shifts remained open. This gap was partially filled in [18], which treats the case of non-cancellative shifts of standard paraproduct type. So the case of general Journé operators, which includes non-cancellative 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 [15] for a wonderful discussion of this issue. We fill this gap in Section 7.4, where we prove two-weight bounds of the type

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

where $S_D$ is a non-cancellative shift. The same is proved for cancellative shifts in Section 7.3.

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

$$[b, S_D^{i,j}]f = \sum [P_{b}, S_D^{i,j}]f + \sum [P_{b}, S_D^{i,j}]f + R_{i,j}f.$$

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.
we outline below the general strategy we will employ. This applies to Theorem
things become very technical, so bounding the remainder terms is no easy task. To help guide the reader,

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

Martikainen representation theorem

\[ \| [b, S^2_{D_1}] : L^p(\mu) \to L^p(\lambda) \| \lesssim \| b \|_{bmo(\nu)} \]

with at most polynomial bounds in \( i, j \).

\[ \begin{align*}
\text{Cancellative Shifts:} & \quad \text{Theorem 7.2} \\
\text{Non-Cancellative} & \quad \text{Shifts:} \\
& \quad \text{Two-weight bounds} \\
& \quad \text{for paraproducts:} \quad \text{Section 6}
\end{align*} \]

\[ \begin{align*}
\text{Full standard paraproduct:} & \quad \text{Theorem 7.3} \\
\text{Full mixed paraproduct:} & \quad \text{Theorem 7.4} \\
\text{Partial paraproduct:} & \quad \text{Theorem 7.5}
\end{align*} \]

\[ \begin{align*}
\| \langle O(b, f), g \rangle \| & \lesssim \| b \|_{BMO(\nu)} \| f \|_{L^p(\mu)} \| g \|_{L^p(\lambda')} \]
\end{align*} \]

Figure 1. Strategy for Theorem 1.1

Once the paraproducts are bounded, all that is left is to bound the so-called “remainder term” \( R_{i,j}f \), of the form \( \Pi_b f - \Pi_{f,b} \), 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 \( O(b, f) \), involving both \( b \in bmo(\nu) \) and \( f \in L^p(\mu) \). We want to show \( \| O(b, f) : L^p(\mu) \to L^p(\lambda) \| \lesssim \| b \|_{bmo(\nu)} \). Using duality this amounts to showing that

\[ | \langle O(b, f), g \rangle | \lesssim \| b \|_{BMO(\nu)} \| f \|_{L^p(\mu)} \| g \|_{L^p(\lambda')} \]

2. Some of these operators \( O(b, f) \) 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 \( \langle b, \delta_{Q_1} \times \mathbb{1}_{Q_2}/|Q_2| \rangle \). 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 \( \langle O(b, f), g \rangle \). If \( O(b, f) \) involves full Haar coefficients of \( b \), we use duality with \( product BMO \) and obtain

\[ | \langle O(b, f), g \rangle | \lesssim \| b \|_{BMO(\nu)} \| S_D \phi(f, g) \|_{L^1(\nu)} \]

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

\[ | \langle O(b, f), g \rangle | \lesssim \| b \|_{bmo(\nu)} \| S_{D_1} \phi(f, g) \|_{L^1(\nu)} \]

where \( S_{D_1} \) is the dyadic square function in the first variable. Obviously this is replaced with \( S_{D_2} \) if the Haar coefficient on \( b \) is in the second variable.

3. Then the next goal is to show that

\[ S_D \phi(f, g) \lesssim (O_1 f)(O_2 g) \]

where \( O_{1,2} \) 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 $R_{1,2}^\tau f$ seems intractable using this method, since it involves average terms $\langle b \rangle_{Q_t \times Q_s}$ instead of Haar coefficients of $b$. So they key here is to decompose these terms in some convenient form.

In Section 7.4, dealing with non-cancellative 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 are really dealing with terms of the form $|\langle O(a, b, f), g \rangle|$, where now the operator $O$ involves a function $b$ in the weighted little 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 $\phi$ 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 $O_{1,2}$ to achieve step 5, 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 non-rigorous about the notations 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 7.4.

1. The full standard paraproduct – Theorem 7.3. This case only requires simple martingale transforms ($a_\tau$ and $g_\tau$, which have all non-negative Haar coefficients), and otherwise follows the strategy outlined above. However, we already start to see the operators $O_{1,2}$ becoming strange compositions of “standard” operators and unweighted paraproducts, such as

$$S \mathcal{D} \phi \leq (M_3 \Pi^*_a, g_\tau)(S \mathcal{D} f).$$

2. The full mixed paraproduct – Theorem 7.4. Here we introduce the idea of combining the remainder term $\Pi g f b - \Pi f b$ with a third term $T$, and we analyze $(\Pi g f b - T)$ and $(T - \Pi f b)$ separately. This allows us to express the remainder as

$$\sum |P_{ab}| f + T_{ab}^{(1,0)} f - T_{ab}^{(0,1)} f,$$

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 $T_{ab}$ operators satisfy

$$|\langle T_{a,b} f, g \rangle| \lesssim \|b\|_{bmo(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda)}.$$ Here is where we employ the strategy outlined earlier, combined with a martingale transform $a_\tau$ applied to $a$. Interestingly, this transform depends on the particular argument $f$ of $[b, S \mathcal{D}]f$. This will be absorbed in the end by the BMO norm of the symbol for $S \mathcal{D}$, 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 $P_{ab} F$, where $F$ is a term depending on $a$ and $f$. So we are done if we can show that $\|F\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)}$. Without getting too technical about the notations, 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

$$\langle f, g \rangle = \sum \langle \Pi^*_a P_{QR} f, g \rangle,$$

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 $H^1 - \text{BMO}$ 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 $F_\tau$ 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 BMO$_T$ classes in between Chang-Fefferman’s product BMO and Cotlar-Sadosky’s little BMO. In the same way as in [17] we deduce a corollary from Theorem 1.1:
we review the necessary background, both one- and bi-

types of dyadic square functions we will

of Theorem 7.5 to prove the lower bound result. Section

first appeared in R. Fefferman and E. M. Stein [1].

need throughout the rest of the paper. In Section 7, we discuss the weighted and Bloom BMO spaces in the

biparameter setting, and use some of these results in Section

dedicated to biparameter paraproducts, which will be crucial in the final Section

one-weight bounds for the one-parameter dyadic shifts.

paraproduct type is given in Proposition 7.5.

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

following one-weight result for Journé operators, originally due to R. Fefferman:

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

$$\|b\|_{BMO(\nu)} \lesssim \sup_{1 \leq k, l \leq n} \|\langle b, R^1_k \rangle R^2_l \|_{L^p(\mu) \to L^\nu(\lambda)},$$

where $R^1_k$ and $R^2_l$ are the Riesz transforms acting in the first and second variable, 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 $\mathbb{R}^d = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$. Then $T$ is bounded $L^p(w) \to L^p(w)$ for all $w \in A_p(\mathbb{R}^d)$, $1 < p < \infty$.

A version of Theorem 1.4 first appeared in R. Fefferman and E. M. Stein [6], with restrictive assumptions on the kernel. Subsequently the kernel assumptions were weakened significantly by R. Fefferman in [7], 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, R. Fefferman improved his own result in [8], 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 [7].

Our proof in Section 7.5 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 Theorem 7.5 is only based on one-weight bounds for the biparameter dyadic shifts, of the form

$$\|S^2_D : L^p(w) \to L^p(w)\| \lesssim 1.$$  

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 non-cancellative shifts of partial paraproduct type is given in Proposition 7.6. Once we have (1.1), the proof of Theorem 1.4 follows immedi-
ately 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- and bi-

parameter, 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 the final Section 7, which proves the upper

bound by an appeal to Martikainen’s [14] representation theorem. Finally, we prove Theorem 1.4.

2. Background and Notation

In this section we review some of the basic building blocks of one-parameter dyadic harmonic analysis on

$\mathbb{R}^n$, followed by their biparameter versions for $\mathbb{R}^d := \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$.

2.1. Dyadic Grids on $\mathbb{R}^n$. Let $D_0 := \{2^{-k}([0,1]^n + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^n\}$ 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 $D_\omega$:

$$D_\omega := \{Q + \omega : Q \in D_0\},$$

where $Q + \omega := Q + \sum_{j:2^{-j} < \ell(Q)} 2^{-j} \omega_j,$
and $l(Q)$ denotes the side length of a cube $Q$. The indexing parameter $\omega$ is rarely relevant in what follows: it only appears when we are dealing with $E_\omega$ – expectation with respect to the standard probability measure on the space of parameters $\omega$. 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 any dyadic grid $D$ on $\mathbb{R}^n$:

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

For every $Q \in D$ and every non-negative integer $k$ we define:

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

### 2.2. The Haar system on $\mathbb{R}^n$.

Recall that every dyadic interval $I$ in $\mathbb{R}$ is associated with two Haar functions:

$$h_I^0 := \frac{1}{\sqrt{|I|}} (1_I - 1_{I_1}) \quad \text{and} \quad h_I^1 := \frac{1}{\sqrt{|I|}} 1_I,$$

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

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

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

$$f = \sum_{Q \in D} \hat{f}(Q') h_Q,$$

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

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

$$\langle f \rangle_Q = \sum_{P \in D, P \supseteq Q} \hat{f}(P') h_{P}^\epsilon(Q). \quad (2.1)$$

- Then, if $Q \subseteq R \in D$:

$$\langle f \rangle_Q - \langle f \rangle_R = \sum_{P \in D, Q \subseteq P \subseteq R} \hat{f}(P') h_{P}^\epsilon(Q). \quad (2.2)$$

- For $Q \in D$:

$$1_Q(f - \langle f \rangle_Q) = \sum_{P \in D, P \subseteq Q} \hat{f}(P') h_{P}^\epsilon. \quad (2.3)$$

- 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 $\epsilon$ and $\delta$, and is not the identically $\tilde{1}$ signature. Then

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

Again to simplify notation, we assume throughout this paper that we only write $h_Q^{\epsilon + \delta}$ for distinct signatures $\epsilon$ and $\delta$. 
Given a dyadic grid $\mathcal{D}$, we define the dyadic square function on $\mathbb{R}^n$ by:

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

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

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

2.3. $A_p(\mathbb{R}^n)$ 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

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

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:

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

for all $w \in A_p(\mathbb{R}^n)$, $1 < p < \infty$, where throughout this paper “$A \lesssim B$” denotes $A \leq cB$ 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 non-negative integers $i$ and $j$, define

$$S_{\mathcal{D}}^{i,j}f(x) := \left( \sum_{R \in \mathcal{D}} \left( \sum_{P \in (R)} |\hat{f}(P')|^2 \right)^{1/2} \left( \sum_{Q \in (R)} \frac{1_{Q}(x)}{|Q|} \right) \right)^{1/2}.$$  

It was shown in [11] that

$$\|S_{\mathcal{D}}^{i,j} : L^p(w) \to L^p(w)\| \lesssim 2^{\frac{j}{2}(i+j)},$$

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

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

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

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

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

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 $L^2(\mathbb{R}^n)$, defined by

$$h_{R}(x_1, x_2) := h_{Q_1}^i(x_1) \otimes h_{Q_2}^j(x_2),$$

for all $R = Q_1 \times Q_2 \in \mathcal{D}$ and $\bar{e} = (e_1, e_2)$. We often write

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

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

$$\hat{f}(Q_1^i \times Q_2^j) := \left\langle f, h_{Q_1}^i \otimes h_{Q_2}^j \right\rangle = \int_{\mathbb{R}^n} f(x_1, x_2)h_{Q_1}^i(x_1)h_{Q_2}^j(x_2)\, dx_1\, dx_2.$$
While the averaging formula (2.1) has a straightforward biparameter analogue:

\[
\langle f \rangle_{Q_1 \times Q_2} = \sum_{P_1 \subseteq Q_1 \atop P_2 \subseteq Q_2} \frac{\hat{f}(P_1^1 \times P_2^2) h^1_{P_1}(Q_1) h^2_{P_2}(Q_2)}{\| \hat{f}(P_1^1 \times P_2^2) \|_1}
\]

the expression in (2.3) takes a slightly messier form in two parameters: for any \( R = Q_1 \times Q_2 \)

\[
I_R(f - \langle f \rangle_R) = \sum_{P_1 \subseteq Q_1 \atop P_2 \subseteq Q_2} \frac{\hat{f}(P_1^1 \times P_2^2) h^1_{P_1} \otimes h^2_{P_2}}{\| \hat{f}(P_1^1 \times P_2^2) \|_1} + \sum_{P_1 \subseteq Q_1 \atop P_2 \subseteq Q_2} \left\langle f, \frac{1}{|Q_1|} h^1_{P_1} \otimes \frac{1}{|Q_2|} h^2_{P_2} \right\rangle I_{Q_1} \otimes I_{Q_2}
\]

\[
= \sum_{P_1 \subseteq Q_1 \atop P_2 \subseteq Q_2} \frac{\hat{f}(P_1^1 \times P_2^2) h^1_{P_1} \otimes h^2_{P_2}}{\| \hat{f}(P_1^1 \times P_2^2) \|_1} I_{Q_1} \otimes I_{Q_2} + \sum_{P_1 \subseteq Q_1 \atop P_2 \subseteq Q_2} \left\langle f, \frac{1}{|Q_1|} h^1_{P_1} \otimes \frac{1}{|Q_2|} h^2_{P_2} \right\rangle I_{Q_1} \otimes I_{Q_2}
\]

where for any cubes \( Q_i \in D_i \):

\[
m_{Q_i}(f(x)) := \frac{1}{|Q_i|} \int_{Q_i} f(x_1, x_2) \, dx_1, \text{ and } m_{Q_2}(f(x)) := \frac{1}{|Q_2|} \int_{Q_2} f(x_1, x_2) \, dx_2.
\]

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

2.5. \( A_p(\mathbb{R}^n) \) Weights. A weight \( w(x_1, x_2) \) on \( \mathbb{R}^n \) belongs to the class \( A_p(\mathbb{R}^n) \), for some \( 1 < p < \infty \), provided that

\[
[w]_{A_p} := \sup_R \langle w \rangle_R \left( \frac{1}{w} \right)^{p-1} < \infty,
\]

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

\[
M_S f(x_1, x_2) := \sup_R \langle |f| \rangle_R I_R(x_1, x_2),
\]

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 [1] for all \( w \in A_p(\mathbb{R}^n) \):

\[
w\{ x \in \mathbb{R}^n : M_S f(x) > \lambda \} \lesssim \int_{\mathbb{R}^n} \left( \frac{|f(x)|}{\lambda} \right)^{p} \left( 1 + \log^+ \frac{|f(x)|}{\lambda} \right)^{k-1} \, dw(x).
\]

Moreover, \( w \) belongs to \( A_p(\mathbb{R}^n) \) if and only if \( w \) belongs to the one-parameter classes \( A_p(\mathbb{R}^n_1) \) in each variable separately and uniformly:

\[
[w]_{A_p(\mathbb{R}^n)} \simeq \max \left\{ \text{ess sup}_{x_1 \in \mathbb{R}^{n_1}} [w(x_1, \cdot)]_{A_p(\mathbb{R}^{n_2})}, \text{ ess sup}_{x_2 \in \mathbb{R}^{n_2}} [w(\cdot, x_2)]_{A_p(\mathbb{R}^{n_1})} \right\}.
\]

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

\[
\| f \|_{L^p(w)} = \sup \{ \| \langle f, g \rangle \| : g \in L^{p'}(w'), \| g \|_{L^{p'}(w')} \leq 1 \}.
\]

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 \( w \in A_p(\mathbb{R}^n), 1 < p < \infty \), 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:

\[
[m_{Q_i} w]_{A_p(\mathbb{R}^{n_i})} \leq [w]_{A_p(\mathbb{R}^n)},
\]

for all \( Q_i \subset \mathbb{R}^{n_i}, i \in \{1, 2\}, i \neq j \).
Proof. Fix a cube $Q_1 \subset \mathbb{R}^n$. Then for every $x_2 \in \mathbb{R}^{n_2}$,
\[ |Q_1| = \int_{Q_1} 1 \, dx_1 \leq \left( \int_{Q_1} w(x_1, x_2) \, dx_1 \right)^{1/p} \left( \int_{Q_1} w'(x_1, x_2) \, dx_1 \right)^{1/p'}, \]
and so
\[ (m_{Q_1} w)(x_2) := (m_{Q_1} w)^{1/p'}(x_2) \leq m_{Q_1} w'(x_2). \]
Then for all cubes $Q_2 \subset \mathbb{R}^{n_2}$,
\[ (m_{Q_1} w)_Q (m_{Q_1} w)^{p-1}_{Q_2} \leq \langle w \rangle_{Q_1 \times Q_2} (w)^{p-1}_{Q_1 \times Q_2} \leq [w]_{A_p(\mathbb{R}^d)}, \]
proving the result for $m_{Q_1} w$. The other case follows symmetrically. \qed

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 $w \in A_p(\mathbb{R}^d)$, then there exist positive constants $C, \epsilon, \delta > 0$ (depending only on $\bar{n}, p,$ and $[w]_{A_p(\mathbb{R}^d)}$), such that

i). For all rectangles $R \subset \mathbb{R}^d$,
\[ \left( \frac{1}{|R|} \int_R w(x)^{1+\epsilon} \, dx \right)^{1/\epsilon} \leq C \left( \frac{\int_R w(x) \, dx}{|R|} \right)^{\delta}. \]

ii). For all rectangles $R \subset \mathbb{R}^d$ and all measurable subsets $E \subset R$,
\[ \frac{w(E)}{w(R)} \leq C \left( \frac{|E|}{|R|} \right)^{\delta}. \]

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

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 [9]):

For any $1 < p < \infty$ and $B > 1$, there exist positive constants
\begin{equation}
D = D(n, p, B) \text{ and } \beta = \beta(n, p, B)
\end{equation}
such that for all $v \in A_p(\mathbb{R}^n)$ with $[v]_{A_p(\mathbb{R}^d)} \leq B$, the reverse Hölder condition
\begin{equation}
\left( \frac{1}{|Q|} \int_Q v(t)^{1+\beta} \, dt \right)^{1/\beta} \leq D \left( \frac{\int_Q v(t) \, dt}{|Q|} \right)
\end{equation}
holds for all cubes $Q \subset \mathbb{R}^n$.
It is easy to see that if a weight $w$ satisfies the reverse Hölder condition (2.10) with constants $D, \beta$, then it also satisfies it with any constants $C, \epsilon$ with $C \geq D$ and $\epsilon \leq \beta$.

Now let $w \in A_p(\mathbb{R}^d)$, set $B := [w]_{A_p(\mathbb{R}^d)}$, 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
\[ C^2 := \max(D_1, D_2) \text{ and } \epsilon := \min(\beta_1, \beta_2). \]
For almost all $x_1 \in \mathbb{R}^{n_1}$, the weight $w(x_1, \cdot) \in A_p(\mathbb{R}^{n_2})$ with $[w(x_1, \cdot)]_{A_p(\mathbb{R}^{n_2})} \leq B$, so $w(x_1, \cdot)$ satisfies reverse Hölder with constants $D_2, \beta_2$ and therefore also with constants $\sqrt{C}, \epsilon$. So
\[ \frac{1}{|R|} \int_R w(x)^{1+\epsilon} \, dx = \frac{1}{|Q_1|} \int_{Q_1} \left( \frac{1}{|Q_2|} \int_{Q_2} w(x_1, x_2)^{1+\epsilon} \, dx_2 \right) \, dx_1 \]
\[ \leq \frac{1}{|Q_1|} \int_{Q_1} \left( \frac{\sqrt{C}}{|Q_2|} \int_{Q_2} w(x_1, x_2) \, dx_2 \right)^{1+\epsilon} \, dx_1 \]
\[ = \frac{C(1+\epsilon)/2}{|Q_1|} \int_{Q_1} (m_{Q_2} w(x_1))^1 \, dx_1. \]
By Proposition 2.1, the weight $m_{Q,w} \in A_p(\mathbb{R}^n)$ with $[m_{Q,w}]_{A_p(\mathbb{R}^n)} \leq B$, so this weight satisfies reverse H"older with constants $D_1, \beta_1$ -- and therefore also with constants $\sqrt{C}, \epsilon$. Then the last inequality above gives that
\[
\left( \frac{1}{|R|} \int_R w(x)^{1+\epsilon} \, dx \right)^{\frac{1}{1+\epsilon}} \leq \frac{C}{|Q|} \int_Q m_{Q,w}(x_1) \, dx_1 = \frac{C}{|R|} \int_R w(x) \, dx.
\]

\[\square\]

3. Biparameter Dyadic Square Functions

Throughout this section, fix dyadic rectangles $\mathcal{D} := \mathcal{D}_1 \times \mathcal{D}_2$ on $\mathbb{R}^n$. The dyadic square function associated with $\mathcal{D}$ is then defined in the obvious way:
\[
S_{\mathcal{D}} f(x_1, x_2) := \left( \sum_{R \in \mathcal{D}} |\hat{f}(R^n)|^2 \frac{1}{|R|} \right)^{1/2}.
\]

We also want to look at the dyadic square functions in each variable, namely
\[
S_{\mathcal{D}_1} f(x_1, x_2) := \left( \sum_{Q_1 \in \mathcal{D}_1} |H_{Q_1} f(x_2)|^2 \frac{1}{|Q_1|} \right)^{1/2}; \quad S_{\mathcal{D}_2} f(x_1, x_2) := \left( \sum_{Q_2 \in \mathcal{D}_2} |H_{Q_2} f(x_1)|^2 \frac{1}{|Q_2|} \right)^{1/2},
\]
where for every $Q_i \in \mathcal{D}_i$ and signatures $\epsilon_i$, we denote
\[
H_{Q_i}^\epsilon f(x_2) := \int_{\mathbb{R}^n_1} f(x_1, x_2) h_{Q_i}^{\epsilon, 2}(x_1) \, dx_1; \quad H_{Q_i}^\epsilon f(x_1) := \int_{\mathbb{R}^n_2} f(x_1, x_2) h_{Q_i}^{\epsilon, 2}(x_2) \, dx_2.
\]

Then for any $w \in A_p(\mathbb{R}^n)$:
\[
\|f\|_{L^p(w)} \simeq \|S_{\mathcal{D}} f\|_{L^p(w)} \simeq \|S_{\mathcal{D}_1} f\|_{L^p(w)} \simeq \|S_{\mathcal{D}_2} f\|_{L^p(w)}.
\]

More generally, define the shifted biparameter square function, for pairs $\vec{i} = (i_1, i_2)$ and $\vec{j} = (j_1, j_2)$ of non-negative integers, by:
\[
S_{\mathcal{D}_1}^{\vec{i}, \vec{j}} f := \left( \sum_{R_i \in \mathcal{D}_1} \left( \sum_{P_i \in (R_i)_{\vec{i}}} |\hat{f}(P_1 \times P_2)| \right)^2 \left( \sum_{Q_j \in (R_j)_{\vec{j}}} \frac{1}{|Q_j|} \right)^{1/2} \right)^{1/2}.
\]

We claim that:
\[
\|S_{\mathcal{D}_1}^{\vec{i}, \vec{j}} f\|_{L^p(w)} \rightarrow L^p(w) \| \lesssim 2^{\vec{i}+\vec{j}} \| \lesssim 2^{\vec{i}+\vec{j}},
\]
for all $w \in A_p(\mathbb{R}^n)$, $1 < p < \infty$. 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 [9]):

**Proposition 3.1.** Suppose that an operator $T$ satisfies $\|T : L^2(w) \rightarrow L^2(w)\| \leq A_{C_n}[w]_{A_p}$ for all $w \in A_2(\mathbb{R}^n)$, for some constants $A, C_n$, where the latter only depends on the dimension. Then:
\[
\left\| \left( \sum_j |Tf_j|^2 \right)^{1/2} \right\|_{L^p(w)} \leq A_{C_n}[w]_{A_p}^{\max(1, \frac{\epsilon}{1-\epsilon})} \left( \sum_j |f_j|^2 \right)^{1/2} \right\|_{L^p(w)}
\]
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 $(S_{\mathcal{D}_1}^{\vec{i}, \vec{j}} f)^2 = \sum_{R_i \in \mathcal{D}_1} (S_{\mathcal{D}_2}^{\vec{i}, \vec{j}} F_{R_i})^2$, where
\[
F_{R_i}(x_1, x_2) := \sum_{P_2 \in \mathcal{D}_2} \left( \sum_{P_1 \in (R_i)_{\vec{i}}} |\hat{f}(P_1 \times P_2)| \right) \left( \sum_{Q_j \in (R_j)_{\vec{j}}} \frac{1}{|Q_j|} \right)^{1/2} h_{P_2}^{\epsilon, 2}(x_2).
\]

Then
\[
\|S_{\mathcal{D}_1}^{\vec{i}, \vec{j}} f\|_{L^p(w)}^p = \int_{\mathbb{R}^n_1} \int_{\mathbb{R}^n_2} \left( \sum_{R_i \in \mathcal{D}_1} (S_{\mathcal{D}_2}^{\vec{i}, \vec{j}} F_{R_i}(x_1, x_2))^2 \right)^{p/2} w(x_1, x_2) \, dx_2 \, dx_1.
\]
For almost all fixed \( x_1 \in \mathbb{R}^n \), \( w(x_1, \cdot) \) is in \( A_p(\mathbb{R}^n) \) uniformly, so we may apply Proposition 3.1 and (2.4) to the inner integral and obtain:

\[
\| S_{\mathcal{D}}^{i,j} f \|_{L^p(w)}^p \lesssim 2^{\frac{pn}{2}(i_1+j_2)} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left( \sum_{R_1 \in \mathcal{D}_1} |F_{R_1}(x_1, x_2)|^2 \right)^{p/2} w(x_1, x_2) \, dx_2 \, dx_1.
\]

Now, we can express the integral above as

\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left( S_{\mathcal{D}_1}^{i_1,j_1} f_{\tau}(x_1, x_2) \right)^p w(x_1, x_2) \, dx_1 \, dx_2 \lesssim 2^{\frac{pn}{2}(i_1+j_1)} \| f_{\tau} \|_{L^p(w)},
\]

where \( f_{\tau} = \sum_{P_1 \times P_2} |\hat{f}(P_1 \times P_2^w)| h_{P_1}^{\epsilon} \otimes h_{P_2}^{\epsilon} \) is just a biparameter martingale transform applied to \( f \), and therefore \( \| f \|_{L^p(w)} \lesssim \| f_{\tau} \|_{L^p(w)} \) by passing to the square function. \( \Box \)

### 3.1. Mixed Square and Maximal Functions.

We will later encounter mixed operators such as:

\[
[SM] f(x_1, x_2) := \left( \sum_{Q_1 \in \mathcal{D}_1} \left( M_{D_2 \mathcal{Q}_1} (h_{Q_1}^{\epsilon}, f)(x_2) \right)^2 \frac{1 \mathbb{1}_{Q_1}(x_1)}{|Q_1|} \right)^{1/2},
\]

\[
[MS] f(x_1, x_2) := \left( \sum_{Q_2 \in \mathcal{D}_2} \left( M_{D_1 \mathcal{Q}_2} (h_{Q_2}^{\epsilon}, f)(x_1) \right)^2 \frac{1 \mathbb{1}_{Q_2}(x_2)}{|Q_2|} \right)^{1/2}.
\]

Next we show that these operators are bounded \( L^p(w) \rightarrow L^p(w) \) for all \( w \in A_p(\mathbb{R}^n) \). 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 \( M_{D_2} \) and \( M_{D_1} \) 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 \( \| T : L^2(v) \rightarrow L^2(v) \| \leq C \) for all \( v \in A_2(\mathbb{R}^n) \). Define the following operators on \( \mathbb{R}^n \):

\[
[ST] f(x_1, x_2) := \left( \sum_{Q_1 \in \mathcal{D}_1} \left( T(h_{Q_1}^{\epsilon}, f)(x_2) \right)^2 \frac{1 \mathbb{1}_{Q_1}(x_1)}{|Q_1|} \right)^{1/2},
\]

\[
[TS] f(x_1, x_2) := \left( \sum_{Q_2 \in \mathcal{D}_2} \left( T(h_{Q_2}^{\epsilon}, f)(x_1) \right)^2 \frac{1 \mathbb{1}_{Q_2}(x_2)}{|Q_2|} \right)^{1/2},
\]

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

**Proof.**

\[
\| [ST] f \|_{L^p(w)}^p = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left( \sum_{Q_1 \in \mathcal{D}_1} \left( T(h_{Q_1}^{\epsilon}, f)(x_2) \right)^2 \frac{1 \mathbb{1}_{Q_1}(x_1)}{|Q_1|} \right)^{p/2} w(x_1, x_2) \, dx_2 \, dx_1
\]

\[
\lesssim \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left( \sum_{Q_1 \in \mathcal{D}_1} (h_{Q_1}^{\epsilon})^2(x_2) \frac{1 \mathbb{1}_{Q_1}(x_1)}{|Q_1|} \right)^{p/2} w(x_1, x_2) \, dx_2 \, dx_1
\]

\[
= \| S_{\mathcal{D}_{1}} f \|_{L^p(w)}^p \lesssim \| f \|_{L^p(w)}^p,
\]

where the first inequality follows as before from Proposition 3.1. The proof for \([TS]\) is symmetrical. \( \Box \)

More generally, define shifted versions of these mixed operators:

\[
[ST]_{i_1,j_1} f(x_1, x_2) := \left( \sum_{R_1 \in \mathcal{D}_1} \left( \sum_{P_1 \in (R_1)_{i_1}} T(h_{P_1}^{\epsilon}, f)(x_2) \right)^2 \sum_{Q_2 \in (R_2)_{j_1}} \frac{1 \mathbb{1}_{Q_2}(x_2)}{|Q_2|} \right)^{1/2},
\]
\[ [TS]^{i_2,j_2} f(x_1, x_2) := \left( \sum_{R \in D_2} \left( \sum_{P \in (R_2)_2} \frac{T(H_{P_2}^i f)(x_1)}{|Q_2|} \right)^2 \right)^{1/2}. \]

Under the same assumptions on \( T \), it is easy to see that
\[ \| [ST]^{i_1,j_1} : L^p(w) \rightarrow L^p(w) \| \lesssim 2^{\frac{1}{p}(1 + j_1)} \quad \text{and} \quad \| [TS]^{i_2,j_2} : L^p(w) \rightarrow L^p(w) \| \lesssim 2^{\frac{1}{p}(1 + j_2)}, \]
for all \( w \in A_p(\mathbb{R}^n) \). Specifically,
\[ \| [ST]^{i_1,j_1} f \|_{L^p(w)} = \int \left| S_{D_1}^{i_1,j_1} F(x_1, x_2) \right|^p \, dw, \]
where \( F(x_1, x_2) := \sum_{P_1 \in D_1} T(H_{P_1}^i f)(x_2) h_{P_1}^j(x_1) \), so
\[ \| [ST]^{i_1,j_1} f \|_{L^p(w)} \lesssim 2^{\frac{1}{p}(1 + j_1)} \| F \|_{L^p(w)}. \]
Now, \( \| F \|_{L^p(w)} \approx \| S_{D_1} F \|_{L^p(w)} = \| [ST] f \|_{L^p(w)} \lesssim \| f \|_{L^p(w)}. \)

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 \( BMO(w) \) space if
\[ \| b \|_{BMO(w)} := \sup_{Q} \frac{1}{w(Q)} \int_Q |b(x) - \langle b \rangle_Q| \, dx < \infty, \]
where the supremum is over all cubes \( Q \) in \( \mathbb{R}^n \). If \( w = 1 \), we obtain the unweighted \( BMO(\mathbb{R}^n) \) space. The dyadic version \( BMO_D(w) \) is obtained by only taking supremum over \( Q \in D \) for some given dyadic grid \( D \) on \( \mathbb{R}^n \). If the weight \( w \in A_p(\mathbb{R}^n) \) for some \( 1 < p < \infty \), Muckenhoupt and Wheeden show in [16] that
\[ \| b \|_{BMO(w)} \approx \| b \|_{BMO_D(w)} = \sup_{Q} \left( \frac{1}{w(Q)} \int_Q |b - \langle b \rangle_Q|^p \, dw \right)^{1/p}, \]
where \( w' \) is the conjugate weight to \( w \). Moreover, if \( w \in A_2(\mathbb{R}^n) \), Wu’s argument in [21] shows that \( BMO_D(w) \approx H_D^1(w)^* \), where the dyadic Hardy space \( H_D^1(w) \) is defined by the norm
\[ \| \phi \|_{H_D^1(w)} := \| S_D \phi \|_{L^1(w)}. \]
Then
\[ | \langle b, \phi \rangle | \lesssim \| b \|_{BMO_D(w)} \| S_D \phi \|_{L^1(w)}, \quad \text{for all} \ w \in A_2(\mathbb{R}^n). \]

Now suppose \( \mu \) and \( \lambda \) 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 [11], the weight \( \nu \in A_2(\mathbb{R}^n) \), which means we may use (4.2) with \( \nu \). A two-weight John-Nirenberg theorem for the Bloom BMO space \( BMO(\nu) \) is also proved in [11], namely
\[ \| b \|_{BMO(\nu)} \approx \| b \|_{BMO(\mu, \lambda, \nu)} \approx \| b \|_{BMO(\nu', \mu', \nu')}, \]
where
\[ \| b \|_{BMO(\mu, \lambda, \nu)} := \sup_{Q} \left( \frac{1}{\mu(Q)} \int_Q |b - \langle b \rangle_Q|^p \lambda \right)^{1/p}, \]
\[ \| b \|_{BMO(\nu', \mu', \nu')} := \sup_{Q} \left( \frac{1}{\nu'(Q)} \int_Q |b - \langle b \rangle_Q|^p \mu' \right)^{1/p}. \]

We now look at weighted BMO spaces in the product setting \( \mathbb{R}^d = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \). Suppose \( w(x_1, x_2) \) is a weight on \( \mathbb{R}^n \). Then we have three BMO spaces:
- **Weighted Little \( bmo(w) \)**: is the space of all locally integrable functions \( b \) on \( \mathbb{R}^d \) such that
  \[ \| b \|_{bmo(w)} := \frac{1}{w(R)} \int_R (|b - \langle b \rangle_R|)^p \, dx < \infty, \]
where the supremum is over all rectangles \( R = Q_1 \times Q_2 \) in \( \mathbb{R}^d \). Given a choice of dyadic rectangles \( D = D_1 \times D_2 \), we define the dyadic weighted little \( bmo_D(w) \) by taking supremum over \( R \in D \).
• Weighted Product $BMO_{\mathcal{D}(w)}$: is the space of all locally integrable functions $b$ on $\mathbb{R}^n$ such that:

$$\|b\|_{BMO_{\mathcal{D}(w)}} := \sup_{\Omega} \left( \frac{1}{w(\Omega)} \sum_{R \subset \Omega \cap \mathcal{D}} |\hat{b}(R)|^2 \frac{1}{\langle w \rangle_R} \right)^{1/2},$$

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

• Weighted Rectangular $BMO_{\mathcal{D},Rec(w)}$: is defined in a similar fashion to the unweighted case – just like product BMO, but taking supremum over rectangles instead of over open sets:

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

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

We have the inclusions

$$bmo_{\mathcal{D}(w)} \subseteq BMO_{\mathcal{D}(w)} \subseteq BMO_{\mathcal{D},Rec(w)}.$$  

Let us look more closely at some of these spaces.

### 4.1 Weighted Product $BMO_{\mathcal{D}(w)}$

As in the one parameter case, we define the dyadic weighted Hardy space $H^1_{\mathcal{D}(w)}$ to be the space of all $\phi \in L^1(\mathcal{D})$ such that $S_{\mathcal{D}} \phi \in L^1(\mathcal{D})$, a Banach space under the norm

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

The following result exists in the literature under various forms, but we include a proof here for completeness.

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

$$(3.3) \quad |\langle b, \phi \rangle| \lesssim \|b\|_{BMO_{\mathcal{D}(w)}} \|S_{\mathcal{D}} \phi\|_{L^1(\mathcal{D})},$$

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

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

$$(4.4) \quad \sum_{j \in \mathbb{Z}} 2^j w(U_j) \lesssim \|S_{\mathcal{D}} \phi\|_{L^1(\mathcal{D})},$$

which comes from the measure theoretical fact that for any integrable function $f$ on a measure space $(X, \mu)$:

$$\|f\|_{L^1(\mu)} \approx \sum_{j \in \mathbb{Z}} 2^j \mu \{x \in X : |f(x)| > 2^j\}.$$  

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

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

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

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

so $\theta < \langle 1_{U_j} \rangle_R \leq M_S 1_{U_j}(x)$, for all $x \in R$.

Therefore

$$\bigcup_{R \in \mathcal{R}_j} R \subset V_j.$$  

Using (2.7), we have that

$$w(V_j) \lesssim \int_{U_j} \frac{1}{\theta^p} \left( 1 + \log^+ \frac{1}{\theta} \right)^{k-1} \, dw \simeq w(U_j).$$
Now suppose $R \in \mathcal{D}$ but $R \not\subset \bigcup_{j \in \mathbb{Z}} R_j$. Then $w(R \cap \{S_D \phi \leq 2^j\}) \geq \frac{1}{2}w(R)$ for all $j \in \mathbb{Z}$, and so
\[
w(R \cap \{S_D \phi = 0\}) = w\left( \bigcap_{j=1}^{\infty} R \cap \{S_D \phi \leq 2^{-j}\} \right) \geq \frac{1}{2}w(R).
\]
Then $|\{S_D \phi = 0\}| \geq |R \cap \{S_D \phi = 0\}| \geq \theta |R| > 0$, and we may write
\[
|\hat{\phi}(R)|^2 = \int_{\{S_D \phi = 0\}} |\hat{\phi}(R)|^2 \frac{1}{|R \cap \{S_D \phi = 0\}|} \, dx \leq \frac{1}{\theta} \int_{\{S_D \phi = 0\}} (S_D \phi)^2 \, dx = 0.
\]
So
\[
(4.7) \quad \hat{\phi}(R) = 0, \text{ for all } R \in \mathcal{D}, R \not\subset \bigcup_{j \in \mathbb{Z}} R_j.
\]
Finally, if $R \in \bigcap_{j \in \mathbb{Z}} R_j$, then
\[
0 = w(R \cap \{S_D \phi = \infty\}) = \lim_{j \to \infty} w(R \cap \{S_D \phi > 2^j\}) \geq \frac{1}{2}w(R),
\]
a contradiction. In light of this and (4.7),
\[
\sum_{R \in \mathcal{D}} |\hat{b}(R)||\hat{\phi}(R)| = \sum_{j \in \mathbb{Z}} \sum_{R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}} |\hat{b}(R)||\hat{\phi}(R)|
\]
\[
\leq \sum_{j \in \mathbb{Z}} \left( \sum_{R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}} |\hat{b}(R)|^2 \frac{1}{\langle w \rangle_R} \right)^{1/2} \left( \sum_{R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}} |\hat{\phi}(R)|^2 \langle w \rangle_R \right)^{1/2}
\]
To estimate the first term, we simply note that
\[
\sum_{R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}} |\hat{b}(R)|^2 \frac{1}{\langle w \rangle_R} \leq \sum_{R \in \mathcal{R}_j} |\hat{b}(R)|^2 \frac{1}{\langle w \rangle_R} \leq \sum_{R \subset V_j, R \in \mathcal{D}} \hat{b}(R)|^2 \frac{1}{\langle w \rangle_R} \leq \|b\|_{BMO_D(w)}^2 w(V_j),
\]
where the second inequality follows from (4.5). For the second term, remark that any $R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}$ satisfies $R \subset V_j$ and $w(R \setminus U_{j+1}) \geq \frac{1}{2}w(R)$. Then
\[
\sum_{R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}} |\hat{\phi}(R)|^2 \langle w \rangle_R \leq \sum_{R \in \mathcal{R}_j \setminus \mathcal{R}_{j+1}} w(R \setminus U_{j+1}) \cdot \frac{1}{|R|} \leq 2 \int_{V_j \setminus U_{j+1}} \hat{\phi}(R)|^2 \frac{1}{|R|} \, dw
\]
\[
\leq 2 \int_{V_j \setminus U_{j+1}} (S_D \phi)^2 \, dw \lesssim 2^{2j}w(V_j),
\]
since $S_D \phi \leq 2^{j+1}$ off $U_{j+1}$. Finally, we have by (4.6):
\[
\sum_{R \in \mathcal{D}} |\hat{b}(R)||\hat{\phi}(R)| \lesssim \|b\|_{BMO_D(w)} \sum_{j \in \mathbb{Z}} 2^{j}w(V_j) \simeq \|b\|_{BMO_D(w)} \sum_{j \in \mathbb{Z}} 2^{j}w(U_j).
\]
Combining this with (4.4), we obtain (4.3).
To see the converse, let $L \in \mathcal{H}_D^*(w)$. Then $L$ is given by $L \phi = \langle b, \phi \rangle$ for some function $b$. Fix an open set $\Omega$ with $w(\Omega) < \infty$. Then
\[
\left( \sum_{R \subset \Omega, R \in \mathcal{D}} \left| \frac{\hat{b}(R)}{\langle w \rangle_R} \right|^2 \right)^{1/2} \leq \sup_{\|\phi\|_{L^2(\omega)}} \sum_{R \subset \Omega, R \in \mathcal{D}} \left( \frac{\hat{b}(R)}{\langle w \rangle_R} \right)|\hat{\phi}(R)|,
\]
where $\|\phi\|_{L^2(\omega)}^2 := \sum_{R \subset \Omega, R \in \mathcal{D}} |\hat{\phi}(R)|^2 \langle w \rangle_R$. By a simple application of Hölder’s inequality,
\[
\left( \sum_{R \subset \Omega, R \in \mathcal{D}} \left| \frac{\hat{b}(R)}{\langle w \rangle_R} \right|^2 \right)^{1/2} \lesssim \|L\| \cdot \|\phi\|_{H_D(w)} \leq \|L\| \cdot w(\Omega)^{1/2} \cdot \|\phi\|_{L^2(\omega)}^2,
\]
so \( \|b\|_{BMO_{\mathcal{D}}(w)} \lesssim \|L\|_w. \)

4.2. Weighted little \( bmo_{\mathcal{D}}(w) \). In this case, we also want to look at each variable separately. Specifically, we look at the space \( BMO(w_1, x_2) \): 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 \( w(\cdot, x_2) \).

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

The norm in this space is given by

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

where

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

The space \( BMO(w_2, x_1) \) 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 \( \mathbb{R}^n = \mathbb{R}^n_1 \otimes \mathbb{R}^n_2 \). Then \( b \in L^1_{\text{loc}}(\mathbb{R}^n) \) is in \( bmo(w) \) if and only if \( b \) is in the one-parameter weighted BMO spaces \( BMO(w_1, x_2) \) separately in each variable, uniformly:

\[ \|b\|_{bmo(w)} \lesssim \max \left\{ \text{ess sup}_{x_1 \in \mathbb{R}^n_1} \|b(x_1, \cdot)\|_{BMO(w_2, x_1)}; \text{ess sup}_{x_2 \in \mathbb{R}^n_2} \|b(\cdot, x_2)\|_{BMO(w_1, x_2)} \right\}. \]

**Remark 4.3.** In the unweighted case \( bmo(\mathbb{R}^n) \), if we fixed \( x_2 \in \mathbb{R}^n_2 \), we would look at \( b(\cdot, x_2) \) in the space \( BMO(\mathbb{R}^n_1) \) – 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 \( b \in bmo(w) \). Then for all cubes \( Q_1, Q_2 \):

\[ \|b\|_{bmo(w)} \geq \frac{1}{w(Q_1 \times Q_2)} \int_{Q_1} \int_{Q_2} |b(x_1, x_2) - (b)_{Q_1 \times Q_2}| \, dx_1 \, dx_2 \]

\[ \geq \frac{1}{w(Q_1 \times Q_2)} \int_{Q_1} \int_{Q_2} |b(x_1, x_2) - (b)_{Q_1 \times Q_2}| \, dx_2 \, dx_1, \]

so

\[ \int_{Q_1} |m_{Q_1} b(x_1) - (b)_{Q_1 \times Q_2}| \, dx_1 \leq \frac{w(Q_1 \times Q_2)}{|Q_2|} \|b\|_{bmo(w)}. \]

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 \( Q_1 \):

\[ \langle f_{Q_2} \rangle_{Q_1} \leq \frac{1}{|Q_1|} \int_{Q_1} \int_{Q_2} |b(x_1, x_2) - (b)_{Q_1 \times Q_2}| \, dx_1 \, dx + \frac{1}{|Q_1|} \int_{Q_1} \int_{Q_2} |m_{Q_2} b(x_1) - (b)_{Q_1 \times Q_2}| \, dx_1 \]

\[ \leq \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{bmo(w)} + \frac{|Q_2|}{|Q_1|} \int_{Q_1} |m_{Q_2} b(x_1) - (b)_{Q_1 \times Q_2}| \, dx_1 \]

\[ \leq 2 \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{bmo(w)} = 2 \langle w(Q_2, \cdot) \rangle_{Q_1} \|b\|_{bmo(w)}, \]

where the last inequality follows from (4.8). By the Lebesgue differentiation theorem:

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

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 \( \|b(x_1, \cdot)\|_{BMO(w_2, x_1)} = \sup_{Q_2} \frac{1}{w(Q_2, x_1)} f_{Q_2}(x_1) \) 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

\[ f_{Q_2}(x_1) \leq 2 \|b\|_{bmo(w)} w(Q_2, x_1) \text{ for all } x_1 \in \mathbb{R}^n_1 \setminus N(Q_2). \]
In order to obtain the inequality we want, holding for a.a. \(x_1\), let \(N := \cup N(Q_2)\) where \(Q_2\) are the cubes in \(\mathbb{R}^{n_2}\) with rational side length and centers with rational coordinates. Then \(N\) is a null set and \(f_{Q_2}(x_1) \leq 2\|b\|_{bmo(w)} w(\widetilde{Q}_2, x_1)\) 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
\[
\text{ess sup}_{x_1 \in \mathbb{R}^{n_1}} \|b(x_1, \cdot)\|_{BMO(w_2, x_1)} \leq 2\|b\|_{bmo(w)}.
\]
The result for the other variable follows symmetrically.

Conversely, suppose
\[
\|b(x_1, \cdot)\|_{BMO(w_2, x_1)} \leq C_1 \text{ for a.a. } x_1, \text{ and } \|b(\cdot, x_2)\|_{BMO(w_1, x_2)} \leq C_2 \text{ for a.a. } x_2.
\]
Then for any \(R = Q_1 \times Q_2\):
\[
\int_R \|b - \langle b \rangle_R\| dx \leq \int_{Q_1} \int_{Q_2} \|b(x_1, x_2) - m_{Q_2}(x_1)\| dx + \int_{Q_1} \|m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2}\|_R dx_1
\]
\[
\leq \int_{Q_1} C_2 w(Q_2, x_1) dx_1 + \int_{Q_1} \int_{Q_2} \|b(x_1, x_2) - m_{Q_2} b(x_2)\| dx_2 dx_1
\]
\[
\leq C_2 w(R) + \int_{Q_2} C_1 w(Q_1, x_2) dx_2
\]
\[
= (C_1 + C_2) w(R),
\]
so
\[
\|b\|_{bmo(w)} \leq 2 \max \left\{ \text{ess sup}_{x_1 \in \mathbb{R}^{n_1}} \|b(x_1, \cdot)\|_{BMO(w_2, x_1)}; \text{ess sup}_{x_2 \in \mathbb{R}^{n_2}} \|b(\cdot, x_2)\|_{BMO(w_1, x_2)} \right\}.
\]

\[\square\]

**Corollary 4.4.** Let \(w \in A_2(\mathbb{R}^n)\) and \(b \in bmoP(w)\). Then
\[
|\langle b, \phi \rangle| \lesssim \|b\|_{bmoP(w)} \|S_{\mathcal{D}_i} \phi\|_{L^1(w)},
\]
for all \(i \in \{1, 2\}\).

**Proof.** This follows immediately from the one-parameter result in (4.2) and the proposition above:
\[
|\langle b, \phi \rangle| \leq \int_{\mathbb{R}^{n_2}} |\langle b(x_1, \cdot), \phi(x_1, \cdot) \rangle|_{\mathbb{R}^{n_2}} dx_1
\]
\[
\lesssim \int_{\mathbb{R}^{n_2}} \|b(x_1, \cdot)\|_{BMO_{A_2(w(x_1), j)}} \|S_{\mathcal{D}_2} \phi(x_1, \cdot)\|_{L^1(w(x_1), j)} dx_1
\]
\[
\lesssim \|b\|_{bmo(w)} \|S_{\mathcal{D}_2} \phi\|_{L^1(w)},
\]
and similarly for \(S_{\mathcal{D}_1}\). \[\square\]

We now look at the little bmo version of (4.1).

**Proposition 4.5.** If \(w \in A_p(\mathbb{R}^n)\) for some \(1 < p < \infty\), then
\[
\|b\|_{bmo(w)} \simeq \|b\|_{bmo(w; p')} := \sup_R \left( \frac{1}{w(R)} \int_R |b - \langle b \rangle_R|^{p'} \, dw^{p'} \right)^{1/p'}.
\]

**Proof.** By Proposition 4.2 and (4.1):
\[
\|b\|_{bmo(w)} \simeq \max \left\{ \text{ess sup}_{x_1 \in \mathbb{R}^{n_1}} \|b(x_1, \cdot)\|_{BMO(w(x_1), \cdot; p')}; \text{ess sup}_{x_2 \in \mathbb{R}^{n_2}} \|b(\cdot, x_2)\|_{BMO(w(\cdot, x_2); p')} \right\}.
\]
Suppose first that \(b \in bmo(w; p')\). Note that for some function \(g\) on \(\mathbb{R}^n\) and a cube \(Q_2\) in \(\mathbb{R}^{n_2}\), we have
\[
\int_{Q_2} |g(x_1, x_2)|^{p'} w^{p'}(x_1, x_2) dx_2 \geq \frac{1}{w(Q_2, x_1)^{p'-1}} \int_{Q_2} g(x_1, x_2) dx_2^{p'}.
\]
Then
\[
\|b\|_{\text{bmo}(w;p')}^{p'} \geq \frac{1}{w(R)} \left( \frac{1}{w(Q_2, x_1)^{p'-1}} \right) \left( \int_{Q_2} b(x_1, x_2) - \langle b \rangle_{Q_1 \times Q_2} dx_2 \right)^{p'} dx_1
\]
\[
= \frac{1}{w(R)} \int_{Q_1} m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2} \left( \frac{|Q_2|^{p'}}{w(Q_2, x_1)^{p'-1}} \right) dx_1
\]
\[
\geq \frac{1}{w(R)} \int_{Q_1} m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2} \left| w'(Q_2, x_1) dx_1 \right|
\]
where the last inequality follows from
\[
\frac{|Q_2|^{p'}}{w(Q_2, x_1)^{p'-1}} = |Q_2| \frac{1}{w(x_1, \cdot)^{p'-1}} \geq |Q_2| \frac{\langle w'(x_1, \cdot) \rangle_{Q_2}}{w(x_1, \cdot)^{p'-1}} \simeq w'(Q_2, x_1).
\]

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
\[
\langle f_{Q_2} \rangle_{Q_1} \lesssim \frac{1}{|Q_1|} \int_{Q_1} \int_{Q_2} \left( |b(x_1, x_2) - \langle b \rangle_{Q_1 \times Q_2}|^{p'} + |m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2}|^{p'} \right) w'(x_1, x_2) dx_2 dx_1
\]
\[
\lesssim \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{\text{bmo}(w;p')}^{p'} + \frac{1}{|Q_1|} \int_{Q_1} m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2} \left| w'(Q_2, x_1) dx_1 \right|
\]
\[
\lesssim \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{\text{bmo}(w;p')}^{p'}.
\]

Then for almost all $x_1$:
\[
f_{Q_2}(x_1) = \lim_{Q_1 \rightarrow x_1} \langle f_{Q_2} \rangle_{Q_1} \lesssim \lim_{Q_1 \rightarrow x_1} \frac{w(Q_1 \times Q_2)}{|Q_1|} \|b\|_{\text{bmo}(w;p')}^{p'} = w(Q_2, x_1) \|b\|_{\text{bmo}(w;p')}^{p'}.
\]

Taking again rational cubes, we obtain
\[
\|b(x_1, \cdot)\|_{\text{BMO}(w(x_1, \cdot);p')} = \sup_{Q_2} \left( \frac{1}{w(Q_2, x_1)} f_{Q_2}(x_1) \right)^{1/p'} \lesssim \|b\|_{\text{bmo}(w;p')},
\]
for almost all $x_1$.

Conversely, if $b \in \text{bmo}(w)$, then there exist $C_1$ and $C_2$ such that
\[
\|b(x_1, \cdot)\|_{\text{BMO}(w(x_1, \cdot);p')} \leq C_1 \text{ a.a. } x_1, \text{ and } \|b(\cdot, x_2)\|_{\text{BMO}(w(\cdot, x_2);p')} \leq C_2 \text{ a.a. } x_2.
\]

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

The first integral is easily seen to be bounded by
\[
\int_{Q_1} \|b(x_1, \cdot)\|_{\text{BMO}(w(x_1, \cdot);p')}^{p'} w(Q_2, x_1) dx_1 \leq C'_1 w(Q_1 \times Q_2).
\]

The second integral is equal to:
\[
\int_{Q_1} \left| m_{Q_2} b(x_1) - \langle b \rangle_{Q_1 \times Q_2} \right|^{p'} w'(Q_2, x_1) dx_1
\]
\[
\leq \int_{Q_1} \frac{w'(Q_2, x_1)}{|Q_2|^{p'}} \left( \int_{Q_2} \left| b(x_1, x_2) - m_{Q_2} b(x_2) \right| dx_2 \right)^{p'} dx_1
\]
\[
\leq \int_{Q_1} \frac{w'(Q_2, x_1) w(Q_2, x_1)^{p'-1}}{|Q_2|^{p'}} \int_{Q_2} \left| b(x_1, x_2) - m_{Q_2} b(x_2) \right|^{p'} w'(x_1, x_2) dx_2 dx_1.
\]
We may express the first term as \( \langle w'(x_1, \cdot) \rangle_{Q_2} (w(x_1, \cdot))_{Q_2}^{p'-1} \lesssim |w|_{A_p}^{p'-1} \) for almost all \( x_1 \). Then, the integral is further bounded by

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

Finally, this gives

\[
\| b \|_{bmo(w; p')} \lesssim (C_{Q_1}^p + C_{Q_2}^p)^{1/p'} \lesssim \max(C_1, C_2) \simeq \| b \|_{bmo(w)}.
\]

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

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

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

where

\[
\| b \|_{bmo(\mu, \lambda, p)} := \sup_{R} \left( \frac{1}{\mu(R)} \int_{R} |b - \langle b \rangle_R|^p \, d\lambda \right)^{1/p},
\]

\[
\| b \|_{bmo(\lambda', \mu', p')} := \sup_{R} \left( \frac{1}{\lambda'(R)} \int_{R} |b - \langle b \rangle_R|^p \, d\mu' \right)^{1/p'}.
\]

Remark that it also easily follows that \( \nu \in A_2(\mathbb{R}^d) \).

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 [3]. Let \( \{X_k(x)\} \) and \( \{Y_l(y)\} \) both be orthonormal bases for the space of spherical harmonics of degree \( n \) in \( \mathbb{R}^n \) respectively. Then \( \sum_k |X_k(x)|^2 = c_n |x|^{2n} \) and thus

\[
1 = \frac{1}{c_n} \sum_k X_k(x - x') X_k(x - x')
\]

and similarly for \( Y_l \).

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

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

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

\[
\Gamma_Q(x, y) = \text{sign}(b(x, y) - \langle b \rangle_Q) \mathbf{1}_Q(x, y).
\]

So

\[
\frac{|b(x, y) - \langle b \rangle_Q|}{|Q|} \Gamma_Q(x, y) = \frac{b(x, y) - \langle b \rangle_Q}{|Q|} \Gamma_Q(x, y)
\]

\[
= \int_Q (b(x, y) - b(x', y')) \Gamma_Q(x, y) \, dx' dy'
\]

\[
\lesssim \sum_{k,l} \int_Q (b(x, y) - b(x', y')) \frac{X_k(x - x')}{|x - x'|^{2n}} X_k(x - x') \frac{Y_l(y - y')}{|y - y'|^{2n}} Y_l(y - y') \Gamma_Q(x, y) \, dx' dy'
\]

\[
= \sum_{k,l} \int_{\mathbb{R}^{2n}} \frac{b(x, y) - b(x', y')}{|x - x'|^{2n}} \frac{X_k(x - x')}{|y - y'|^{2n}} X_k(x - x') \frac{Y_l(y - y')}{|y - y'|^{2n}} Y_l(y - y') \Gamma_Q(x, y) \, dx' dy'
\]

\[
\cdot \sum_{|\alpha| + |\beta| = n} \sum_{|\gamma| + |\delta| = n} x^{(k)}_{\alpha \beta} x^\alpha x^\beta \cdot \sum_{|\gamma| + |\delta| = n} y^{(l)}_{\gamma \delta} y^\gamma y^\delta \Gamma_Q(x, y) \mathbf{1}_Q(x', y') \, dx' dy'.
\]
Note that
\[
\int_{\mathbb{R}^{2n}} \frac{b(x, y) - b(x', y')}{|x - x'|^{2n}|y - y'|^{2n}} X_k(x - x') Y_l(y - y') \cdot x^{\alpha} y^{\beta} y^{i_1}_Q(x', y') \, dx \, dy' = [b, T_k T_l] (x^{\alpha} y^{i_1} y^{i_2}_Q(x', y')).
\]

Here \(T_k\) and \(T_l\) are the Calderón-Zygmund operators that correspond to the kernels
\[
\frac{X_k(x)}{|x|^{2n}} \text{ and } \frac{Y_l(y)}{|y|^{2n}}.
\]

Observe that these have the correct homogeneity due to the homogeneity of the \(X_k\) and \(Y_l\). With this notation, the above becomes
\[
\frac{|b(x, y) - \langle b \rangle_Q| |1_Q(x, y)|}{|b(\langle b \rangle_Q Q)|} \leq \sum_{k, l} \sum_{|\alpha| + |\beta| = n} \sum_{|\gamma| + |\delta| = n} x^{(k)}_{\alpha \beta} x^{(l)}_{\gamma \delta} y^{i_1}_Q(x, y) [b, T_k T_l] (x^{\alpha} y^{\beta} y^{i_2}_Q(x', y')) (x, y).
\]

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

We disregarded the coefficients of the \(X\) 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 \([b, T_k T_l]\) writes as a linear combination of terms such as \(M[b, R^1_{\alpha} R^2_{\beta}]\) where \(M\) and \(N\) are compositions of Riesz transforms: in a first step write \([b, T_k T_l]\) as a linear combination of terms of the form \([b, R^1_{(n)} R^1_{(n)}]\) where

\[
R^1_{(n)} = \prod_s R^1_{i_s(k)}
\]

is a composition of \(n\) Riesz transforms acting in the variabe 1 with a choice \(i(k) = (i_s^{(k)})_{s=1}^n \in \{1, \ldots, n\}^n\) for each \(k\) and similar for \(R^1_{(n)}\) acting in variable 2. 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 \(R^1_{(n-1)} R^1_{(n-1)}\) and repeat. It then follows that for each \(k, l\) the commutator \([b, T_k T_l]\) writes as a linear combination of terms such as \(M[b, R^1_{\alpha} R^2_{\beta}]\) where \(M\) and \(N\) are compositions of Riesz transforms. It is decisive that \(T_k\) and \(T_l\) are homogeneous polynomials in Riesz transforms of the same degree. We required that all commutators of the form \([b, R^1_{\alpha} R^2_{\beta}]\) are bounded, 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:

\[
\|b\|_{bmo(\nu)} \simeq \|b\|_{bmo(\mu, \lambda, p)} := \sup_R \left( \frac{1}{\mu(R)} \int_R |b - \langle b \rangle_R|^p \, d\lambda \right)^{1/p} \lesssim \sup_{1 \leq k, l \leq n} \| [b, R^1_{\alpha} R^2_{\beta}] \|_{L^p(\mu) \to L^p(\lambda)}.
\]
6. Biparameter Paraproducts

Decomposing two functions \( b \) and \( f \) on \( \mathbb{R}^d \) 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

\[
\Pi_b f := \sum_{Q \in \mathcal{D}} \hat{b}(Q) \langle f \rangle_Q h_{Q}^s, \quad \Pi_b^* f := \sum_{Q \in \mathcal{D}} \hat{b}(Q) \langle f \rangle_Q \frac{1}{|Q|} h_{Q}^s, \quad \text{and} \quad \Gamma_b f := \sum_{Q \in \mathcal{D}} \hat{b}(Q) \langle f \rangle_Q \frac{1}{\sqrt{|Q|}} h_{Q}^{s+\delta}.
\]

In [11], it was shown that, when \( b \in BMO(\nu) \), the operators \( \Pi_b, \Pi_b^* \), and \( \Gamma_b \) are bounded \( L^p(\mu) \to L^p(\lambda) \).

6.1. Product BMO Paraproducts. In the biparameter setting \( \mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \), 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:

\[
\Pi_{b(0,1)} f := \sum_{Q_1 \times Q_2} \hat{b}(Q_1^s \times Q_2^s) \langle f \rangle_{Q_1 \times Q_2} h_{Q_1}^{s} \otimes h_{Q_2}^{s},
\]

\[
\Pi_{b(1,0)} f := \sum_{Q_1 \times Q_2} \hat{b}(Q_1^s \times Q_2^s) \langle f \rangle_{Q_1 \times Q_2} \frac{1}{|Q_1|} \otimes \frac{1}{|Q_2|} h_{Q_1}^{s} \otimes h_{Q_2}^{s},
\]

\[
\Gamma_{b(0,1)} f := \sum_{Q_1 \times Q_2} \hat{b}(Q_1^s \times Q_2^s) \langle f \rangle_{Q_1 \times Q_2} \frac{1}{\sqrt{|Q_1|}} \otimes \frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{s+\delta} \otimes h_{Q_2}^{s},
\]

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 \( \text{BMO}_\mathcal{D} \): in a proof using duality, one would separate out the \( b \) function and be left with the biparameter square function \( S_\mathcal{D} \). They are:

\[
\Gamma_{b(1,0)} f := \sum_{Q_1 \times Q_2} \hat{b}(Q_1^s \times Q_2^s) \langle f \rangle_{Q_1 \times Q_2} \frac{1}{\sqrt{|Q_1|}} \otimes \frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{s} \otimes h_{Q_2}^{s+\delta},
\]

\[
\Gamma_{b(1,0)} f := \sum_{Q_1 \times Q_2} \hat{b}(Q_1^s \times Q_2^s) \langle f \rangle_{Q_1 \times Q_2} \frac{1}{|Q_1|} \otimes \frac{1}{|Q_2|} h_{Q_1}^{s} \otimes h_{Q_2}^{s+\delta},
\]

\[
\Gamma_{b(1,0)} f := \sum_{Q_1 \times Q_2} \hat{b}(Q_1^s \times Q_2^s) \langle f \rangle_{Q_1 \times Q_2} \frac{1}{|Q_2|} h_{Q_1}^{s} \otimes h_{Q_2}^{s+\delta}.
\]

Proposition 6.1. If \( \nu := \mu^{1/p} \lambda^{1-1/p} \) for \( A_p(\mathbb{R}^d) \) weights \( \mu \) and \( \lambda \), and \( \mathcal{P}_b \) denotes any one of the nine paraproducts defined above, then

\[
\| \mathcal{P}_b : L^p(\mu) \to L^p(\lambda) \| \lesssim \| b \|_{\text{BMO}_\mathcal{D}(\nu)},
\]

where \( \| b \|_{\text{BMO}_\mathcal{D}(\nu)} \) denotes the norm of \( b \) in the dyadic weighted product \( \text{BMO}_\mathcal{D}(\nu) \) space on \( \mathbb{R}^d \).

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:

\[
\| (\mathcal{P}_b f, g) \| \lesssim \| b \|_{\text{BMO}_\mathcal{D}(\nu)} \| f \|_{L^p(\mu)} \| g \|_{L^p(\lambda)}.
\]

1. Write \( (\mathcal{P}_b f, g) = \langle b, \phi \rangle \), where \( \phi \) depends on \( f \) and \( g \). By (4.3), \( \| (\mathcal{P}_b f, g) \| \lesssim \| b \|_{\text{BMO}_\mathcal{D}(\nu)} \| S_\mathcal{D} \phi \|_{L^p(\nu)} \).

2. Show that \( S_\mathcal{D} \phi \lesssim (O_1 f)(O_2 g) \), where \( O_1 \) and \( O_2 \) are operators satisfying a one-weight bound \( L^p(w) \to L^p(w) \), for all \( w \in A_p(\mathbb{R}^d) \) – these operators will usually be a combination of maximal and square functions.
3. Then the $L^1(\nu)$-norm of $S_D\phi$ can be separated into the $L^p(\mu)$ and $L^{p'}(\lambda')$ norms of these operators $O_i$, by a simple application of Hölder’s inequality:

$$\|S_D\phi\|_{L^1(\nu)} \lesssim \|O_1 f\|_{L^p(\mu)} \|O_2 g\|_{L^{p'}(\lambda')} \lesssim \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')} ,$$

and the result follows.

Remark also that we will not have to treat the adjoints $P_b^\ast$ 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 $P_b^\ast$ is bounded $L^p(\mu) \to L^p(\lambda)$.

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

$$\langle \Pi_b f, g \rangle = \langle b, \phi \rangle , \text{ where } \phi := \sum_{Q_1 \times Q_2} \langle f \rangle_{Q_1 \times Q_2} \hat{g}(Q_1^1 \times Q_2^2)h_{Q_1}^c \otimes h_{Q_2}^c .$$

Then

$$(S_D\phi)^2 \leq \sum_{Q_1 \times Q_2} \langle f \rangle_{Q_1 \times Q_2}^2 \hat{g}(Q_1^1 \times Q_2^2)^2 \frac{1}{|Q_1|} \otimes \frac{1}{|Q_2|} \leq (Msf)^2 \cdot (S_Dg)^2 ,$$

so

$$| \langle \Pi_b f, g \rangle | \lesssim \|b\|_{BMO_D(\nu)} \|Msf\|_{L^p(\mu)} \|S_Dg\|_{L^{p'}(\lambda')} \lesssim \|b\|_{BMO_D(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')} .$$

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

$$| \langle \Pi_b f, g \rangle | \lesssim \|b\|_{BMO_D(\nu)} \|Msf\|_{L^{p'}(\lambda')} \|S_Dg\|_{L^p(\mu)} \lesssim \|b\|_{BMO_D(\nu)} \|f\|_{L^{p'}(\lambda')} \|g\|_{L^p(\mu)} ,$$

proving that $\|\Pi_b : L^{p'}(\lambda') \to L^p(\mu')\| = \|\Pi_b^\ast : L^p(\mu) \to L^{p'}(\lambda)\| \lesssim \|b\|_{BMO_D(\nu)}$. For $\Gamma_b$:

$$\langle \Gamma_b f, g \rangle = \langle b, \phi \rangle , \text{ where } \phi := \sum_{Q_1 \times Q_2} \hat{f}(Q_1^1 \times Q_2^2)\hat{g}(Q_1^1 \times Q_2^2) \frac{1}{\sqrt{|Q_1|}} \otimes \frac{1}{\sqrt{|Q_2|}} h_{Q_1}^{\epsilon_1 + \delta_1} \otimes h_{Q_2}^{\epsilon_2 + \delta_2} ,$$

from which it easily follows that $S_D\phi \lesssim S_D f \cdot S_Dg$.

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

$$\phi := \sum_{Q_1 \times Q_2} \langle f, h_{Q_1}^c \otimes \frac{1}{|Q_2|} \rangle \langle g, \frac{1}{|Q_1|} \otimes h_{Q_2}^c \rangle h_{Q_1}^c \otimes h_{Q_2}^c .$$

Then

$$(S_D\phi)^2 = \sum_{Q_1 \times Q_2} \langle f, h_{Q_1}^c \otimes \frac{1}{|Q_2|} \rangle^2 \langle g, \frac{1}{|Q_1|} \otimes h_{Q_2}^c \rangle^2 \frac{1}{|Q_1|} \otimes \frac{1}{|Q_2|} \lesssim \langle f, H_{Q_1}^c f \rangle \langle g, H_{Q_2}^c g \rangle \frac{1}{|Q_1|} \otimes \frac{1}{|Q_2|} \lesssim \left( \sum_{Q_1} (M_{D_2} H_{Q_1}^c f)^2(x_2) \frac{\|Q_1\|}{|Q_1|} \right) \left( \sum_{Q_2} (M_{D_2} H_{Q_2}^c g)^2(x_1) \frac{\|Q_2\|}{|Q_2|} \right) \lesssim (SM)^2 f \cdot [MS]^2 g ,$$

where $[SM]$ and $[MS]$ are the mixed square-maximal operators in Section 3.1. 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 $\Gamma_{b(0,1)}$:

$$\phi = \sum_{Q_1 \times Q_2} \langle H_{Q_1}^c f \rangle_{Q_2} \frac{1}{\sqrt{|Q_1|}} \hat{g}(Q_1^1 \times Q_2^2)h_{Q_1}^c \otimes h_{Q_2}^c ,$$

so $S_D\phi \lesssim [SM] f \cdot S_Dg$. Note that $\Gamma_{b(1,0)}$ works the same way, except we bound $S_D\phi$ by $[MS] f \cdot S_Dg$, and the remaining two paraproducts follow by duality. \qed
6.2. 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 $S_D$, in one of the variables:

$$
\pi_b^{(0)} f := \sum_{Q_1 \times Q_2} \left\langle b, h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle \left\langle f, \frac{1}{|Q_1|} \otimes h^{\gamma}_{Q_2} \right\rangle h^{\gamma}_{Q_1} \otimes h^{\gamma}_{Q_2}
$$

$$
\pi^*_b^{(0)} f := \sum_{Q_1 \times Q_2} \left\langle b, h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle \left\langle \hat{f}(Q_1^i \times Q_2^j), \frac{1}{|Q_1|} \otimes h^{\gamma}_{Q_2} \right\rangle h^{\gamma}_{Q_1} \otimes h^{\gamma}_{Q_2}
$$

$$
\pi_b^{(1)} f := \sum_{Q_1 \times Q_2} \left\langle b, \frac{1}{|Q_1|} \otimes h^{\gamma}_{Q_2} \right\rangle \left\langle f, h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle h^{\gamma}_{Q_1} \otimes h^{\gamma}_{Q_2}
$$

$$
\pi^*_b^{(1)} f := \sum_{Q_1 \times Q_2} \left\langle b, \frac{1}{|Q_1|} \otimes h^{\gamma}_{Q_2} \right\rangle \left\langle \hat{f}(Q_1^i \times Q_2^j), h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle h^{\gamma}_{Q_1} \otimes h^{\gamma}_{Q_2}
$$

$$
\gamma_b^{(0)} f := \sum_{Q_1 \times Q_2} \left\langle b, h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle \left\langle \hat{f}(Q_1^i \times Q_2^j), \frac{1}{|Q_1|} \otimes h^{\gamma}_{Q_2} \right\rangle h^{\gamma}_{Q_1} \otimes h^{\gamma}_{Q_2} + \gamma_b^{(0)} f
$$

$$
\gamma^*_b^{(0)} f := \sum_{Q_1 \times Q_2} \left\langle b, \frac{1}{|Q_1|} \otimes h^{\gamma}_{Q_2} \right\rangle \left\langle \hat{f}(Q_1^i \times Q_2^j), h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle h^{\gamma}_{Q_1} \otimes h^{\gamma}_{Q_2} + \gamma_b^{(0)} f.
$$

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

$$
||p_b : L^p(\mu) \to L^p(\lambda)|| \lesssim ||b||_{bmo_{\nu}(\nu)},
$$

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

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

$$
\langle \pi_{b(0)} f, g \rangle = \langle b, \phi \rangle, \text{ where } \phi := \sum_{Q_1 \times Q_2} \left\langle H^{\gamma}_{Q_2} f \right\rangle_{Q_1} \left\langle \hat{g}(Q_1^i \times Q_2^j), h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle.
$$

Then

$$
(S_D \phi)^2 \leq \left( \sum_{Q_1 \times Q_2} \left\langle H^{\gamma}_{Q_2} f \right\rangle_{Q_1}^2 \frac{1}{|Q_1|} \right) \left( \sum_{Q_1 \times Q_2} \left\langle \hat{g}(Q_1^i \times Q_2^j), h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle^2 \right) \frac{1}{|Q_1|} \frac{1}{|Q_2|} \frac{1}{|Q_1|} \frac{1}{|Q_2|}
$$

$$
= [MS]^2 f \cdot S^2_D g,
$$

and so

$$
||\langle \pi_{b(0)} f, g \rangle|| \lesssim ||b||_{bmo_{\nu}(\nu)} ||S_D \phi||_{L^1(\nu)} \lesssim ||b||_{bmo_{\nu}(\nu)} ||f||_{L^p(\mu)} ||g||_{L^p(\lambda)}.
$$

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

$$
\phi := \sum_{Q_1 \times Q_2} \hat{f}(Q_1^i \times Q_2^j) \frac{1}{|Q_1|} \left\langle \hat{g}(Q_1^i \times Q_2^j), h^{\gamma}_{Q_1} \otimes \frac{1}{|Q_2|} \right\rangle
$$

from which it easily follows that $S_D \phi \leq S_D f \cdot S_D g$. The proof for $\gamma_{b(1)}$ is symmetrical. □

7. Commutators with Journé Operators

7.1. Definition of Journé Operators. We begin with the definition of biparameter Calderón-Zygmund operators, or Journé operators, on $\mathbb{R}^d := \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}$, as outlined in [14]. As shown later in [10], these conditions are equivalent to the original definition of Journé [13].
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 \( \text{spt}(f_i) \cap \text{spt}(g_i) = \emptyset \) for \( i = 1, 2 \), we assume the kernel representation

\[
\langle T f, g \rangle = \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} K(x, y) f(y) g(x) \, dx \, dy.
\]

The kernel \( K : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \setminus \{(x, y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : x_1 = y_1 \text{ or } x_2 = y_2\} \to \mathbb{C} \) is assumed to satisfy:

1. Size condition:

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

2. Hölder conditions:

2a. If \(|y_1 - y_1'| \leq \frac{1}{2} |x_1 - y_1| \) and \(|y_2 - y_2'| \leq \frac{1}{2} |x_2 - y_2|\):

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

2b. If \(|x_1 - x_1'| \leq \frac{1}{2} |x_1 - y_1| \) and \(|x_2 - x_2'| \leq \frac{1}{2} |x_2 - y_2|\):

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

2c. If \(|y_1 - y_1'| \leq \frac{1}{2} |x_1 - y_1| \) and \(|x_2 - x_2'| \leq \frac{1}{2} |x_2 - y_2|\):

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

2d. If \(|x_1 - x_1'| \leq \frac{1}{2} |x_1 - y_1| \) and \(|y_2 - y_2'| \leq \frac{1}{2} |x_2 - y_2|\):

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

3. Mixed size and Hölder conditions:

3a. If \(|x_1 - x_1'| \leq \frac{1}{2} |x_1 - y_1|\), then \(|K(x, y) - K((x_1', x_2), y)| \leq C \frac{1}{|x_1 - y_1|^{n_1+\delta}} \frac{1}{|x_2 - y_2|^{n_2}}.
\]

3b. If \(|y_1 - y_1'| \leq \frac{1}{2} |x_1 - y_1|\), then \(|K(x, y) - K(x, (y_1', y_2))| \leq C \frac{1}{|x_1 - y_1|^{n_1+\delta}} \frac{1}{|x_2 - y_2|^{n_2}}.
\]

3c. If \(|x_2 - x_2'| \leq \frac{1}{2} |x_2 - y_2|\), then \(|K(x, y) - K((x_1', x_2), y)| \leq C \frac{1}{|x_1 - y_1|^{n_1}} \frac{1}{|x_2 - x_2'|^{n_2+\delta}}.
\]

3d. If \(|y_2 - y_2'| \leq \frac{1}{2} |x_2 - y_2|\), then \(|K(x, y) - K((x_1, y_1'), y)| \leq C \frac{1}{|x_1 - y_1|^{n_1}} \frac{1}{|y_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 \( \text{spt}(f_1) \cap \text{spt}(g_1) = \emptyset \), we assume the kernel representation:

\[
\langle T f, g \rangle = \int_{\mathbb{R}^{n_1}} \int_{\mathbb{R}^{n_2}} K_{f_2, g_2}(x_1, y_1) f_1(y_1) g_1(x_1) \, dx_1 \, dy_1,
\]

where the kernel \( K_{f_2, g_2} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \setminus \{(x_1, y_1) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_1} : x_1 = y_1\} \) satisfies the following size condition:

\[
|K_{f_2, g_2}(x_1, y_1)| \leq C(f_2, g_2) \frac{1}{|x_1 - y_1|^{n_1}},
\]

and Hölder conditions:

If \(|x_1 - x_1'| \leq \frac{1}{2} |x_1 - y_1|\), then \(|K_{f_2, g_2}(x_1, y_1) - K_{f_2, g_2}(x_1', y_1)| \leq C(f_2, g_2) \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n_1+\delta}}.
\]

If \(|y_1 - y_1'| \leq \frac{1}{2} |x_1 - y_1|\), then \(|K_{f_2, g_2}(x_1, y_1) - K_{f_2, g_2}(x_1, y_1')| \leq 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:

\[
C(\mathbb{I}_{Q_1}, \mathbb{I}_{Q_2}) + C(\mathbb{I}_{Q_1}, u_{Q_2}) + C(u_{Q_2}, \mathbb{I}_{Q_2}) \leq C|Q_2|.
\]
for all cubes $Q_2 \subset \mathbb{R}^{n_2}$ and all “$Q_2$-adapted zero-mean” functions $u_{Q_2}$ that is, $\text{spt}(u_{Q_2}) \subset Q_2$, $|u_{Q_2}| \leq 1$, and $\int u_{Q_2} = 0$. We assume the symmetrical representation with kernel $K_{f_1, g_1}$ in the case $\text{spt}(f_2) \cap \text{spt}(g_2) = \emptyset$.

II. Boundedness and Cancellation Assumptions:

1. Assume $T_1, T^*_1, T_1(1)$ and $T_1^*(1)$ are in product $\text{BMO}(\mathbb{R}^n)$, 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 $|\langle T(1_{Q_1} \otimes 1_{Q_2}), 1_{Q_1} \otimes 1_{Q_2} \rangle| \leq C|Q_1||Q_2|$, 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:

- $|\langle T(a_{Q_1} \otimes 1_{Q_2}), 1_{Q_1} \otimes 1_{Q_2} \rangle| \leq C|Q_1||Q_2|$, where $Q_i$ stands for the summation above.
- $|\langle T(1_{Q_1} \otimes 1_{Q_2}), a_{Q_1} \otimes 1_{Q_2} \rangle| \leq C|Q_1||Q_2|$, where $Q_i$ stands for the summation above.
- $|\langle T(1_{Q_1} \otimes b_{Q_2}), 1_{Q_1} \otimes 1_{Q_2} \rangle| \leq C|Q_1||Q_2|$, where $Q_i$ stands for the summation above.

7.2. Biparameter Dyadic Shifts and Martikainen’s Representation Theorem. Given dyadic rectangles $D = D_1 \times D_2$ and pairs of non-negative 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:

$$S^\vec{i} \vec{j} f := \sum_{R_i \in D_i} \sum_{R_2 \in D_2} \sum_{P_i \in (R_i)_{1}} a_{P_i, R_i, R_2, Q_1, Q_2} \hat{f}(P_1 \times P_2) h_{Q_1} \otimes h_{Q_2},$$

where

$$|a_{P_i, R_i, R_2, Q_1, Q_2}| \leq \sqrt{\frac{|Q_1|}{|R_1|}} \sqrt{\frac{|Q_2|}{|R_2|}} = 2^{-\frac{n_1}{2}(i_1+j_1)}2^{-\frac{n_2}{2}(i_2+j_2)}.$$

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

We use the simplified notation

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

for the summation above.

First note that

$$S^\vec{i} \vec{j}_D(f) = \sum_{R_i \times R_2 \in D_{1,2}} \left( \sum_{P_i \in (R_i)_{1}} a_{P_i, R_i, R_2, Q_1, Q_2} \hat{f}(P_1 \times P_2) \right)^2 \frac{1_{Q_1}}{|Q_1|} \otimes \frac{1_{Q_2}}{|Q_2|} \lesssim 2^{-n_1(i_1+j_1)}2^{-n_2(i_2+j_2)} \left(S^\vec{i} \vec{j} f \right)^2,$$

where $S^\vec{i} \vec{j}$ is the shifted biparameter square function in (3.1). Then, by (3.2):

$$\|S^\vec{i} \vec{j} f\|_{L^p(w)} \lesssim 2^{-\frac{n_1}{2}(i_1+j_1)}2^{-\frac{n_2}{2}(i_2+j_2)} \|S^\vec{i} \vec{j} f\|_{L^p(w)} \lesssim \|f\|_{L^p(w)},$$

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

Next, we state Martikainen’s Representation Theorem [14]:

**Theorem 7.1** (Martikainen). For a biparameter singular integral operator $T$ as defined in Section 7.1, there holds for some biparameter shifts $S^\vec{i} \vec{j}_D$ that:

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

where non-cancellative 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 $[b, S_D]$ with the dyadic shifts, with the requirements that the bounds be independent of the choice of $D$ 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 non-cancellative case in Section 7.4.
7.3. Cancellative Case.

**Theorem 7.2.** Let $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2$ be dyadic rectangles in $\mathbb{R}^n = \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ and $\mathcal{S}^i_j$ be a cancellative dyadic shift as defined in (7.1). If $\mu, \lambda \in A_\nu(\mathbb{R}^n)$, $1 < p < \infty$, and $\nu = \mu^{1/p} \lambda^{-1/p}$, then

$$
\|b, \mathcal{S}^i_j : L^p(\mu) \to L^p(\lambda)\| \lesssim \left( (1 + \max(i_1, j_1))(1 + \max(i_2, j_2)) \right) \|b\|_{bmo(\nu)},
$$

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

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

$$
b f = \sum P_b f + \sum p_b f + \Pi f b,
$$

where $P_b$ runs through the nine paraproducts associated with $BMO(\nu)$ in Section 6.1, and $p_b$ runs through the six paraproducts associated with $bmo(\nu)$ in Section 6.2. Then

$$
[b, \mathcal{S}^i_j]f = \sum [P_b, \mathcal{S}^i_j]f + \sum [p_b, \mathcal{S}^i_j]f + \mathcal{R}_{i,j}f,
$$

where

$$
\mathcal{R}_{i,j}f := \Pi \mathcal{S}^i_j b - \mathcal{S}^i_j \Pi f b.
$$

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),

$$
\left\| \sum [P_b, \mathcal{S}^i_j] + \sum [p_b, \mathcal{S}^i_j] : L^p(\mu) \to L^p(\lambda) \right\| \lesssim \|b\|_{bmo(\nu)},
$$

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

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

from which the result follows.

A straightforward calculation shows that

$$
\mathcal{R}_{i,j} = \sum_{\mathcal{R}, \mathcal{P}, \mathcal{Q}} a_{\mathcal{PQR}} f(P_1 \times P_2) \left( \langle b \rangle \mathcal{Q}_1 \mathcal{Q}_2 - \langle b \rangle \mathcal{P}_1 \mathcal{P}_2 \right) h_{\mathcal{Q}_1} \otimes h_{\mathcal{Q}_2}.
$$

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

$$
\langle b \rangle \mathcal{Q}_1 \mathcal{Q}_2 - \langle b \rangle \mathcal{P}_1 \mathcal{P}_2 = \left( \langle b \rangle \mathcal{Q}_1 \mathcal{Q}_2 - \langle b \rangle \mathcal{R}_1 \mathcal{R}_2 \right) + \left( \langle b \rangle \mathcal{R}_1 \mathcal{R}_2 - \langle b \rangle \mathcal{P}_1 \mathcal{P}_2 \right).
$$

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 $R_2 = Q_2^{(j_2)}$:

$$
\langle b \rangle \mathcal{Q}_1 \mathcal{Q}_2 - \langle b \rangle \mathcal{R}_1 \mathcal{R}_2 = \sum_{1 \leq k_1 \leq j_1, 1 \leq k_2 \leq j_2} \tilde{b}(\mathcal{Q}_1^{(k_1)} \times \mathcal{Q}_2^{(k_2)}) h_{\mathcal{Q}_1^{(k_1)}}(\mathcal{Q}_1) h_{\mathcal{Q}_2^{(k_2)}}(\mathcal{Q}_2)
$$

$$
+ \sum_{1 \leq k_1 \leq j_1} \left( b, h_{\mathcal{Q}_1^{(k_1)}} \otimes \frac{\mathcal{R}_2}{\mathcal{R}_1} h_{\mathcal{Q}_2^{(k_1)}}(\mathcal{Q}_1) \right) + \sum_{1 \leq k_2 \leq j_2} \left( b, \frac{\mathcal{R}_1}{\mathcal{R}_2} \otimes h_{\mathcal{Q}_2^{(k_2)}} h_{\mathcal{Q}_2^{(k_2)}}(\mathcal{Q}_2) \right).
$$

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

$$
\mathcal{R}_{i,j}^1 = \sum A_{k_1, k_2} f + \sum B_{k_1}^{(0,1)} f + \sum B_{k_2}^{(1,0)} f,
$$

where

$$
A_{k_1, k_2} f := \sum_{\mathcal{R}, \mathcal{P}, \mathcal{Q}} a_{\mathcal{PQR}} f(P_1 \times P_2) \tilde{b}(\mathcal{Q}_1^{(k_1)} \times \mathcal{Q}_2^{(k_2)}) h_{\mathcal{Q}_1^{(k_1)}}(\mathcal{Q}_1) h_{\mathcal{Q}_2^{(k_2)}}(\mathcal{Q}_2) h_{\mathcal{Q}_1} \otimes h_{\mathcal{Q}_2}.
$$
$$B_{k_1}^{(0,1)} f := \sum_{\mathbf{R}, \mathbf{P}, \mathbf{Q}} a_{\mathbf{PQR}} \hat{f}(P_1 \times P_2) \left( b, h_{Q_1^{(i_1)}} \otimes \frac{1}{|R_2|} h_{Q_2^{(i_2)}}(Q_1) h_{Q_1^{(i_1)}} \otimes h_{Q_2^{(i_2)}} \right),$$

and

$$B_{k_2}^{(1,0)} f := \sum_{\mathbf{R}, \mathbf{P}, \mathbf{Q}} a_{\mathbf{PQR}} \hat{f}(P_1 \times P_2) \left( b, \frac{|R_1|}{|R_1|} \otimes h_{Q_2^{(i_2)}}(Q_2) h_{Q_1^{(i_1)}} \otimes h_{Q_2^{(i_2)}} \right).$$

We show that these operators satisfy:

$$\|A_{k_1, k_2} : L^p(\mu) \rightarrow L^p(\lambda)\| \lesssim \|b\|_{BMO(\nu)}, \text{ for all } k_1, k_2,$$

$$\|B_{k_1}^{(0,1)} : L^p(\mu) \rightarrow L^p(\lambda)\| \lesssim \|b\|_{BMO(\nu)}, \text{ for all } k_1, \text{ and } \|B_{k_2}^{(1,0)} : L^p(\mu) \rightarrow L^p(\lambda)\| \lesssim \|b\|_{BMO(\nu)}, \text{ for all } k_2.$$
where the last operator is the shifted square function in (3.1). Then, from (3.2):

$$\|A_{k_1,k_2} : L^p(\mu) \to L^p(\lambda)\| \lesssim \|\hat{b}\|_{BMO(\nu)} \|S_{D}\phi\|_{L^1(\nu)}$$

$$\lesssim \|\hat{b}\|_{BMO(\nu)} 2^{-\frac{1}q(i_1+j_1-k_1)} 2^{-\frac{1}q(i_2+j_2-k_2)} \|M_Sg\|_{L^{p'}(\lambda')} \|S_{D}^{(i_1,i_2),(j_1-k_1,j_2-k_2)} f\|_{L^p(\mu)}$$

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

$$\phi = \sum_{R,P,Q} a_{PQR} \hat{f}(P \times P_1)h_{Q_1(k_1)}(Q_1)g_{Q_1}h_{Q_1(k_1)} \otimes \frac{\|R_2\|}{|R_2|}.$$ 

Then

$$S_D^{2,1} f \lesssim 2^{-n_1(i_1+j_1)} 2^{-n_2(i_2+j_2)} \sum_{R_i \in D_{k_1}} \sum_{Q_i \in D_{k_2}} \frac{1}{|N_1|} \left( \sum_{P_i \in (R_i)_{j_1}} \sum_{P_j \in (R_i)_{j_2}} \left| \hat{f}(P_1 \times P_2) \right| \right) \left( \sum_{Q_i \in (R_2)_{j_1}} \left( \sum_{Q_j \in (R_2)_{j_2}} |M_{D_1}(H_{Q_2}g)| \right) \|R_2\| \right) \left( \frac{\|R_2\|}{|R_2|} \right)^2,$$

and the summation above is bounded by:

$$\left( \sum_{R_i \in D_{k_1}} \sum_{Q_i \in D_{k_2}} \frac{1}{|N_1|} \left( \sum_{P_i \in (R_i)_{j_1}} \sum_{P_j \in (R_i)_{j_2}} \left| \hat{f}(P_1 \times P_2) \right| \right) \left( \sum_{Q_i \in (R_2)_{j_1}} \left( \sum_{Q_j \in (R_2)_{j_2}} |M_{D_1}(H_{Q_2}g)| \right) \|R_2\| \right) \right) \left( \frac{\|R_2\|}{|R_2|} \right)^2,$$

which is exactly

$$\left( S_{D}^{(i_1,i_2),(j_1-k_1,0)} f \right)^2 \left( |MS|^{j_2,0} g \right)^2.$$

From (3.2) and (3.3), we obtain exactly $\|S_{D}\phi\|_{L^1(\nu)} \lesssim \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda')}$, and the proof is complete. 

7.4. The Non-Cancellative Case. Following Martikainen’s proof in [14], we are left with three types of terms to consider – all of paraproduct type:

- The full standard paraproduct: $\Pi_a$ and $\Pi^*_a$,
- 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 $BMO(\mathbb{R}^n)$, with $\|a\|_{BMO(\mathbb{R}^n)} \leq 1$, and

- The partial paraproducts, defined for every $i_1, j_1 \geq 0$ as:

$$S_{D}^{i_1,j_1} f := \sum_{R_1 \in D_{i_1}} \sum_{Q_1 \in (R_1)_{j_1}} \hat{a}_{P_1,Q_1,R_1}(R_1^{Q_1}) \hat{f}(P_1^{R_1} \times R_1^{Q_1}) h_{Q_1}^{R_1} \otimes \frac{1}{|R_2|},$$

where, for every fixed $P_1, Q_1, R_1, a_{P_1,Q_1,R_1}(x)$ is a $BMO(\mathbb{R}^n)$ function with

$$\|a_{P_1,Q_1,R_1}\|_{BMO(\mathbb{R}^n)} \leq \frac{\sqrt{|P_1|} \sqrt{|Q_1|}}{|R_1|} = 2^{-\frac{1}q(i_1+j_1)},$$

and

$$\hat{a}_{P_1,Q_1,R_1}(R_1^{Q_1}) := \left( a_{P_1,Q_1,R_1} h_{R_1}^{Q_1} \right)_{\mathbb{R}^n} := \int_{\mathbb{R}^n} a_{P_1,Q_1,R_1}(x) h_{R_1}^{Q_1}(x) \, dx,$$

The symmetrical partial paraproduct $S_{D}^{i_2,j_2}$ is defined analogously.

We treat each case separately.
7.4.1. 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 $a \in BMO_\mathcal{D}(\mathbb{R}^n)$ with $\|a\|_{BMO_\mathcal{D}(\mathbb{R}^n)} \leq 1$. We prove that

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

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

**Proof.** Remark first that

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

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

$$\Pi_a(bf) - \Pi_a bf = \sum_{R \in \mathcal{D}} \hat{a}(R) \langle (bf)_R - (b)_R \langle f \rangle_R \rangle h_R$$

$$= \Pi_a \left( \sum_p P_a f + \sum_p P_b f + \Pi_f b \right) - \Pi_a bf,$$

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

$$\Pi_a bf - \Pi_a \Pi_a b = \Pi_a P_a f + \sum_{R \in \mathcal{D}} \hat{a}(R) \langle (bf)_R - (b)_R \langle f \rangle_R \rangle h_R.$$

Noting that

$$[b, \Pi_a]f = \sum P_b \Pi_a f + \sum p_b \Pi_a f - \sum P_a p_b f - \sum P_a \Pi_a f - \Pi_a \Pi_a b - \Pi_a \Pi_a f,$$

we obtain

$$[b, \Pi_a]f = \sum P_b \Pi_a f + \sum p_b \Pi_a f - \sum \hat{a}(R) \langle (bf)_R - (b)_R \langle f \rangle_R \rangle h_R.$$

The first terms are easily handled:

$$\| P_b \Pi_a f \|_{L^p(\nu)} \lesssim \|b\|_{BMO_\mathcal{D}(\nu)} \| \Pi_a f \|_{L^p(\mu)} \lesssim \|b\|_{BMO_\mathcal{D}(\nu)} \|a\|_{BMO_\mathcal{D}(\mathbb{R}^n)} \|f\|_{L^p(\mu)},$$

$$\| P_b \Pi_a f \|_{L^p(\nu)} \lesssim \|b\|_{bmo_\mathcal{D}(\nu)} \| \Pi_a f \|_{L^p(\mu)} \lesssim \|b\|_{bmo_\mathcal{D}(\nu)} \|a\|_{BMO_\mathcal{D}(\mathbb{R}^n)} \|f\|_{L^p(\mu)}.$$ 

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

$$\langle bf \rangle_R - \langle b \rangle_R \langle f \rangle_R = \frac{1}{|R|} \sum_{Q \subset Q_1 \times Q_2} \hat{b}(P_1 \times P_2) \hat{f}(P_1 \times P_2)$$

$$+ \frac{1}{|R|} \sum_{Q \subset Q_1} \left< b, h_{P_1} \otimes \mathbb{1}_{Q_2} \right|_{Q_1} \langle f, h_{P_1} \otimes h_{Q_2} \rangle + \frac{1}{|R|} \sum_{Q \subset Q_2} \left< b, \mathbb{1}_{Q_1} \otimes h_{P_2} \right|_{Q_2} \langle f, h_{P_1} \otimes h_{Q_2} \rangle.$$

Therefore

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

where:

$$\Lambda_{a,b} f := \sum_{Q_1 \times Q_2} \hat{a}(Q_1 \times Q_2) \mathbb{1}_{Q_1} \left( \sum_{Q \subset Q_1} \hat{b}(P_1 \times P_2) \hat{f}(P_1 \times P_2) \right) h_{Q_1} \otimes h_{Q_2},$$

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

$$\lambda_{a,b}^{(1,0)} f := \sum_{Q_1 \times Q_2} \hat{a}(Q_1 \times Q_2) \mathbb{1}_{Q_2} \left( \sum_{Q \subset Q_2} \left< b, \mathbb{1}_{Q_1} \otimes h_{P_2} \right|_{Q_2} \langle f, h_{P_1} \otimes h_{Q_2} \rangle \right) h_{Q_1} \otimes h_{Q_2},$$
\[ \lambda_{a,b}^{(1,0)} f := \sum_{Q_1 \times Q_2} \hat{a}(Q_1 \times Q_2) \frac{1}{|Q_1||Q_2|} \left( \sum_{P_2 \subset Q_2} \left( b, \frac{\mathbb{1}_{Q_1}}{|Q_1|} \otimes h_{P_2} \right) f, \mathbb{1}_{Q_1} \otimes h_{P_2} \right) h_{Q_1} \otimes h_{Q_2}. \]

To analyze the term \( \Lambda_{a,b} \), we write \( \langle \Lambda_{a,b}, f, g \rangle = \langle b, \phi \rangle \), where
\[
\phi = \sum_{P_1 \times P_2} \hat{f}(P_1) \times P_2 \left( \sum_{Q_1 \supset P_1} \hat{a}(Q_1 \times Q_2) \hat{g}(Q_1 \times Q_2) \frac{1}{|Q_1||Q_2|} \right) h_{P_1} \otimes h_{P_2} = \sum_{R \in \mathcal{D}} \hat{f}(R) \left( \sum_{T \in \mathcal{D}, T \supset R} \hat{a}(T) \hat{g}(T) \frac{1}{|T|} \right) h_{R}. \]

So \( |\langle \Lambda_{a,b}, f, g \rangle| \lesssim \|b\|_{BMO_{\mathcal{D}}} \|S_{\mathcal{D}} \phi\|_{L^1(\nu)} \), and
\[
S_{\mathcal{D}}^2 \phi = \sum_{R \in \mathcal{D}} |\hat{f}(R)|^2 \left( \sum_{T \in \mathcal{D}, T \supset R} |\hat{a}(T)\hat{g}(T)| \frac{1}{|T|} \right)^2 \frac{1}{|R|} \leq \sum_{R \in \mathcal{D}} |\hat{f}(R)|^2 \left( \sum_{T \in \mathcal{D}, T \supset R} |\hat{a}_r(T)\hat{g}_r(T)| \frac{1}{|T|} \right)^2 \frac{1}{|R|},
\]
where \( a_r := \sum_{R \in \mathcal{D}} |\hat{a}(R)||h_R |) \) and \( g_r := \sum_{R \in \mathcal{D}} |\hat{g}(R)||h_R |) \) are martingale transforms which do not increase either the \( BMO \) norm of \( a \), or the \( L^p \) norm of \( g \). Now note that
\[
\langle \Pi^*_a, g_r \rangle_R = \sum_{T \supset R} \hat{a}_r(T)\hat{g}(T) \frac{1}{|T|} + \sum_{T \supset R} \hat{a}_r(T)\hat{g}_r(T) \frac{1}{|T|},
\]
and since all the Haar coefficients of \( a_r \) and \( g_r \) are non-negative, we may write
\[
\sum_{T \supset R} \hat{a}_r(T)\hat{g}_r(T) \frac{1}{|T|} \leq \langle \Pi^*_a, g_r \rangle_R.
\]

Then
\[
S_{\mathcal{D}}^2 \phi \leq \sum_{R \in \mathcal{D}} |\hat{f}(R)|^2 \langle \Pi^*_a, g_r \rangle_R^2 \frac{1}{|R|} \leq (MS \Pi^*_a, g_r)^2 S_{\mathcal{D}}^2 f,
\]
and
\[
\|S_{\mathcal{D}} \phi\|_{L^1(\nu)} \leq \|MS \Pi^*_a, g_r\|_{L^p(\lambda)} \|S_{\mathcal{D}}^2 f\|_{L^p(\mu)} \leq \|\Pi^*_a, g_r\|_{L^p(\lambda)} \|f\|_{L^p(\mu)} \leq \|a_r\|_{BMO_{\mathcal{D}}} \|g_r\|_{L^p(\lambda)} \|f\|_{L^p(\mu)},
\]
which gives us the desired estimate
\[
\|\Lambda_{a,b} : L^p(\lambda) \to L^p(\mu)\| \lesssim \|a\|_{BMO_{\mathcal{D}}} \|b\|_{BMO_{\mathcal{D}}}.
\]

Finally, we analyze the term \( \lambda_{a,b}^{(0,1)} \), with the last term being symmetrical. We have \( \langle \lambda_{a,b}^{(0,1)}, f, g \rangle = \langle b, \phi \rangle \) with
\[
\phi = \sum_{P_1} \left( \sum_{Q_1 \supset P_1} f, h_{P_1} \otimes \mathbb{1}_{P_2} \right) \frac{1}{|P_1|} \sum_{Q_1 \supset P_1} \hat{a}(Q_1 \times P_2) \hat{g}(Q_1 \times P_2) \frac{1}{|Q_1|} \frac{|P_2|}{|P_1|} h_{P_1},
\]
and \( |\langle \lambda_{a,b}^{(0,1)}, f, g \rangle| \lesssim \|b\|_{bmo_{\mathcal{D}}} \|S_{\mathcal{D}} \phi\|_{L^1(\nu)} \). Now
\[
S_{\mathcal{D}}^2 \phi \leq \sum_{P_1} \left( \sum_{Q_1 \supset P_1} f, h_{P_1} \otimes \mathbb{1}_{P_2} \right) \frac{1}{|P_1|} \sum_{Q_1 \supset P_1} \hat{a}_r(Q_1 \times P_2) \hat{g}_r(Q_1 \times P_2) \frac{1}{|Q_1|} \frac{|P_2|}{|P_1|} \frac{|P_1|}{|P_2|},
\]
where we are using the same martingale transforms as above. Note that
\[
\langle \Pi^*_a, g_r \rangle_{R^{n_1}}(x_2) = \sum_{P_2} \frac{\mathbb{1}_{P_2}(x_2)}{|P_2|} \sum_{Q_1} \hat{a}_r(Q_1 \times P_2) \hat{g}_r(Q_1 \times P_2) \frac{|Q_1 \cap P_1|}{|Q_1||P_1|}.
\]
and again since all terms are non-negative:

\[
S_{D_1}^2 \phi \leq \sum_{P_1} M_{D_1}^2(H P_1 f)(x_2) \left( \sum_{Q_1 \supset P_1} \sum_{P_2} \hat{a}_r(Q_1 \times P_2) \hat{g}_r(Q_1 \times P_2) \frac{1}{|Q_1|} \frac{\mathbb{1}_{P_2}(x_2)}{|P_2|} \right)^2 \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} \\
\leq \sum_{P_1} M_{D_1}^2(H P_1 f)(x_2) \left( \left( \mathbb{1}_{P_1}^{\ast}, g_r \right)_{R^1} \right)^2 \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} \\
\leq \left( M_{D_1}(\Pi_a^{\ast}, g_r)(x_1, x_2) \right)^2 \sum_{P_1} M_{D_2}^2(H P_1 f)(x_2) \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} = \left( M_{D_1}(\Pi_a^{\ast}, g_r)(x_1, x_2) \right)^2 \left( |SM| f(x_1, x_2) \right)^2.
\]

Then

\[
\|S_{D_1} \phi\|_{L^1(\nu)} \lesssim \|\Pi_a^{\ast}, g_r\|_{L^p(\lambda')} \|[SM] f\|_{L^p(\mu)} \lesssim \|a\|_{BMO_2(R^d)} \|g\|_{L^p(\lambda')} \|f\|_{L^p(\mu)},
\]

and so

\[
\left\| \lambda^{(0,1)}_{a,b} : L^p(\mu) \to L^p(\lambda) \right\| \lesssim \|a\|_{BMO_2(R^d)} \|b\|_{bmo_2(\nu)},
\]

and the proof is complete. \( \square \)

7.4.2. The full mixed paraproduct. We are now dealing with \([b, \Pi_a^{(0,1)}]\), where

\[
\Pi_a^{(0,1)} f := \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \left( f, h_{P_1}^1 \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right) \frac{\mathbb{1}_{P_1}}{|P_1|} \otimes h_{P_2}^2.
\]

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

\[
\| [b, \Pi_a^{(0,1)}] : L^p(\mu) \to L^p(\lambda) \| \lesssim \|a\|_{BMO_2(R^d)} \|b\|_{bmo_2(\nu)}.
\]

Note that the case \([b, \Pi_a^{(1,0)}]\) follows symmetrically.

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

\[
R_{a,b}^{(0,1)} f := \Pi_{a^{(0,1)}} f b - \Pi_a^{(0,1)} \Pi_f b.
\]

Explicitly, these terms are:

\[
\Pi_{a^{(0,1)}} f b = \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \left( f, h_{P_1}^1 \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right) \left( \sum_{Q_1 \supset P_1} \langle b \rangle_{Q_1 \times P_2} h_{Q_1}^1(P_1) h_{Q_1}^2(x_1) \right) h_{P_2}^2(x_2),
\]

\[
\Pi_a^{(0,1)} \Pi_f b = \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \left( \sum_{Q_2 \supset P_2} \hat{f}(P_1 \times Q_2^1) \langle b \rangle_{P_1 \times Q_2^1} h_{Q_2}^2(P_2) \right) \frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} \otimes h_{P_2}^2(x_2).
\]

Consider now a third term

\[
T := \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \langle b \rangle_{P_1 \times P_2} \left( f, h_{P_1}^1 \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right) \frac{\mathbb{1}_{P_1}}{|P_1|} \otimes h_{P_2}^2(x_2).
\]

Using the one parameter formula:

\[
\frac{\mathbb{1}_{P_1}(x_1)}{|P_1|} = \sum_{Q_1 \supset P_1} h_{Q_1}^1(P_1) h_{Q_1}^2(x_1),
\]

we write \( T \) as

\[
T = \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \left( f, h_{P_1}^1 \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right) \left( \sum_{Q_1 \supset P_1} \langle b \rangle_{P_1 \times P_2} h_{Q_1}^1(P_1) h_{Q_1}^2(x_1) \right) h_{P_2}^2(x_2),
\]

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

\[
\Pi_{a^{(0,1)}} f b - T = \sum_{P_1 \times P_2} \hat{a}(P_1 \times P_2) \left( f, h_{P_1}^1 \otimes \frac{\mathbb{1}_{P_2}}{|P_2|} \right) \left( \sum_{Q_1 \supset P_1} \langle b \rangle_{Q_1 \times P_2} \left( \langle b \rangle_{Q_1 \times P_2} - \langle b \rangle_{P_1 \times P_2} \right) h_{Q_1}^1(P_1) h_{Q_1}^2(x_1) \right) h_{P_2}^2(x_2).
\]
Using (2.2):
\[
(b)_{Q_1 \times P_2} - (b)_{P_1 \times P_2} = - \sum_{R_1 : P_1 \subset R_1 \subset Q_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1).
\]
Then the term in parentheses above becomes
\[
(7.4) \quad - \sum_{Q_1 \supset P_1} \left( \sum_{R_1 : P_1 \subset R_1 \subset Q_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \right) \text{ s.t. } h_{Q_1}(x_1) = \sum_{R_1 \supset P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \frac{1}{|R_1|}.
\]
Next, we analyze this term depending on the relationship between $R_1$ and $Q_1$:

Case 1: $R_1 \subset Q_1$: Then we may rewrite the sum as
\[
\sum_{R_1 \supset P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \sum_{Q_1 \supset R_1} \left\langle h_{Q_1}^{\delta_1}(x_1) \right\rangle = \sum_{R_1 \supset P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \frac{1}{|R_1|}.
\]
This then leads to
\[
= \sum_{R_1 \times P_2} \left\langle f, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle \left( \sum_{R_1 \supset P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \frac{1}{|R_1|} \right) h_{R_2}^{\tau_2}(x_2)
\]
\[
= \sum_{R_1 \times P_2} \left\langle f, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle \left( \sum_{R_1 \supset P_1} \left\langle b, h_{R_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{R_1}^{\tau_1}(P_1) \frac{1}{|R_1|} \right) h_{R_2}^{\tau_2}(x_2)
\]
\[
= \gamma_{b(0,1)} \Pi_{a(0,1)} f.
\]
Case 2(a): $R_1 = Q_1$ and $\tau_1 \neq \delta_1$: Then (7.4) becomes:
\[
- \sum_{Q_1 \supset P_1} \left\langle b, h_{Q_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle \frac{1}{|Q_1|} h_{Q_1}^{\tau_1 + \delta_1}(P_1) h_{Q_1}^{\delta_1}(x_1),
\]
which leads to
\[
\sum_{Q_1 \times P_2} \left\langle b, h_{Q_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle \frac{1}{|Q_1|} h_{Q_1}^{\tau_1}(x_1) h_{P_2}^{\tau_2}(x_2) \sum_{P_1 \subset Q_1} \left\langle f, h_{P_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{P_1}^{\tau_1 + \delta_1}(P_1)
\]
\[
= \sum_{Q_1 \times P_2} \left\langle f, h_{Q_1}^{\tau_1 + \delta_1} \otimes h_{P_2}^{\tau_2} \right\rangle \frac{1}{|Q_1|} h_{Q_1}^{\delta_1}(x_1) h_{P_2}^{\tau_2}(x_2)
\]
\[
= \gamma_{b(0,1)} \Pi_{a(0,1)} f.
\]
Case 2(b): $R_1 = Q_1$ and $\tau_1 = \delta_1$: Then (7.4) becomes:
\[
\sum_{Q_1 \supset P_1} \left\langle b, h_{Q_1}^{\delta_1} \otimes \frac{1}{|P_2|} \right\rangle \frac{1}{|Q_1|} h_{Q_1}^{\delta_1},
\]
which gives rise to the term
\[
T^{(0,1)}_{a,b} f := \sum_{Q_1 \times P_2} \left\langle b, h_{Q_1}^{\delta_1} \otimes \frac{1}{|P_2|} \right\rangle h_{Q_1}^{\delta_1}(x_1) h_{P_2}^{\delta_2}(x_2) \frac{1}{|Q_1|} \sum_{P_1 \subset Q_1} \left\langle f, h_{P_1}^{\tau_1} \otimes \frac{1}{|P_2|} \right\rangle h_{P_1}^{\tau_1 + \delta_1}(P_1)
\]
We have proved that
\[
\Pi_{a(0,1)} f - T = - \pi_{b(0,1)} \Pi_{a(0,1)} f - \gamma_{b(0,1)} \Pi_{a(0,1)} f - T^{(0,1)}_{a,b} f.
\]
Expressing $T$ instead as
\[
T = \sum_{P_1 \times P_2} \left\langle f, h_{P_1}^{\tau_1} \otimes h_{P_2}^{\tau_2} \right\rangle \sum_{Q_2 \supset P_2} \left\langle b, h_{Q_2}^{\delta_1} \otimes \frac{1}{|P_2|} \right\rangle h_{Q_2}^{\delta_1}(P_2) \frac{1}{|P_1|} \otimes h_{P_2}^{\tau_2},
\]
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
\[
T_{a,b}^{(1,0)} f := \sum_{p_1 \times p_2} \tilde{a}(P_{p_1}^1 \times P_{p_2}^2) \frac{\|p_1 x_1}{|p_1|} \otimes h_{p_2}^2(x_2) \left( \sum_{Q_2 \supseteq p_2} b \frac{p_1}{|p_1|} \otimes h_{Q_2}^2 \right) \tilde{f}(P_{p_1}^1 \times P_{p_2}^2) \left( \frac{1}{|Q_2|} \right).
\]
Then
\[
R_{a,b}^{(0,1)} f = \Pi_{a;(0,1)}\pi_{b;(1,0)} f + \Pi_{a;(0,1)}\gamma_{b;(1,0)} f - \pi_{b;(1,0)} f - \gamma_{b;(1,0)} f = \tilde{T}_{a,b}^{(0,1)} f - \tilde{T}_{a,b}^{(1,0)} f.
\]
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
\[
\phi = \sum_{Q_1 \times p_2} \tilde{g}(Q_1^1 \times P_{p_2}^2) \frac{1}{|Q_1|} \left( \sum_{P_1 \subseteq Q_1} \tilde{a}(P_{p_1}^1 \times P_{p_2}^2) \left( f, h_{p_1}^1 \otimes \frac{p_1}{|p_1|} \right) / |p_2| \right) h_{Q_2}^2 \otimes \frac{p_2}{|p_2|}.
\]
Then $| \langle T_{a,b}^{(0,1)} f, g \rangle | \lesssim \|b\|_{\text{bmo}(\nu)} \|S_D \phi\|_{L^1(\nu)}$, and
\[
S_D^2 \phi = \sum_{Q_1} \left( \sum_{p_2} \tilde{g}(Q_1^1 \times P_{p_2}^2) \left( \frac{1}{|Q_1|} \sum_{P_1 \subseteq Q_1} \tilde{a}(P_{p_1}^1 \times P_{p_2}^2) \left( f, h_{p_1}^1 \otimes \frac{p_1}{|p_1|} \right) / |p_2| \right) \|p_2 \| / |p_2| \right) \frac{1}{|Q_1|}.
\]
Now,
\[
\langle \Pi_{a;(0,1)} f, \frac{p_1}{|p_1|} \otimes h_{Q_1}^2 \rangle = \sum_{P_1 \subseteq Q_1} \tilde{a}(P_{p_1}^1 \times P_{p_2}^2) \left( f, h_{p_1}^1 \otimes \frac{p_1}{|p_1|} \right) / |p_2| = 0.
\]
Define the martingale transform $a \mapsto a_\tau = \sum_{P_1 \times p_2} \tau_{p_1,p_2}^1 \tau_{p_1,p_2}^2 \tilde{a}(P_{p_1}^1 \times P_{p_2}^2)$, where
\[
\tau_{p_1,p_2}^1, \tau_{p_1,p_2}^2 = \begin{cases} +1, & \text{if } \left( f, h_{p_1}^1 \otimes \frac{p_1}{|p_1|} \right) \geq 0 \\ -1, & \text{otherwise.} \end{cases}
\]
Remark 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 $a_\tau$. Then we have
\[
\frac{1}{|Q_1|} \sum_{p_2} \tilde{a}(P_{p_1}^1 \times P_{p_2}^2) \left( f, h_{p_1}^1 \otimes \frac{p_1}{|p_1|} \right) / |p_2| \lesssim \left( \Pi_{a_\tau;(0,1)} f, \frac{p_1}{|p_1|} \otimes h_{Q_1}^2 \right).
\]
Returning to the square function estimate, we now have
\[
S^2_D \phi \lesssim \sum_{Q_1} \left( \sum_{p_2} \|\tilde{g}(Q_1^1 \times P_{p_2}^2)\|^2 / |p_2| \right) \left( \sum_{p_2} \|H_{p_2}^2 \Pi_{a_\tau;(0,1)} f\|^2 / |p_2| \right) \frac{1}{|Q_1|} \leq S^2_D \|M_S \Pi_{a_\tau;(0,1)} f\|^2 / |Q_1|,
\]
Finally,
\[
\|S_D \phi\|_{L^1(\nu)} \lesssim \|S_D g\|_{L^{p'}(\nu)} \left( \|M_S \Pi_{a_\tau;(0,1)} f\|^2 \right)_{L^p(\mu)} \lesssim \|a\|_{\text{BMO}(\mathbb{R}^d)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\nu)} \lesssim \|a\|_{\text{BMO}(\mathbb{R}^d)} \|b\|_{\text{bmo}(\nu)},
\]
showing that
\[
\|T_{a,b}^{(0,1)} : L^p(\mu) \rightarrow L^p(\lambda)\| \lesssim \|a\|_{\text{BMO}(\mathbb{R}^d)} \|b\|_{\text{bmo}(\nu)}.
\]
The estimate for $T_{a,b}^{(1,0)}$ follows similarly. \qed
7.4.3. The partial paraproducts. We work with

\[ S_{P}^{i,j} f := \sum_{R_1 \times R_2} \sum_{P_i \in (R_1)_{i_1}} \sum_{Q_j \in (R_1)_{j_2}} \hat{a}_{P_i Q_j R_1} (R_2^2) \hat{f}(P_1' \times R_2^2) h_{Q_j R_1}^{\delta_j} \otimes \frac{1}{|R_2|}, \]

where \( i_1, j_1 \) are non-negative integers, and for every \( P_1, Q_1, R_1 \):

\[ a_{P_i Q_j R_1} (x_2) \in BMO(\mathbb{R}^{n_2}) \quad \text{with} \quad \| a_{P_i Q_j R_1} \|_{BMO(\mathbb{R}^{n_2})} \leq 2^{-\frac{\nu}{p}(i_1 + j_1)}. \]

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

\[ \left\| [b, S_{P}^{i,j}] : L^p(\mu) \rightarrow L^p(\lambda) \right\| \lesssim \| b \|_{bmo(\nu)}. \]

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

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

\[ \left\| S_{P}^{i,j} : L^p(w) \rightarrow L^p(w) \right\| \lesssim 1. \tag{7.5} \]

**Proof.** Let \( f \in L^p(w) \) and \( g \in L^p(w') \), and show that \( \left\| S_{P}^{i,j} f, g \right\| \lesssim \| f \|_{L^p(w)} \| g \|_{L^p(w')} \):

\[ \left\| \left\langle S_{P}^{i,j} f, g \right\rangle \right\| \leq \sum_{R_1} \sum_{P_i \in (R_1)_{i_1}} \sum_{Q_j \in (R_1)_{j_2}} \| a_{P_i Q_j R_1} \|_{BMO(\mathbb{R}^{n_2})} \| S_{P} \phi_{P_i Q_j R_1} \|_{L^1(\mathbb{R}^{n_2})} \]

\[ \leq 2^{-\frac{\nu}{p}(i_1 + j_1)} \sum_{R_1} \sum_{P_i \in (R_1)_{i_1}} \sum_{Q_j \in (R_1)_{j_2}} \| S_{P} \phi_{P_i Q_j R_1} \|_{L^1(\mathbb{R}^{n_2})}, \]

where for every \( P_1, Q_1, R_1 \):

\[ \phi_{P_i Q_j R_1} (x_2) := \sum_{R_2} \hat{f}(P_1 \times R_2) \left\langle g, h_{Q_j R_1} \otimes \frac{1}{|R_2|} \right\rangle h_{R_2} (x_2). \]

Now,

\[ S_{P}^{i,j} \phi_{P_i Q_j R_1} = \sum_{R_2} \left| \hat{H}_{P_1} f(R_2) \right|^2 \langle [H_{Q_j} g] \rangle_{R_2} \frac{1}{|R_2|} h_{R_2} (x_2). \]

So

\[ \sum_{R_1} \sum_{P_i \in (R_1)_{i_1}} \sum_{Q_j \in (R_1)_{j_2}} \| S_{P} \phi_{P_i Q_j R_1} \|_{L^1(\mathbb{R}^{n_2})} \leq \sum_{R_1} \sum_{P_i \in (R_1)_{i_1}} \int_{\mathbb{R}^{n_2}} (M_{D} H_{Q_j} g(x_2)) (S_{P} H_{P_1} f(x_2)) \frac{1}{|R_2|} h_{R_2} (x_2) \]

\[ = \int_{\mathbb{R}^{n_2}} \int_{\mathbb{R}^{n_1}} \sum_{R_1} \sum_{P_i \in (R_1)_{i_1}} \sum_{Q_j \in (R_1)_{j_2}} (M_{D} H_{Q_j} g(x_2)) (S_{P} H_{P_1} f(x_2)) \frac{1}{|R_1|} \frac{1}{|R_2|} h_{R_2} (x_2) \]

\[ \leq \int_{\mathbb{R}^{n_2}} \left( \sum_{R_1} \left( \sum_{P_i \in (R_1)_{i_1}} S_{P} H_{P_1} f(x_2) \right)^2 \frac{1}{|R_1|} \right)^{1/2} \left( \sum_{R_1} \left( \sum_{Q_j \in (R_1)_{j_2}} M_{D} H_{Q_j} g(x_2) \right)^2 \frac{1}{|R_1|} \right)^{1/2} \]

\[ = \int_{\mathbb{R}^{n_2}} \left[ S S_{D} \right]^{i,j,0} f \cdot [S M_{D}]^{j,i} g w^{1/p} w^{-1/p} \ dx. \]
Then, from the estimates in (3.3):
\[
\| \langle S^{i,j_1}_D f, g \rangle \| \leq 2^{\frac{3}{2}(i_1+j_1)} \| [SS_{D_2}]^{i_1,0} f \|_{L^p(w)} \| [SM_{D_2}]^{j_1,0} g \|_{L^{p'}(w')}
\]
\[
\lesssim 2^{\frac{3}{2}(i_1+j_1)} 2^{\frac{3}{2}m} \| f \|_{L^p(w)} 2^{\frac{3}{2}m} \| g \|_{L^{p'}(w')},
\]
and the result follows.

**Proof of Theorem 7.5.** In light of (7.5), we only need to bound the remainder term
\[
R^{i_1,j_1} f := \Pi_S^{i_1,j_1} f - S^{i_1,j_1}_D f.
\]
The proof is somewhat similar to that of the full mixed paraproducts, in that we combine each of these terms:
\[
\Pi_S^{i_1,j_1} f = \sum_{R_1 \times R_2} \sum_{P_1 \in (R_1)} \sum_{Q_1 \in (R_1)} \tilde{a}_{P_1Q_1R_1}(R_2^2) \hat{f}(P_1^i \times R_2^2) \left( \sum_{Q_2 \supseteq R_2} \langle b \rangle_{Q_1 \times Q_2} h^{R_2}_{Q_2}(R_2) h^{R_2}_{Q_2}(x_2) \right) h^{R_1}_{Q_1}(x_1),
\]
\[
S^{i_1,j_1}_D f = \sum_{R_1 \times R_2} \sum_{P_1 \in (R_1)} \sum_{Q_1 \in (R_1)} \tilde{a}_{P_1Q_1R_1}(R_2^2) \hat{f}(P_1^i \times R_2^2) \langle b \rangle_{P_1 \times R_2} h^{R_1}_{Q_1}(x_1) \otimes \frac{\| R_2 \|_{R_2}}{\| R_2 \|_{R_2}}.
\]
with a third term:
\[
T := \sum_{R_1 \times R_2} \sum_{P_1 \in (R_1)} \sum_{Q_1 \in (R_1)} \tilde{a}_{P_1Q_1R_1}(R_2^2) \hat{f}(P_1^i \times R_2^2) \langle b \rangle_{Q_1 \times R_2} h^{R_1}_{Q_1} \otimes \frac{R_2}{R_2}.
\]
As before, expanding the indicator function in $T$ into its Haar series, we may combine $T$ with $\Pi_S^{i_1,j_1} f$:
\[
\Pi_S^{i_1,j_1} f - T = \sum_{R_1 \times R_2} \sum_{P_1 \in (R_1)} \sum_{Q_1 \in (R_1)} \tilde{a}_{P_1Q_1R_1}(R_2^2) \hat{f}(P_1^i \times R_2^2) \langle b \rangle_{Q_1 \times R_2} h^{R_1}_{Q_1}(x_1),
\]
where
\[
T_b(x_2) = \sum_{Q_2 \supseteq R_2} \left( \langle b \rangle_{Q_1 \times Q_2} - \langle b \rangle_{Q_1 \times P_2} \right) h^{R_2}_{Q_2}(R_2) h^{R_2}_{Q_2}(x_2)
\]
\[
= \sum_{Q_2 \supseteq R_2} \sum_{P_2 \subseteq Q_2} \langle b \rangle_{Q_1} \otimes h^{R_2}_{P_2} \left( \langle b \rangle_{Q_1} \otimes h^{R_2}_{P_2} \right) h^{R_2}_{Q_2}(R_2) h^{R_2}_{Q_2}(x_2).
\]
We analyze this term depending on the relationship of $P_2$ with $Q_2$.

**Case 1: $P_2 \subset Q_2$:** Then
\[
T_b(x_2) = \sum_{P_2 \supseteq R_2} \left( \langle b \rangle_{Q_1} \otimes h^{R_2}_{P_2} \right) h^{R_2}_{P_2}(R_2) \frac{\| P_2 \|_{P_2}}{\| P_2 \|_{P_2}},
\]
which gives the operator
\[
\sum_{Q_1 \times P_2} \left( \langle b \rangle_{Q_1} \otimes h^{R_2}_{P_2} \right) h^{R_1}_{Q_1}(x_1) \frac{\| P_2 \|_{P_2}}{\| P_2 \|_{P_2}} \left( \sum_{R_2 \subseteq P_2} \sum_{P_1 \in (Q_1)} \tilde{a}_{P_1Q_1R_1}(R_2^2) \hat{f}(P_1^i \times R_2^2) h^{R_2}_{P_2}(R_2) \right)
\]
\[
= \sum_{Q_1 \times P_2} \left( \langle b \rangle_{Q_1} \otimes h^{R_2}_{P_2} \right) h^{R_1}_{Q_1}(x_1) \frac{\| P_2 \|_{P_2}}{\| P_2 \|_{P_2}} \left( \sum_{P_1 \in (Q_1)} \Pi^*_a(P_1Q_1R_1) (H^{R_1}_{P_1} f, h^{R_2}_{P_2}) \right)
\]
\[
= \Pi^*_{b,(1,0)} F,
\]
where
\[
F := \sum_{Q_1 \times P_1} \Pi^*_a(P_1Q_1R_1) (H^{R_1}_{P_1} f)(x_2) h^{R_1}_{Q_1}(x_1).
\]
Now
\[
\| \Pi^*_{b,(1,0)} F \|_{L^p(\lambda)} \lesssim \| b \|_{bmo(\nu)} \| F \|_{L^p(\mu)},
\]
so we are done if we can show that

\[(7.6) \quad \|F\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)}.
\]

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

\[
| \langle F, g \rangle | \leq \sum_{Q_1} \sum_{P_i \in (Q_1^{(1)})_{i_1}} \left| \left\langle \Pi_{a_{P_i, Q_1, R_1}}(H_{P_i}^r f), H_{Q_1}^g \right\rangle \right|_{R^{n_2}}.
\]

Notice that we may write

\[
\left\langle \Pi_{a_{P_i, Q_1, R_1}}(H_{P_i}^r f), H_{Q_1}^g \right\rangle_{R^{n_2}} = \left\langle a_{P_i, Q_1, R_1}, \phi_{P_i, Q_1, R_1} \right\rangle_{R^{n_2}},
\]

where

\[
\phi_{P_i, Q_1, R_1}(x_2) = \sum_{R_2} \overline{H_{P_i}^r f(R_2^\delta)} \left\langle H_{Q_1}^g \right\rangle_{R_2} \eta_{R_2}^\delta(\gamma_2).
\]

Then

\[
| \langle F, g \rangle | \leq \sum_{Q_1} \sum_{P_i \in (Q_1^{(1)})_{i_1}} \|a_{P_i, Q_1, R_1}\|_{BMO(R^{n_2})} \|S_{D_2} \phi_{P_i, Q_1, R_1}\|_{L^1(R^{n_2})} \leq 2^{-2m(i_1 + j_1)} \sum_{R_1} \sum_{P_i \in (R_1^{(1)})_{i_1}} \int_{R^{n_2}} \left( \frac{\|H_{P_i}^r f(R_2^\delta)\|^2}{\|H_{Q_1}^g\|_{R_2}^2} \right)^{1/2} \frac{\|R_1(x_2)\|_{R_2}}{|R_1|} \frac{\|R_1(x_1)\|_{R_1}}{|R_1|} dx.
\]

The integral above is bounded by

\[
\int_{R^{n_2}} \left( \sum_{R_1} \sum_{P_i \in (R_1^{(1)})_{i_1}} (S_{D_2} H_{P_i}^r f)(x_2) \right)^2 \frac{\|R_1(x_1)\|_{R_1}}{|R_1|} \frac{\|R_1(x_1)\|_{R_1}}{|R_1|} dx = \int_{R^{n_2}} \left( \sum_{R_1} \sum_{P_i \in (R_1^{(1)})_{i_1}} (S_{D_2} H_{P_i}^r f)(x_2) \right)^2 \frac{\|R_1(x_1)\|_{R_1}}{|R_1|} \frac{\|R_1(x_1)\|_{R_1}}{|R_1|} dx \leq 2^{-2m(i_1 + j_1)} \|f\|_{L^p(\mu)} \|g\|_{L^{\rho'}(\mu')},
\]

by (3.3).

The desired estimate in (7.6) is now proved.

Case 2(a): \(P_2 = Q_2\) and \(\tau_2 \neq \delta_2\): Then

\[
T_b(x_2) = \sum_{Q_2 \supseteq R_2} \left( b_{Q_1} \boxtimes h_{Q_2}^2 \right) \frac{1}{|Q_2|} h_{Q_2}^{\tau_2 + \delta_2}(R_2) h_{Q_2}^{\delta_2}(x_2),
\]

giving rise to the operator

\[
\sum_{Q_1 \times Q_2} \left( b_{Q_1} \boxtimes h_{Q_2}^2 \right) \left( \sum_{P_i \in (Q_1^{(1)})_{i_1}} \Pi_{a_{P_i, Q_1, R_1}}(H_{P_i}^r f), h_{Q_2}^{\tau_2 + \delta_2} \right) \frac{1}{|Q_2|} h_{Q_1}^\delta \boxtimes h_{Q_2} = \gamma_{b(1,0)} F,
\]

which is handled as in the previous case.

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

\[
T' := \sum_{Q_1 \times Q_2} \left( b_{Q_1} \boxtimes h_{Q_2}^2 \right) h_{Q_1}^\delta \boxtimes h_{Q_2} \sum_{P_i \in (Q_1^{(1)})_{i_1}} \frac{1}{|Q_2|} \sum_{R_2 \supseteq Q_2} \hat{a}_{P_i, Q_1, R_1}(R_2^\delta) H_{P_i}^r f(R_2^\delta).
\]

Now define

\[
F_r := \sum_{Q_1} \left( \sum_{P_i \in (Q_1^{(1)})_{i_1}} \Pi_{a_{P_i, Q_1, R_1}}(H_{P_i}^r f)(x_2) \right) h_{Q_1}^\delta(x_1),
\]
just as we defined $F$ before, except now to every function $a_{P_1Q_1R_1}$ we apply the martingale transform

$$a_{P_1Q_1R_1} \mapsto a_{P_1Q_1R_1}' = \sum_{R_2} a_{R_2} \widehat{a}_{P_1Q_1R_1}(R_2')h_{R_2}'^\tau,$$

where $\tau_{R_2}' := \begin{cases} +1, & \text{if } H_{P_1}^\tau f(R_2') \geq 0, \\ -1, & \text{otherwise.} \end{cases}$

Since this does not increase the $BMO(\mathbb{R}^n)$ norms of the $a_{P_1Q_1R_1}$ functions, the estimate (7.6) still holds:

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

Moreover, note that $L^p(\mu)$.

Theorem 7.5 proves that $\Pi_{\tilde{i},j}^\tau b \to T$ obeys the desired bound, and the case $T - S_{\tilde{i},j}^\tau \Pi_{\tilde{i},j} b$ is handled similarly.

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

$$\|S_{\tilde{i}}^\tau b\|_{L^p(\mu)} \to L^p(\mu)} \| \lesssim 1,$$

for all $w \in A_p(\mathbb{R}^d)$. For the cancellative shifts, this was proved in (7.2). For the non-cancellative shifts, the first two types are simply paraproducts with symbol $\|a\|_{BMO(\mathbb{R}^n)} \leq 1$, while the third type, a partial paraproduct, was proved to be bounded on $L^p(\mu)$ in Proposition 7.6.

Theorem 1.4 now follows trivially from Martikainen’s representation Theorem 7.1: take $f \in L^p(\mu)$ and $g \in L^p(\mu')$. Then

$$\langle Tf, g \rangle \leq C_{T} E_{\omega_1} E_{\omega_2} \sum_{i_1, i_2 \in \mathbb{Z}^+} 2^{\max(i_1, i_2)\delta/2} 2^{\max(i_2, j_2)\delta/2} \| \langle S_{\tilde{i},j}^\tau f, g \rangle \|$$

$$\lesssim \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu')} \sum_{i_1, i_2 \in \mathbb{Z}^+} 2^{\max(i_1, i_2)\delta/2} 2^{\max(i_2, j_2)\delta/2}$$

$$\approx \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu')}.$$
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 \( b : \mathbb{R}^d \rightarrow \mathbb{C} \) with \( \vec{d} = (d_1, \ldots, d_t) \). Take a partition \( \mathcal{I} = \{ I_s : 1 \leq s \leq l \} \) of \( \{1, 2, \ldots, t\} \) so that \( \bigcup_{1 \leq s \leq l} I_s = \{1, 2, \ldots, t\} \). We say that \( b \in \text{BMO} \mathcal{I}(\mathbb{R}^d) \) if for any choices \( v = (v_s), v_s \in I_s \), \( b \) 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 \( \vec{x} = (x_1, \ldots, x_t) \in \mathbb{R}^d \), we define \( \vec{x}_v \) by removing those variables indexed by \( v_s \), the little product BMO norm becomes

\[
\|b\|_{\text{BMO}_v} = \max_{v} \left\{ \sup_{\vec{x}_v} \|b(\vec{x}_v)\|_{\text{BMO}} \right\}
\]

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

In [17] 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 (On-Petermichl-Strouse).** Let \( \vec{j} = (j_1, \ldots, j_t) \) with \( 1 \leq j_k \leq d_k \) and let for each \( 1 \leq s \leq l \), \( \vec{j}^{(s)} = (j_k)_{k \in I_s} \) be associated a tensor product of Riesz transforms \( \vec{R}_{\vec{j}^{(s)}} = \bigotimes_{k \in I_s} R_{k,j_k} \); here \( R_{k,j_k} \) are \( j_k \) Riesz transforms acting on functions defined on the \( k \)th variable. We have the two-sided estimate

\[
\|b\|_{\text{BMO} \mathcal{I}(\mathbb{R}^d)} \lesssim \sup_{\vec{j}} \left[ \|\vec{R}_{\vec{j} \downarrow 1} ; \ldots ; \|\vec{R}_{\vec{j} \downarrow l} (b) \| \right]_{L^p(\mathbb{R}^d)} \lesssim \|b\|_{\text{BMO} \mathcal{I}(\mathbb{R}^d)}.
\]

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

\[
\|T_1, \ldots, T_t (b) \|_{L^2(\mathbb{R}^d)} \lesssim \|b\|_{\text{BMO} \mathcal{I}(\mathbb{R}^d)}.
\]

This estimate was generalised somewhat in [18] in that the paraproduct free condition was slightly weakened, the considerations in this present text in combination with arguments from [5] and [17] 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 \( \mathbb{R}^d, \vec{d} = (d_1, \ldots, d_t) \) with a partition \( \mathcal{I} = (I_s)_{1 \leq s \leq l} \) of \( \{1, \ldots, t\} \) as discussed before. Let \( b \in \text{BMO} \mathcal{I}(\mathbb{R}^d) \) and let \( T_s \) denote a multi-parameter Journé operator acting on function defined on \( \bigotimes_{k \in I_s} \mathbb{R}^{d_k} \). Then we have the estimate below

\[
\|T_1, \ldots, T_t (b) \|_{L^p(\mathbb{R}^d)} \lesssim \|b\|_{\text{BMO} \mathcal{I}(\mathbb{R}^d)}.
\]

Acknowledgement: We are grateful to Jill Pipher and Yumeng Ou for helpful suggestions, especially in pointing out to us the connection with [7] and [8]. The first and second author 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**

[1] R.J. Bagby and D.S. Kurtz, \( L(\log L) \) spaces and weights for the strong maximal function, J. Anal. Math. **44** (1984), no. 1, 21–31. \( \dagger \)

[2] S. Bloom, A commutator theorem and weighted BMO, Trans. Amer. Math. Soc. **292** (1985), no. 1, 103–122. \( \dagger \)

[3] R. R. Coifman, R. Rochberg, and Guido Weiss, Factorization theorems for Hardy spaces in several variables, Ann. of Math. (2) **103** (1976), no. 3, 611–635. \( \dagger \), 18

[4] M. Cotlar and C. Sadosky, Two distinguished subspaces of product BMO and Nehari-AAK theory for Hankel operators on the torus, Integr. Equ. Oper. Theory **26** (1996), no. 3. \( \dagger \)

[5] L. Dalenc and Y. Ou, Upper bound for multi-parameter iterated commutators, Preprint (2014). \( \dagger \)

[6] R. Fefferman and E. M. Stein, Singular integrals on product spaces, Adv. Math. **45** (1982), no. 2. \( \dagger \)

[7] R. Fefferman, Harmonic Analysis on Product Spaces, Ann. of Math. **126** (1987), no. 1. \( \dagger \), 5, 37
A. Grau de la Herran, Comparison of T1 conditions for multi-parameter operators, Proc. Amer. Math. Soc 144 (2016), no. 6. ↑22

I. Holmes, M. T. Lacey, and B. D. Wick, Commutators in the two-weight setting, Math. Ann. 367 (2017), no. 1-2, 51–80, available at http://arxiv.org/abs/1506.05747. ↑1, 2, 7, 12, 20

T. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, Ann. of Math. (2) 175 (2012), no. 3, 1473–1506. ↑1, 2

J-L. Journé, Calderón-Zygmund operators on product spaces, Rev. Mat. Iberoamericana 1 (1985), no. 3. ↑22

H. Martikainen and T. Orponen, Some obstacles in characterising the boundedness of bi-parameter singular integrals, Math. Z. 282 (2016). ↑5

B. Muckenhoupt and R. L. Wheeden, Weighted bounded mean oscillation and the Hilbert transform, Studia Math. 54 (1975/76), no. 3. ↑12

Y. Ou, S. Petermichl, and E. Strouse, Higher order Journé commutators and characterizations of multi-parameter BMO, Adv. Math. 291 (2016), 24–58. ↑2, 4, 37

Y. Ou and S. Petermichl, Little bmo and Journé commutators, available at http://yumengou.mit.edu/sites/default/files/documents/LittleBMOJourneYOSP.pdf. ↑2, 37

S. Petermichl, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol, C. R. Acad. Sci. Paris Sér. I Math. 330 (2000), no. 6. ↑1

J. L. Rubio de Francia, J. Ruiz, and J. L. Torrea, Calderón-Zygmund theory for operator-valued kernels, Adv. Math. 62 (1986), 7–48. ↑

S. Wu, A wavelet characterization for weighted Hardy Spaces, Rev. Mat. Iberoamericana 8 (1992), no. 3. ↑12