TWO WEIGHT INEQUALITIES FOR ITERATED COMMUTATORS WITH CALDERÓN–ZYGMUND OPERATORS

IRINA HOLMES and BRETT D. WICK

Communicated by Hari Bercovici

ABSTRACT. Given a Calderón–Zygmund operator T, a classic result of Coifman, Rochberg and Weiss relates the norm of the commutator $[b, T]$ with the BMO norm of b. We focus on a weighted version of this result, obtained by Bloom and later generalized by Lacey and the authors, which relates $\|[b, T] : L^p(\mathbb{R}^n; \mu) \to L^p(\mathbb{R}^n; \lambda)\|$ to the norm of b in a certain weighted BMO space determined by A_p weights μ and λ. We extend this result to higher iterates of the commutator and recover a one-weight result of Chung, Pereyra and Perez in the process.

KEYWORDS: *Commutators, Calderón–Zygmund operators, bounded mean oscillation, weights.*

MSC (2010): 42, 42A, 42B, 42B20, 42B25, 42A50, 42A40.

1. INTRODUCTION

A Calderón–Zygmund operator associated to a kernel $K(x, y)$ is an integral operator:

$$Tf(x) := \int_{\mathbb{R}^n} K(x, y) f(y) \, \mathrm{d}y, \quad x \notin \mathrm{supp} f,$$

where the kernel satisfies the standard size and smoothness estimates

$$|K(x, y)| \leqslant \frac{C}{|x - y|^n},$$

$$|K(x + h, y) - K(x, y)| + |K(x, y + h) - K(x, y)| \leqslant C \frac{|h|^\delta}{|x - y|^{n + \delta}},$$

for all $|x - y| > 2|h| > 0$ and a fixed $\delta \in (0, 1]$. We say such a kernel is δ-standard kernel. The prototypes for this important class of operators are the Hilbert transform, in the one-dimensional case, and the Riesz transforms, in the multidimensional case.

Recall that the commutator $[S, T]$ of two operators S and T is defined as $[S, T] := ST - TS$. We are interested in commutators of multiplication by a symbol b with Calderón–Zygmund operators T, denoted $[b, T]$ and defined as:

$$[b, T]f := bTf - T(bf).$$

In the foundational paper [7] Coifman, Rochberg, and Weiss provided a connection between the norm of the commutator $[b, T] : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ and the norm of the function b in BMO. This result was later extended to the case when the commutator acts between two different weighted Lebesgue spaces $L^p(\lambda) := L^p(\mathbb{R}^n; \lambda)$ and $L^p(\mu) := L^p(\mathbb{R}^n; \mu)$. In 1985, Bloom [2] showed that, if μ and λ are A_p weights, then $\|[b, H] : L^p(\mu) \to L^p(\lambda)\|$ is equivalent to $\|b\|_{\mathrm{BMO}(\nu)}$, where H is the Hilbert transform and $\mathrm{BMO}(\nu)$ is the weighted BMO space associated with the weight $\nu = \mu^{1/p}\lambda^{-1/p}$. Here,

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

where $\nu(Q) = \int_Q \mathrm{d}\nu$, and the supremum is over all cubes Q. When there is no weight involved, we will simply denote this space by BMO, which is the classical space of functions with bounded mean oscillation. A new dyadic proof of Bloom's result was given in [11]. This was then generalized to all Calderón–Zygmund operators in [12], where one of the main results is the following theorem.

THEOREM 1.1. *Let T be a Calderón–Zygmund operator on $\mathbb{R}^n$ and $\mu, \lambda \in A_p$ with $1 < p < \infty$. Suppose $b \in \mathrm{BMO}(\nu)$, where $\nu = \mu^{1/p}\lambda^{-1/p}$. Then*

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

where c is a constant depending on the dimension n, the operator T, and μ, λ, and p.

A natural extension of this is to consider higher iterates of this commutator. To see how these arise naturally, we follow an argument of Coifman, Rochberg, and Weiss, [7]. For $b \in \mathrm{BMO}$ and r sufficiently small, consider the operator:

$$S_r(f) = e^{rb}T(e^{-rb}f)$$

Then it is easy to see that $\frac{\mathrm{d}}{\mathrm{d}r}S_r(f)|_{r=0} = [b, T](f)$ and similarly that we have $\frac{\mathrm{d}^n}{\mathrm{d}r^n}S_r(f)|_{r=0} = [b, \ldots, [b, [b, T]], \ldots]$ with the function b appearing n times. For some Calderón–Zygmund operator T, let $C_b^1(T) := [b, T]$, and

$$C_b^k(T) := [b, C_b^{k-1}(T)] \quad \text{for all integers } k \geqslant 1.$$

Using weighted theory and the connection between the space BMO and A_2 weights, it is then easy to see that the norm of the operator $C_b^k(T)$ on $L^2(\mathbb{R}^n)$ depends on the number of iterates and the norm of the function $b \in \mathrm{BMO}$. At this point, a few natural questions arise: (1) What is the norm of the kth iterate as a

function of the norm of $b \in \mathrm{BMO}$? (2) What happens if we attempt to compute the norm of this operator when it acts on $L^p(\mathbb{R}^n; w)$ for a weight $w \in A_p$? (3) Is there an extension of Theorem 1.1 for the iterates?

In the paper [6] Chung, Pereyra, and Perez provide answers to questions (1) and (2) and show that

$$\|C_b^k(T) : L^2(w) \to L^2(w)\| \leqslant c\|b\|_{\mathrm{BMO}}^k [w]_{A_2}^{k+1},$$

where c is a constant depending on n, k and T. In fact, they show that, more generally, if T is any operator bounded on $L^2(\mathbb{R}^n; w)$ with norm $\varphi([w]_{A_2})$ then

$$\|C_b^k(T) : L^2(w) \to L^2(w)\| \leqslant \varphi([w]_{A_2})[w]_{A_2}^k \|b\|_{\mathrm{BMO}}^k,$$

a result which they then extend to the L^p case using Rubio de Francia extrapolation. However, the two weight extension of Theorem 1.1 lies outside the scope of the results in [6]. Additionally, in pag. 1166 of [6] they ask if it is possible to provide a proof of the norm of the iterates of commutators with Calderón–Zygmund operators via the methods of dyadic analysis. The main goal of this paper is to extend Theorem 1.1 to the case of iterates, addressing question (3), and in the process show how to answer the question raised in [6]. This leads to the main result of the paper.

THEOREM 1.2. *Let T be a Calderón–Zygmund operator on $\mathbb{R}^n$ and $\mu, \lambda \in A_p$ with $1 < p < \infty$. Suppose $b \in \mathrm{BMO} \cap \mathrm{BMO}(\nu)$, where $\nu = \mu^{1/p}\lambda^{-1/p}$. Then for all integers $k \geqslant 1$,*

$$\|C_b^k(T) : L^p(\mu) \to L^p(\lambda)\| \leqslant c\|b\|_{\mathrm{BMO}}^{k-1} \|b\|_{\mathrm{BMO}(\nu)},$$

where c is a constant depending on n, k, T, μ, λ, and p.

In particular, if $\mu = \lambda = w \in A_2$,

$$\|C_b^k(T) : L^2(w) \to L^2(w)\| \leqslant c\|b\|_{\mathrm{BMO}}^k [w]_{A_2}^{k+1},$$

where c is a constant depending on n, k and T.

The paper is structured as follows. In Section 2 we discuss the necessary background and notation, such as the Haar system, dyadic shifts, and weighted BMO spaces. Note that most of these concepts were also needed in [12], and are treated in more detail there. In Section 3 we show how, through the Hytönen representation theorem, it suffices to prove our main result for dyadic shifts $\mathbb{S}^{ij}$. The rest of the paper is dedicated to this. In Section 4 we revisit the two-weight proof for the first commutator $[b, \mathbb{S}^{ij}]$ in [12], making some definitions which will be useful later, and obtaining the one-weight result. In Section 5 we look at the second iteration $[b, [b, \mathbb{S}^{ij}]]$ — this will provide the intuition behind the general case of k iterations, and also establish the final tools needed for this. In Section 6 we prove the general result.

2. BACKGROUND AND NOTATION

2.1. THE HAAR SYSTEM. Let $\mathcal{D}^0 = \{2^{-k}([0,1)^n + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^n\}$ be the standard dyadic grid on $\mathbb{R}^n$. For any $\omega = (\omega_j)_{j \in \mathbb{Z}} \in (\{0,1\}^n)^{\mathbb{Z}}$, we let $\mathcal{D}^\omega :=$ $\{Q \dot{+} \omega : Q \in \mathcal{D}^0\}$ be the translate of $\mathcal{D}^0$ by ω, where

$$Q \dot{+} \omega := Q + \sum_{j:2^{-j} < l(Q)} 2^{-j} \omega_j,$$

where $l(Q)$ denotes the side length of any cube Q in $\mathbb{R}^n$. Every dyadic grid $\mathcal{D}^\omega$ is characterized by two fundamental properties, namely: (1) For every $P, Q \in \mathcal{D}^\omega$, $P \cap Q \in \{P, Q, \varnothing\}$, and (2) For every fixed $k \in \mathbb{Z}$, the cubes $Q \in \mathcal{D}^\omega$ with $l(Q) = 2^{-k}$ partition $\mathbb{R}^n$. Let $\mathcal{D}$ be a fixed dyadic grid, $Q \in \mathcal{D}$, and k be a non-negative integer. We let $Q^{(k)}$ denote the kth ancestor of Q in $\mathcal{D}$, i.e. the unique element of $\mathcal{D}$ with side length $2^k l(Q)$ that contains Q, and $Q_{(k)}$ denote the collection of kth descendants of Q in $\mathcal{D}$, i.e. the 2^{kn} disjoint subcubes of Q in $\mathcal{D}$ with side length $2^{-k} l(Q)$.

The Haar system on $\mathcal{D}$ is defined by associating 2^n Haar functions to every $Q = I_1 \times \cdots \times I_n \in \mathcal{D}$, where each I_i is a dyadic interval in $\mathbb{R}$ with length $l(Q)$,

$$h_Q^\epsilon(x) := \prod_{i=1}^n h_{I_i}^{\epsilon_i}(x_i),$$

for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, where $\epsilon \in \{0,1\}^n$ is called the signature of h_Q^ϵ, and $h_{I_i}^{\epsilon_i}$ is one of the one-dimensional Haar functions:

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

We write $\epsilon = 1$ when $\epsilon_i = 1$ for all i. In this case, $h_Q^1 = |Q|^{-1/2}\mathbb{1}_Q$ is said to be non-cancellative, while all the other $2^n - 1$ Haar functions associated with Q are cancellative. Moreover, the cancellative Haar functions on a fixed dyadic grid form an orthonormal basis for $L^2(\mathbb{R}^n)$. We then write for any $f \in L^2(\mathbb{R}^n)$,

$$f = \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \widehat{f}(Q, \epsilon) h_Q^\epsilon,$$

where $\widehat{f}(Q, \epsilon) := \langle f, h_Q^\epsilon \rangle$ and $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $L^2(\mathbb{R}^n)$. Then

$$\langle f \rangle_Q = \sum_{R \in \mathcal{D}, R \supsetneq Q; \epsilon \neq 1} \widehat{f}(R, \epsilon) h_R^\epsilon(Q),$$

where $\langle f \rangle_Q := |Q|^{-1} \int_Q f \, dx$ denotes the average of f over Q.

2.2. A_p WEIGHTS. By a weight on $\mathbb{R}^n$ we mean an almost everywhere positive, locally integrable function w. For some $1 < p < \infty$ with Hölder conjugate q, we say that a weight w belongs to the Muckenhoupt A_p class if

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

where the supremum is over all cubes in $\mathbb{R}^n$. We let $w' := w^{1-q}$, the "conjugate" weight to w. Then $w \in A_p$ if and only if $w' \in A_q$, with $[w']_{A_q} = [w]_{A_p}^{q-1}$.

For a weight w and $1 < p < \infty$, let $L^p(w)$ denote the usual L^p-space with respect to the measure $dw = w\,dx$, i.e. the space of all functions f such that $\|f\|_{L^p(w)}^p = \int_{\mathbb{R}^n} |f|^p \, dw < \infty$. If $w \in A_p$, we then have the duality $(L^p(w))^* \equiv L^q(w')$, in the sense that

$$\|f\|_{L^p(w)} = \sup\{|\langle f, g \rangle| : g \in L^q(w'), \|g\|_{L^q(w')} \leqslant 1\}.$$

We review some of the crucial properties of A_p weights, starting with the maximal function

$$Mf := \sup_Q (\langle |f| \rangle_Q \mathbb{1}_Q),$$

where again the supremum is over all cubes Q in $\mathbb{R}^n$. If $w \in A_p$, then the following bound is sharp [3] (as the referee pointed out, while the sharp inequality first appeared in [3], this result is implicit in the proof of Christ and Fefferman [4]) in the exponent of $[w]_{A_p}$:

$$(2.1) \qquad \|M\|_{L^p(w)} \lesssim [w]_{A_p}^{1/(p-1)},$$

where for some quantities A and B, "$A \lesssim B$" denotes $A \leqslant CB$ for some constant C. This constant will usually depend on the dimension, the exponent p, and the A_p characteristics of the weights involved. Another important tool is the dyadic square function

$$(S_{\mathcal{D}}f)^2 := \sum_{Q \in \mathcal{D}, \epsilon \neq 1} |\widehat{f}(Q, \epsilon)|^2 \frac{\mathbb{1}_Q}{|Q|},$$

for which we have the sharp [8] one-weight inequality

$$(2.2) \qquad \|S_{\mathcal{D}}\|_{L^p(w)} \lesssim [w]_{A_p}^{\max(1/2, 1/(p-1))}.$$

For a dyadic grid $\mathcal{D}$ on $\mathbb{R}^n$ and a pair (i, j) of non-negative integers, define a shifted dyadic square function

$$(2.3) \qquad (\widetilde{S}_{\mathcal{D}}^{i,j}f)^2 := \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \left(\sum_{P \in (Q^{(j)})_{(i)}} |\widehat{f}(P, \epsilon)| \right)^2 \frac{\mathbb{1}_Q}{|Q|}.$$

The following was proved in Lemma 2.2 of [12]:

$$(2.4) \qquad \|\widetilde{S}_{\mathcal{D}}^{i,j}\|_{L^2(w)} \lesssim 2^{(n/2)/(i+j)} [w]_{A_2}.$$

Lastly, we recall the extrapolation property of A_p weights [10]. Suppose an operator T satisfies

$$\|Tf\|_{L^2(w)} \lesssim A[w]_{A_2}^{\alpha} \|f\|_{L^2(w)},$$

for all $w \in A_2$, for some fixed $A > 0$ and $\alpha > 0$. Then,

$$\|Tf\|_{L^p(w)} \lesssim A[w]_{A_p}^{\alpha \max(1, 1/(p-1))} \|f\|_{A_p},$$

for all $1 < p < \infty$ and all $w \in A_p$.

2.3. WEIGHTED BMO. Let w be a weight on $\mathbb{R}^n$. The weighted BMO space $\mathrm{BMO}(w)$ is the space of all locally integrable functions b such that

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

where $w(Q) = \int_Q dw$, and the supremum is over all cubes Q. Note that if we take $w = 1$ we obtain the usual space of functions with bounded mean oscillation, which we simply denote by BMO. If $w \in A_p$, it was shown in [19] that $\| \cdot \|_{\mathrm{BMO}(w)}$ is equivalent to the norm $\| \cdot \|_{\mathrm{BMO}^q(w)}$, defined as

$$\|b\|_{\mathrm{BMO}^q(w)}^q := \sup_Q \frac{1}{w(Q)} \int_Q |b - \langle b \rangle_Q|^q \, dw'.$$

Given a dyadic grid $\mathcal{D}$, we define the dyadic versions of these spaces, $\mathrm{BMO}_{\mathcal{D}}(w)$ and $\mathrm{BMO}_{\mathcal{D}}^q(w)$, by taking the supremum over $Q \in \mathcal{D}$ instead.

Now suppose $\mu, \lambda \in A_p$ and define $\nu := \mu^{1/p} \lambda^{-1/p}$. As shown in [12], ν is then an A_2 weight. The following inequality will be very useful:

(2.5) $|\langle b, \Phi \rangle| \lesssim [\nu]_{A_2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} \|S_{\mathcal{D}} \Phi\|_{L^1(\nu)}.$

This in fact holds for all A_2 weights w, and comes from a duality relationship between $\mathrm{BMO}_{\mathcal{D}}^2(w)$ and the dyadic weighted Hardy space $\mathcal{H}_{\mathcal{D}}^1(w)$. See Section 2.6 of [12] for details.

2.4. PARAPRODUCTS. Recall the paraproducts with symbol b on $\mathbb{R}$:

$$\Pi_b f = \sum_{I \in \mathcal{D}} \widehat{b}(I) \langle f \rangle_I h_I \quad \text{and} \quad \Pi_b^* f = \sum_{I \in \mathcal{D}} \widehat{b}(I) \widehat{f}(I) |I|^{-1} \mathbb{1}_I.$$

These are most useful in dyadic proofs due to the identity $bf = \Pi_b f + \Pi_b^* f + \Pi_f b$. To generalize this property to $\mathbb{R}^n$, we define the multidimensional paraproducts below.

DEFINITION 2.1. For a fixed dyadic grid $\mathcal{D}$ on $\mathbb{R}^n$, define the following paraproduct operators with symbol b:

$$\Pi_b f := \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \widehat{b}(Q, \epsilon) \langle f \rangle_Q h_Q^\epsilon, \quad \Pi_b^* f := \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \widehat{b}(Q, \epsilon) \widehat{f}(Q, \epsilon) \frac{\mathbb{1}_Q}{|Q|}, \quad \text{and}$$

$$\Gamma_b f := \sum_{Q \in \mathcal{D}} \sum_{\epsilon, \eta \neq 1; \epsilon \neq \eta} \widehat{b}(Q, \epsilon) \widehat{f}(Q, \eta) \frac{1}{\sqrt{|Q|}} h_Q^{\epsilon+\eta},$$

where for every $\epsilon, \eta \in \{0,1\}^n$, $\epsilon + \eta$ is defined by letting $(\epsilon + \eta)_i$ be 0 if $\epsilon_i \neq \eta_i$ and 1 otherwise.

Then $bf = \Pi_b f + \Pi_b^* f + \Gamma_b f + \Pi_f b$. Note that, while the first two paraproducts above reduce to the standard one-dimensional ones when $n = 1$, the third paraproduct Γ_b vanishes in this case. This third paraproduct comes from the fact that $h_Q^\epsilon h_Q^\eta = |Q|^{-1/2} h_Q^{\epsilon+\eta}$.

For ease of notation later, we denote

$$\|T\|_{L^p(w;v)} := \|T : L^p(w) \to L^p(v)\|,$$

the operator norm between two weighted L^p-spaces. And, when $w = v$ we will frequently write

$$\|T\|_{L^p(w)} := \|T : L^p(w) \to L^p(w)\|.$$

The following two-weight result was proved in Theorem 3.1 of [12]. We recall first that the adjoints of Π_b, Π_b^*, and Γ_b as $L^p(\mu) \to L^p(\lambda)$ operators are Π_b^*, Π_b, and Γ_b as $L^q(\lambda') \to L^q(\mu')$ operators, respectively.

THEOREM 2.2. Let $\mu, \lambda \in A_p$ for some $1 < p < \infty$, $v = \mu^{1/p} \lambda^{-1/p}$, and suppose $b \in \mathrm{BMO}_{\mathcal{D}}^2(v)$ for a fixed dyadic grid $\mathcal{D}$ on $\mathbb{R}^n$. Then:

(2.6)
$$\|\Pi_b\|_{L^p(\mu;\lambda)} = \|\Pi_b^*\|_{L^q(\lambda';\mu')} \lesssim c \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(v)},$$

(2.7)
$$\|\Pi_b^*\|_{L^p(\mu;\lambda)} = \|\Pi_b\|_{L^q(\lambda';\mu')} \lesssim c \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(v)},$$

(2.8)
$$\|\Gamma_b\|_{L^p(\mu;\lambda)} = \|\Gamma_b\|_{L^q(\lambda';\mu')} \lesssim c \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(v)},$$

where, in each case, c denotes a constant depending on μ, λ, and p.

2.5. DYADIC SHIFTS. Let i, j be non-negative integers and $\mathcal{D}$ a dyadic grid on $\mathbb{R}^n$. A dyadic shift operator with parameters (i, j) is an operator of the form

(2.9)
$$\mathbb{S}_{\mathcal{D}}^{ij} f := \sum_{R \in \mathcal{D}} \sum_{P \in R_{(i)}, Q \in R_{(j)}} \sum_{\epsilon, \eta \in \{0,1\}^n} a_{PQR}^{\epsilon\eta} \widehat{f}(P, \epsilon) h_Q^\eta,$$

where $a_{PQR}^{\epsilon\eta}$ are coefficients with $a_{PQR}^{\epsilon\eta} \leqslant |R|^{-1} \sqrt{|P||Q|}$. The shift is said to be cancellative if all Haar functions in its definition are cancellative, that is $a_{PQR}^{\epsilon\eta} = 0$ whenever $\epsilon = 1$ or $\eta = 1$. Otherwise, it is called non-cancellative.

The following weighted inequality for dyadic shifts, which can be found in [15], [17], [20], will be extremely useful.

THEOREM 2.3. *Let $\mathbb{S}_{\mathcal{D}}^{ij}$ be a dyadic shift operator. Then for any weight $w \in A_2$,*

$$\|\mathbb{S}_{\mathcal{D}}^{ij}\|_{L^2(w)} \lesssim \kappa_{ij}[w]_2, \tag{2.10}$$

where $\kappa_{ij} := \max\{i, j, 1\}$ is the complexity of the shift.

As a first application of this result, we observe that all paraproducts in Definition 2.1 can be expressed in terms of dyadic shifts with parameters $(0, 0)$, that is, shifts of the form

$$\mathbb{S}^{00} f = \sum_{Q \in \mathcal{D}; \epsilon, \eta \in \{0,1\}^n} a_Q^{\epsilon\eta} \widehat{f}(Q, \epsilon) h_Q^{\eta} \quad |a_Q^{\epsilon\eta}| \leqslant 1.$$

For instance, we may write $\Pi_b = \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} \mathbb{S}^{00}$, where $a_Q^{\epsilon\eta} = 0$ if $\epsilon \neq 1$ or $\eta = 1$, and $a_Q^{\epsilon\eta} = \widehat{b}(Q, \eta)|Q|^{-1/2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{-1}$ if $\epsilon = 1$ and $\eta \neq 1$. Similar expressions can be obtained for the other two paraproducts. Then, if P_b is any one of Π_b, Π_b^*, or Γ_b, it follows from (2.10) that

$$\|P_b\|_{L^2(w)} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}[w]_{A_2}, \tag{2.11}$$

for any $w \in A_2$. These one-weight inequalities for paraproducts were obtained in [1] for the one-dimensional case $n = 1$, and, using the Wilson–Haar basis, in [5] for $n \geqslant 1$.

We can also use dyadic shifts to recover the one-weight bound for the martingale transform:

$$T_\sigma f := \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \sigma_{Q,\epsilon} \widehat{f}(Q, \epsilon) h_Q^\epsilon,$$

where $|\sigma_{Q,\epsilon}| \leqslant 1$ for all $Q \in \mathcal{D}$ and $\epsilon \neq 1$. For $w \in A_2$, all martingale transforms T_σ are uniformly bounded on $L^2(w)$. In particular, there is a universal constant C such that

$$\|T_\sigma\|_{L^2(w)} \leqslant C[w]_{A_2}, \tag{2.12}$$

for all σ. This result, obtained in [21] for the one-dimensional case, trivially follows from the observation that $T_\sigma = \mathbb{S}^{00}$, where $a_Q^{\epsilon\eta}$ is defined to be $\sigma_{Q,\epsilon}$ if $\epsilon = \eta \neq 1$ and 0 otherwise.

The following simple consequence of this fact will come in handy later.

PROPOSITION 2.4. *Let $w \in A_2$, $b \in \mathrm{BMO}_{\mathcal{D}}^2(w)$, and T_σ be a martingale transform. Then $T_\sigma b \in \mathrm{BMO}_{\mathcal{D}}^2(w)$, with*

$$\|T_\sigma b\|_{\mathrm{BMO}_{\mathcal{D}}^2(w)} \lesssim [w]_{A_2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(w)}. \tag{2.13}$$

Proof. It is easy to observe that $\mathbb{1}_Q(T_\sigma b - \langle T_\sigma b \rangle_Q) = T_\sigma(\mathbb{1}_Q(b - \langle b \rangle_Q))$, and so

$$\|T_\sigma b\|_{\mathrm{BMO}_{\mathcal{D}}^2(w)} = \sup_{Q \in \mathcal{D}} \frac{1}{w(Q)^{1/2}} \|T_\sigma(\mathbb{1}_Q(b - \langle b \rangle_Q))\|_{L^2(w^{-1})} \lesssim [w]_{A_2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(w)}. \quad \blacksquare$$

3. PROOF OF THE MAIN RESULT

As in the proof of Theorem 1.1 in [12], the backbone of our proof of Theorem 1.2 is the celebrated Hytönen representation theorem [13], [14], [16], which we state below.

THEOREM 3.1. *Let T be a Calderón–Zygmund operator associated with a δ-standard kernel. Then there exist dyadic shift operators $\mathbb{S}_\omega^{ij}$ with parameters (i, j) for all non-negative integers i, j such that*

$$\langle Tf, g \rangle = c \mathbb{E}_\omega \sum_{i,j=0}^{\infty} 2^{-\kappa_{i,j}(\delta/2)} \langle \mathbb{S}_\omega^{ij} f, g \rangle,$$

for all bounded, compactly supported functions f and g, where c is a constant depending on the dimension n and on T. Here all $\mathbb{S}_\omega^{ij}$ with $(i, j) \neq (0, 0)$ are cancellative, but the shifts $\mathbb{S}_\omega^{00}$ may not be cancellative.

It is easy to see that

(3.1)
$$\langle C_b^k(T)f, g \rangle = c \mathbb{E}_\omega \sum_{i,j=0}^{\infty} 2^{-\kappa_{i,j}(\delta/2)} \langle C_b^k(\mathbb{S}_\omega^{ij})f, g \rangle,$$

for all integers $k \geqslant 1$. Thus it suffices to show that $C_b^k(\mathbb{S}_\omega^{ij})$ are uniformly bounded, regardless of ω, with bounds that depend at most polynomially on κ_{ij}. Since our arguments will be independent of choice of ω, we fix a dyadic grid $\mathcal{D}$ and suppress the ω subscript in what follows. We claim the following.

THEOREM 3.2. *Let $\mu, \lambda \in A_p$ for some $1 < p < \infty$ and $b \in \mathrm{BMO}_\mathcal{D}^2(\nu) \cap \mathrm{BMO}_\mathcal{D}^2$, where $\nu = \mu^{1/p} \lambda^{-1/p}$. For any pair (i, j) of non-negative integers, let $\mathbb{S}^{ij} := \mathbb{S}_\mathcal{D}^{ij}$ be a dyadic shift as in the Hytönen representation theorem. Then for any integer $k \geqslant 1$,*

(3.2)
$$\|C_b^k(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)} \leqslant c \kappa_{ij}^k \|b\|_{\mathrm{BMO}_\mathcal{D}^2}^{k-1} \|b\|_{\mathrm{BMO}_\mathcal{D}^2(\nu)},$$

where c is a constant depending on n, p, μ, λ, and k. In particular, if $\mu = \lambda = w \in A_2$,

(3.3)
$$\|C_b^k(\mathbb{S}^{ij})\|_{L^2(w)} \leqslant c \kappa_{ij}^k \|b\|_{\mathrm{BMO}_\mathcal{D}^2}^k [w]_{A_2}^{k+1},$$

where c is a constant depending on n and k.

Then for all ω,

$$\|C_b^k(\mathbb{S}_\omega^{ij})\|_{L^p(\mu;\lambda)} \leqslant t c \kappa_{ij}^k \|b\|_{\mathrm{BMO}}^{k-1} \|b\|_{\mathrm{BMO}(\nu)}, \text{ and } \|C_b^k(\mathbb{S}_\omega^{ij})\|_{L^2(w)} \leqslant c \kappa_{ij}^k \|b\|_{\mathrm{BMO}}^k [w]_{A_2}^{k+1},$$

where we used the equivalence of $\mathrm{BMO}(\nu)$ and $\mathrm{BMO}^2(\nu)$ norms. Then (3.1) gives that

$$\|C_b^k(T)\|_{L^p(\mu;\lambda)} \leqslant c \|b\|_{\mathrm{BMO}}^{k-1} \|b\|_{\mathrm{BMO}(\nu)} \sum_{i,j=0}^{\infty} 2^{-\kappa_{i,j}(\delta/2)} \kappa_{ij}^k \leqslant c' \|b\|_{\mathrm{BMO}}^{k-1} \|b\|_{\mathrm{BMO}(\nu)}.$$

The one-weight result in Theorem 1.2 follows similarly. The rest of the paper is dedicated to proving Theorem 3.2.

4. THE COMMUTATOR $[b, \mathbb{S}^{ij}]$ REVISITED

Recall that the product of two functions can be formally decomposed in terms of the paraproducts in Definition 2.1 as $bf = \mathfrak{P}_b f + \Pi_f b$, where

$$\mathfrak{P}_b := \Pi_b + \Pi_b^* + \Gamma_b.$$

Consequently, the commutator $[b, T]$ with an operator T can be expressed as

$$C_b^1(T)f = [b, T]f = [\mathfrak{P}_b, T]f + (\Pi_{Tf}b - T\Pi_f b).$$

Proving inequalities for $[b, T]$ via dyadic methods usually involves proving some appropriate bounds for the paraproducts — from which the boundedness of the first term $[\mathfrak{P}_b, T]$ usually follows — and then treating the "remainder term" $\mathcal{R}_1 f = \Pi_{Tf} b - T\Pi_f b$ separately.

Now, remark that if we consider the second iteration

$$C_b^2(T)f = [b, C_b^1(T)]f = [\mathfrak{P}_b, C_b^1(T)]f + (\Pi_{C_b^1(T)f}b - C_b^1(T)\Pi_f b),$$

we will encounter the term $\Pi_{\mathcal{R}_1 f} b - \mathcal{R}_1 \Pi_f b$. We can already see that more compact notation for repeatedly performing the operation $T \mapsto \Pi_{T.}b - T\Pi_. b$ would be useful.

DEFINITION 4.1. Given an operator T on some function space and a function b, define the operator $\Theta_b(T)$ by

$$\Theta_b(T)f := \Pi_{Tf}b - T\Pi_f b.$$

More generally, let $\Theta_b^0(T) := T$, and

$$\Theta_b^k(T) := \Theta_b(\Theta_b^{k-1}(T)),$$

for all integers $k \geqslant 1$.

Using this notation, for an operator T,

$$(4.1) \qquad\qquad [b, T] = [\mathfrak{P}_b, T] + \Theta_b(T).$$

In particular,

$$(4.2) \qquad\qquad C_b^1(\mathbb{S}^{ij}) = [\mathfrak{P}_b, \mathbb{S}^{ij}] + \Theta_b(\mathbb{S}^{ij}),$$

so

$$\|C_b^1(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)} \lesssim \|\mathfrak{P}_b\|_{L^p(\mu;\lambda)}(\|\mathbb{S}^{ij}\|_{L^p(\mu)} + \|\mathbb{S}^{ij}\|_{L^p(\lambda)}) + \|\Theta_b(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)}$$

$$(4.3) \qquad\qquad\qquad \lesssim \kappa_{ij} C(\mu, \lambda, p)\|b\|_{\mathrm{BMO}_\mathcal{D}^2(\nu)} + \|\Theta_b(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)},$$

where the first term is bounded using Theorem 2.2 and (2.10). Letting $\mu = \lambda = w \in A_2$ in (4.3) and using the one-weight bounds (2.11) for the paraproducts we have

$$\|C_b^1(\mathbb{S}^{ij})\|_{L^2(w)} \leqslant 2\|\mathfrak{P}_b\|_{L^2(w)}\|\mathbb{S}^{ij}\|_{L^2(w)} + \|\Theta_b(\mathbb{S}^{ij})\|_{L^2(w)}$$
$$\lesssim \kappa_{ij}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}[w]_{A_2}^2 + \|\Theta_b(\mathbb{S}^{ij})\|_{L^2(w)}.$$

So it remains to bound the remainder term. And, based on analysis that will come later, we in fact need to control certain iterates of the remainder term, leading to the following claim.

PROPOSITION 4.2. *Under the same assumptions as in Theorem 3.2, for all integers $k \geqslant 1$,*

$$(4.4) \qquad \|\Theta_b^k(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)} \leqslant c\kappa_{ij}^k\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{k-1}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)},$$

where c is a constant depending on n, p, μ, λ, and k, and

$$(4.5) \qquad \|\Theta_b^k(\mathbb{S}^{ij})\|_{L^2(w)} \leqslant c\kappa_{ij}^k\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^k[w]_{A_2}^{k+1},$$

where c is a constant depending on n and k.

Obviously, letting $k = 1$ in this proposition yields the results in Theorem 3.2 with $k = 1$. The first result (4.4) was proved for $k = 1$ in [12]. In this section we revisit this proof in order to obtain the one-weight result. The latter will follow directly from the two-weight proof in the case of cancellative shifts $\mathbb{S}^{ij}$ with $(i,j) \neq (0,0)$, but the case $(i,j) = (0,0)$ will require some care. However, as we shall see in the next section, the tools we introduce here lay most of the groundwork for the iterated commutators.

4.1. THE CANCELLATIVE SHIFTS. We showed in [12] that

$$(4.6) \qquad \Theta_b(\mathbb{S}^{ij}) = \sum_{R \in \mathcal{D}; \epsilon,\eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon\eta} \widehat{f}(P,\epsilon)(\langle b \rangle_Q - \langle b \rangle_P) h_Q^{\eta},$$

whenever $(i,j) \neq (0,0)$ and the dyadic shift is cancellative. Then, assuming $i \leqslant j$, we expressed $\Theta_b(\mathbb{S}^{ij})$ as

$$\Theta_b(\mathbb{S}^{ij}) = \sum_{l=1}^{j} A_l - \sum_{m=1}^{i} B_m,$$

for some certain operators A_l, B_m. Using the weighted $\mathcal{H}^1$-BMO duality statement in (2.5), we showed that these operators satisfy

$$(4.7) \qquad \|A_l\|_{L^p(\mu;\lambda)} \leqslant [\nu]_{A_2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} 2^{-(n/2)(i+j-l)}\|M\|_{L^q(\lambda')}\|\widetilde{S}_{\mathcal{D}}^{i,j-l}\|_{L^p(\mu)},$$

$$(4.8) \qquad \|B_m\|_{L^p(\mu;\lambda)} \leqslant [\nu]_{A_2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} 2^{-(n/2)(i+j-m)}\|M\|_{L^p(\mu)}\|\widetilde{S}_{\mathcal{D}}^{j,i-m}\|_{L^q(\lambda')},$$

for all integers $1 \leqslant l \leqslant j$ and $1 \leqslant m \leqslant j$, where $\widetilde{S}_{\mathcal{D}}^{i,j}$ is the shifted dyadic square function in (2.3). From here, (4.4) follows easily from (2.1) and (2.4).

Now, if we let $\mu = \lambda = w \in A_2$, (4.7) becomes:

$$\|A_l\|_{L^2(w)} \leqslant \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} 2^{-(n/2)(i+j-l)} \|M\|_{L^2(w^{-1})} \|\widetilde{S}_{\mathcal{D}}^{i,j-l}\|_{L^2(w)}$$

$$\lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} 2^{-(n/2)(i+j-l)} [w^{-1}]_{A_2} 2^{(n/2)(i+j-l)} [w]_{A_2} = \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} [w]_{A_2}^2.$$

Similarly, we obtain $\|B_m\|_{L^2(w)} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} [w]_{A_2}^2$ from (4.8), and (4.5) follows. The proof for $i \geqslant j$ is symmetrical.

4.2. THE CASE $i = j = 0$. As shown in [14], the non-cancellative shift $\mathbb{S}^{00}$ is of the form

(4.9) $$\mathbb{S}^{00} = \mathbb{S}_c^{00} + \Pi_a + \Pi_d^*,$$

where $\mathbb{S}_c^{00}$ is a *cancellative* shift with parameters $(0,0)$, and Π_a, Π_d^* are paraproducts with symbols $a, d \in \mathrm{BMO}_{\mathcal{D}}$ and $\|a\|_{\mathrm{BMO}_{\mathcal{D}}} \leqslant 1$, $\|d\|_{\mathrm{BMO}_{\mathcal{D}}} \leqslant 1$.

In Section 5.2 of [12] we show that $\Theta_b(\mathbb{S}_c^{00}) = 0$ and

(4.10) $$\Theta_b(\Pi_a) = \Pi_a \Pi_b + \Pi_a \Gamma_b + \Lambda_{a,b} - \widetilde{\Lambda}_{a,b},$$

(4.11) $$\Theta_b(\Pi_a^*) = \Lambda_{b,a} - \Pi_b^* \Pi_a^* - \Gamma_b \Pi_a^* - \Pi_b \Pi_a^*,$$

where

(4.12) $$\Lambda_{a,b} f := \sum_{Q \in \mathcal{D}; \epsilon, \eta \neq 1} \widehat{a}(Q, \epsilon) \sum_{R \in \mathcal{D}, R \supset Q} \widehat{b}(R, \eta) \widehat{f}(R, \eta) \frac{1}{|R|} h_Q^{\epsilon},$$

(4.13) and $$\widetilde{\Lambda}_{a,b} := \sum_{Q \in \mathcal{D}; \epsilon, \eta \neq 1} \widehat{a}(Q, \epsilon) \widehat{b}(Q, \eta) \widehat{f}(Q, \eta) \frac{1}{|Q|} h_Q^{\epsilon}.$$

We remark to the reader, that we are using a slightly different definition of $\Lambda_{a,b}$ than in [12]. The $\Lambda_{a,b}$ defined in (4.12) corresponds to $\Lambda_{b,a}^*$ in [12]. However, we shall see later that it is more advantageous for our purposes to work with the definitions above.

We claim the following lemma.

LEMMA 4.3. *Let $\mu, \lambda \in A_p$ for some $1 < p < \infty$. Suppose $a \in \mathrm{BMO}_{\mathcal{D}}^2$ and $b \in \mathrm{BMO}_{\mathcal{D}}^2(\nu)$, where $\mathcal{D}$ is a fixed dyadic grid on $\mathbb{R}^n$ and $\nu = \mu^{1/p} \lambda^{-1/p}$. If T is any one of the operators*

$$\Lambda_{a,b}, \ \widetilde{\Lambda}_{a,b}, \ \Lambda_{b,a}, \ \widetilde{\Lambda}_{b,a}, \ \Theta_b(\Pi_a), \ or \ \Theta_b(\Pi_a^*),$$

then

(4.14) $$\|T\|_{L^p(\mu;\lambda)} \lesssim C(\mu, \lambda, p) \|a\|_{\mathrm{BMO}_{\mathcal{D}}^2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)}.$$

In particular, if $\mu = \lambda = w \in A_2$,

(4.15) $$\|T\|_{L^2(w)} \lesssim \|a\|_{\mathrm{BMO}_{\mathcal{D}}^2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} [w]_{A_2}^2.$$

Proposition 4.2 with $i = j = 0$ and $k = 1$ follows from this immediately, by the assumptions on the BMO norms of a and d. Before we proceed with the proof, we remark that, while (4.14) was proved in some form for the Λ operators in [12], we need a slight modification of that proof in order for it to yield the one-weight result. Roughly speaking, the original proof boils down to $\|S_{\mathcal{D}}\|_{L^p(\mu)}\|M\Pi_a^*\|_{L^q(\lambda')}$, which in the one-weight case gives a factor of $[w]_{A_2}^3$.

Proof of Lemma 4.3. It suffices to prove the result for the Λ operators. For then, from the decomposition in (4.10) we have

$$\|\Theta_b(\Pi_a)\|_{L^p(\mu;\lambda)} \leqslant \|\Pi_a\|_{L^p(\lambda)}\|\Pi_b + \Gamma_b\|_{L^p(\mu;\lambda)} + \|\Lambda_{a,b} - \widetilde{\Lambda}_{a,b}\|_{L^p(\mu;\lambda)}$$
$$\lesssim C(\mu,\lambda,p)\|a\|_{\text{BMO}_{\mathcal{D}}^2}\|b\|_{\text{BMO}_{\mathcal{D}}^2(\nu)},$$

where we used (2.11) for the Π_a term, and Theorem 2.2 for the paraproducts with symbol b. Similarly, we can see from (4.11) that $\Theta_b(\Pi_a^*)$ obeys the same bound. If $\mu = \lambda = w \in A_2$,

$$\|\Theta_b(\Pi_a)\|_{L^2(w)} \leqslant \|\Pi_a\|_{L^2(w)}\|\Pi_b + \Gamma_b\|_{L^2(w)} + \|\Lambda_{a,b} + \widetilde{\Lambda}_{a,b}\|_{L^2(w)}$$
$$\lesssim \|a\|_{\text{BMO}_{\mathcal{D}}^2}\|b\|_{\text{BMO}_{\mathcal{D}}^2}[w]_{A_2}^2,$$

and the same holds for $\Theta_b(\Pi_a^*)$.

Now let us look at $\Lambda_{a,b}$. Let $f \in L^p(\mu)$ and $g \in L^q(\lambda')$. Then

$$\langle \Lambda_{a,b}f, g\rangle = \sum_{Q\in\mathcal{D}, \epsilon\neq 1} \widehat{a}(Q,\epsilon)\widehat{g}(Q,\epsilon) \sum_{R\in\mathcal{D}, R\supset Q; \eta\neq 1} \widehat{b}(R,\eta)\widehat{f}(R,\eta)\frac{1}{|R|}.$$

Let $a_\tau = T_\tau a$ and $b_\sigma = T_\sigma b$, where T_τ and T_σ are martingale transforms with $\tau_{Q,\epsilon} = \pm 1$ and $\sigma_{Q,\epsilon} = \pm 1$ chosen for every pair $(Q \in \mathcal{D}, \epsilon \neq 1)$ such that

$$\tau_{Q,\epsilon}\widehat{a}(Q,\epsilon)\widehat{g}(Q,\epsilon) \geqslant 0 \quad \text{and} \quad \sigma_{Q,\epsilon}\widehat{b}(Q,\epsilon)\widehat{f}(Q,\epsilon) \geqslant 0.$$

Then

$$|\langle \Lambda_{a,b}f, g\rangle| \leqslant \sum_{Q\in\mathcal{D};\epsilon\neq 1} \widehat{a_\tau}(Q,\epsilon)\widehat{g}(Q,\epsilon) \sum_{R\in\mathcal{D}, R\supset Q;\eta\neq 1} \widehat{b_\sigma}(R,\eta)\widehat{f}(R,\eta)\frac{1}{|R|}$$

$$(4.16) \qquad\qquad \leqslant \sum_{Q\in\mathcal{D};\epsilon\neq 1} \widehat{a_\tau}(Q,\epsilon)\widehat{g}(Q,\epsilon)\langle \Pi_{b_\sigma}^* f\rangle_Q,$$

where the last inequality follows from

$$\langle \Pi_{b_\sigma}^* f\rangle_Q = \sum_{P\in\mathcal{D}, P\subsetneq Q;\epsilon\neq 1} \widehat{b_\sigma}(P,\epsilon)\widehat{f}(P,\epsilon)\frac{1}{|Q|} + \sum_{R\in\mathcal{D}, R\supset Q;\eta\neq 1} \widehat{b_\sigma}(R,\eta)\widehat{f}(R,\eta)\frac{1}{|R|}.$$

By the assumptions on σ and τ and the monotone convergence theorem, (4.16) becomes

$$\sum_{Q \in \mathcal{D}, \epsilon \neq 1} \int \widehat{a_\tau}(Q, \epsilon) \widehat{g}(Q, \epsilon) (\Pi_{b_\sigma}^* f)(x) \frac{\mathbb{1}_Q(x)}{|Q|} \, dx$$

$$= \int \sum_{Q \in \mathcal{D}, \epsilon \neq 1} \widehat{a_\tau}(Q, \epsilon) \widehat{g}(Q, \epsilon) (\Pi_{b_\sigma}^* f)(x) \frac{\mathbb{1}_Q(x)}{|Q|} \, dx = \langle \Pi_{a_\tau}^* g, \Pi_{b_\sigma}^* f \rangle.$$

Therefore, by (2.11) and Theorem 2.2,

$$\|\Lambda_{a,b}\|_{L^p(\mu;\lambda)} \lesssim \|\Pi_{a_\tau}^*\|_{L^q(\lambda')} \|\Pi_{b_\sigma}^*\|_{L^p(\mu;\lambda)}$$

(4.17) $$\lesssim C(\mu, \lambda, p) \|a_\tau\|_{\mathrm{BMO}_{\mathcal{D}}^2} \|b_\sigma\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} \lesssim C(\mu, \lambda, p) \|a\|_{\mathrm{BMO}_{\mathcal{D}}^2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)},$$

where the last inequality follows from Proposition 2.4. Letting $\mu = \lambda = w \in A_2$ in (4.17),

$$\|\Lambda_{a,b}\|_{L^2(w)} \lesssim \|a_\tau\|_{\mathrm{BMO}_{\mathcal{D}}^2} [w]_{A_2} \|b_\sigma\|_{\mathrm{BMO}_{\mathcal{D}}^2} [w]_{A_2} \lesssim \|a\|_{\mathrm{BMO}_{\mathcal{D}}^2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} [w]_{A_2}^2,$$

which proves the result for $\Lambda_{a,b}$. An identical argument proves the result for $\widetilde{\Lambda}_{a,b}$. As for $\Lambda_{b,a}$ or $\widetilde{\Lambda}_{b,a}$, the argument follows similarly, with a few modifications:

$$\|\Lambda_{b,a}\|_{L^p(\mu;\lambda)} \lesssim \|\Pi_{a_\tau}^*\|_{L^p(\mu)} \|\Pi_{b_\sigma}^*\|_{L^q(\lambda';\mu')}. \qquad \blacksquare$$

5. THE SECOND ITERATION $[b, [b, \mathbb{S}^{ij}]]$

In this section we take a closer look at what happens when $k = 2$, and develop the rest of the tools we need for the general case. From (4.2) we have:

$$C_b^2(\mathbb{S}^{ij}) = [b, C_b^1(\mathbb{S}^{ij})] = [\mathfrak{P}_b, C_b^1(\mathbb{S}^{ij})] + \Theta_b(C_b^1(\mathbb{S}^{ij})).$$

We show that each term obeys the bounds in Theorem 3.2. The first term can be bounded using (2.11) and Theorem 3.2 with $k = 1$. In order to analyze the second term, we look at some simple properties of Θ_b that will be useful. Suppose S and T are some operators. Obviously Θ_b is linear, that is $\Theta_b(S + cT) = \Theta_b(S) + c\Theta_b(T)$. Moreover,

(5.1) $$\Theta_b(ST) = \Theta_b(S)T + S\Theta_b(T).$$

To see this, note that we can write $\Theta_b(ST)f = \Pi_{S(Tf)}b - S\Pi_{Tf}b + S\Pi_{Tf}b - ST\Pi_f b$. In turn, this yields

(5.2) $$\Theta_b([S, T]) = [\Theta_b(S), T] + [S, \Theta_b(T)].$$

Then

(5.3) $$\Theta_b(C_b^1(\mathbb{S}^{ij})) = [\Theta_b(\mathfrak{P}_b), \mathbb{S}^{ij}] + [\mathfrak{P}_b, \Theta_b(\mathbb{S}^{ij})] + \Theta_b^2(\mathbb{S}^{ij}).$$

The second term in the expression is easily controlled using Proposition 4.2 with $k = 1$. For the first term, remark that

(5.4) $$\Theta_b(\Gamma_b) = 0,$$

which is easily seen by verifying that

$$\Pi_{\Gamma_b f} b = \Gamma_b \Pi_f b = \sum_{R \in \mathcal{D}} \sum_{\epsilon, \eta \neq 1; \epsilon \neq \eta} \widehat{b}(R, \eta) \widehat{f}(R, \eta) \langle b \rangle_R \frac{1}{\sqrt{|R|}} h_R^{\epsilon + \eta}.$$

So $\Theta_b(\mathfrak{P}_b) = \Theta_b(\Pi_b) + \Theta_b(\Pi_b^*)$, and both these terms can be bounded using Lemma 4.3 with $a = b$. To analyze $\Theta_b^2(\mathbb{S}^{ij})$ and prove Proposition 4.2 for $k = 2$, we again need to look at the cancellative and non-cancellative cases separately.

5.1. THE CANCELLATIVE CASE. Using the expression for $\Theta_b(\mathbb{S}^{ij})$ in (4.6), we find that

$$\Pi_{\Theta_b(\mathbb{S}^{ij})f} b = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon \eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P) \langle b \rangle_Q h_Q^{\eta}, \quad \text{and}$$

$$\Theta_b(\mathbb{S}^{ij}) \Pi_f b = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon \eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P) \langle b \rangle_P h_Q^{\eta}.$$

Therefore

$$\Theta_b^2(\mathbb{S}^{ij}) f = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon \eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)^2 h_Q^{\eta}.$$

Now consider the operator

$$U_{(j)} f := \sum_{Q \in \mathcal{D}; \eta \neq 1} (\langle b \rangle_Q - \langle b \rangle_{Q^{(j)}}) \widehat{f}(Q, \eta) h_Q^{\eta},$$

for a non-negative integer j. Then

$$U_{(j)} \Theta_b(\mathbb{S}^{ij}) f = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon \eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)(\langle b \rangle_Q - \langle b \rangle_R) h_Q^{\eta}, \quad \text{and}$$

$$\Theta_b(\mathbb{S}^{ij}) U_{(i)} f = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon \eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)(\langle b \rangle_P - \langle b \rangle_R) h_Q^{\eta}.$$

So

$$\Theta_b^2(\mathbb{S}^{ij}) = U_{(j)} \Theta_b(\mathbb{S}^{ij}) - \Theta_b(\mathbb{S}^{ij}) U_{(i)}.$$

We claim that for any A_2 weight w:

(5.5) $$\|U_{(j)}\|_{L^2(w)} \lesssim j[w]_{A_2} \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}.$$

To see this, remark that

$$|\langle b \rangle_Q - \langle b \rangle_{Q^{(j)}}| \leqslant j 2^n \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2},$$

so $U_{(j)}$ can be expressed as

$$U_{(j)} = j 2^n \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2} T_\sigma,$$

where T_σ is a martingale transform. Then (5.5) follows from (2.12). Finally, this and Proposition 4.2 with $k = 1$ give that

$$\|\Theta_b^2(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)} \lesssim \|\Theta_b(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)}(\|U_{(i)}\|_{L^p(\mu)} + \|U_{(j)}\|_{L^p(\lambda)})$$
$$\lesssim \kappa_{ij}^2 C(\mu, \lambda, p)\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2},$$

and, in the one-weight case,

$$\|\Theta_b^2(\mathbb{S}^{ij})\|_{L^2(w)} \lesssim \kappa_{ij}^2 \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^2 [w]_{A_2}^3.$$

5.2. THE CASE $i = j = 0$. From (4.9) we have:

$$\Theta_b^2(\mathbb{S}^{00}) = \Theta_b^2(\Pi_a) + \Theta_b^2(\Pi_d^*).$$

Using the expression in (4.10) and the properties of Θ_b in (5.1) and (5.4),

(5.6) $\Theta_b^2(\Pi_a) = \Theta_b(\Pi_a)\Pi_b + \Pi_a\Theta_b(\Pi_b) + \Theta_b(\Pi_a)\Gamma_b + \Theta_b(\Lambda_{a,b}) - \Theta_b(\widetilde{\Lambda}_{a,b}).$

Lemma 4.3 and the paraproduct norms immediately control the first three terms, showing that their norms as operators $L^p(\mu) \to L^p(\lambda)$ are bounded (up to a constant) by $\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)}$, and their norms as operators $L^2(w) \to L^2(w)$ are bounded (up to a constant) by $\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^2[w]_{A_2}^3$. For the last two terms, we look at an interesting property of the $\Lambda_{a,b}$ operators.

PROPOSITION 5.1. *For all locally integrable functions a, b, c, the operator $\Lambda_{a,b}$ satisfies:*

(5.7) $\Theta_c(\Lambda_{a,b}) = \Pi_a\Lambda_{c,b},$

(5.8) $\Theta_c(\widetilde{\Lambda}_{a,b}) = 0.$

Proof. To prove the first statement, note that

$$\langle \Lambda_{a,b}f \rangle_Q = \sum_{\substack{Q,R\in\mathcal{D}; R\supsetneq Q \\ \eta\neq 1}} \widehat{b}(R,\eta)\widehat{f}(R,\eta)\frac{1}{|R|} \sum_{\substack{P\in\mathcal{D}; Q\subsetneq P\subset R \\ \epsilon\neq 1}} \widehat{a}(P,\epsilon)h_P^\epsilon(Q)$$

(5.9) $$= \sum_{\substack{Q,R\in\mathcal{D}; R\supsetneq Q \\ \eta\neq 1}} \widehat{b}(R,\eta)\widehat{f}(R,\eta)\frac{1}{|R|}(\langle a \rangle_Q - \langle a \rangle_R).$$

A quick calculation shows that

$$\Theta_c(\Lambda_{a,b})f = \sum_{Q\in\mathcal{D},\epsilon\neq 1} \widehat{a}(Q,\epsilon)\Big[\sum_{\substack{Q,R\in\mathcal{D}; R\supset Q \\ \eta\neq 1}} \widehat{b}(R,\eta)\widehat{f}(R,\eta)\frac{1}{|R|}(\langle c \rangle_Q - \langle c \rangle_R)\Big]h_Q^\epsilon.$$

From (5.9), we recognize the term in parentheses as $\langle \Lambda_{c,b}f \rangle_Q$, and so

$$\Theta_c(\Lambda_{a,b})f = \sum_{Q\in\mathcal{D},\epsilon\neq 1} \widehat{a}(Q,\epsilon)\langle \Lambda_{c,b}f \rangle_Q h_Q^\epsilon = \Pi_a\Lambda_{c,b}f.$$

The second statement follows by

$$\Pi_{\widetilde{\Lambda}_{a,b}f}c = \widetilde{\Lambda}_{a,b}\Pi_f c = \sum_{Q\in\mathcal{D};\,\epsilon,\eta\neq 1}\widehat{a}(Q,\epsilon)\widehat{b}(Q,\eta)\widehat{f}(Q,\eta)\frac{1}{|Q|}\langle c\rangle_Q h_Q^\epsilon. \qquad \blacksquare$$

Returning to (5.6), we can now see that the last two terms in the expression become simply $\Pi_a\Lambda_{b,b}$, which is controlled exactly as the other terms. The result for $\Theta_b^2(\Pi_a^*)$ follows similarly, after noting that $\Theta_b(\Lambda_{b,a}) = \Pi_b\Lambda_{b,a}$. Finally, recall the assumptions on the BMO norms of a and d in (4.9) and see that the results in this section prove Proposition 4.2 for $k = 2$ and $i = j = 0$.

6. THE GENERAL CASE OF HIGHER ITERATIONS

In this section we prove Theorem 3.2. A closer look at recursively expanding the formula

$$C_b^{m+1}(T) = [\mathfrak{P}_b, C_b^m(T)] + \Theta_b(C_b^m(T)),$$

for some operator T, shows that, in order to control $C_b^{m+1}(T)$, we need not only bound the previous iterations $C_b^{m+1-k}(T)$, but really

$$C_b^{m-k}(T),\ \Theta_b(C_b^{m-k}(T)),\ \Theta_b^2(C_b^{m-k}(T)),\ldots,\ \Theta_b^k(C_b^{m-k}(T)),$$

for every $0 \leqslant k \leqslant m - 1$. So, it makes sense to instead prove the following more general statement.

THEOREM 6.1. *Under the same assumptions as Theorem* 3.2, *for any integer* $k \geqslant 1$,

$$(6.1) \qquad \|\Theta_b^M(C_b^k(\mathbb{S}^{ij}))\|_{L^p(\mu;\lambda)} \leqslant c\kappa_{ij}^{M+k}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{M+k-1}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} \quad \textit{for all } M \geqslant 0,$$

where c is a constant depending on n, μ, λ, p, M, and k. In particular, if $\mu = \lambda = w \in A_2$,

$$(6.2) \qquad \|\Theta_b^M(C_b^k(\mathbb{S}^{ij}))\|_{L^2(w)} \leqslant c\kappa_{ij}^{M+k}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{M+k}[w]_{A_2}^{M+k+1} \quad \textit{for all } M \geqslant 0,$$

where c is a constant depending on n, M, and k.

Theorem 3.2 will then follow as a special case of the above result, with $M = 0$. We begin by completing the proof of Proposition 4.2.

Proof of Proposition 4.2. So far, this result has been proved for $k = 1$ and $k = 2$. In case $(i,j) \neq (0,0)$, we generalize the argument in Subsection 5.1. We claim that for all $k \geqslant 2$ and $(i,j) \neq (0,0)$,

$$(6.3) \qquad \Theta_b^k(\mathbb{S}^{ij})f = \sum_{R\in\mathcal{D};\,\epsilon,\eta\neq 1}\ \sum_{P\in R_{(i)},\,Q\in R_{(j)}}a_{PQR}^{\epsilon\eta}\widehat{f}(P,\epsilon)(\langle b\rangle_Q - \langle b\rangle_P)^k h_Q^\eta$$

$$(6.4) \qquad = U_{(j)}\Theta_b^{k-1}(\mathbb{S}^{ij}) - \Theta_b^{k-1}(\mathbb{S}^{ij})U_{(i)}.$$

Then, assuming Proposition 4.2 holds for some $k \geqslant 1$, the result for $k+1$ follows from (5.5). To see this, assume (6.3) holds for some $k \geqslant 2$. Then

$$\Pi_{\Theta_b^k(\mathbb{S}^{ij})f}b = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon\eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)^k \langle b \rangle_Q h_Q^{\eta}, \quad \text{and}$$

$$\Theta_b^k(\mathbb{S}^{ij})\Pi_f b = \sum_{R \in \mathcal{D}; \epsilon, \eta \neq 1} \sum_{P \in R_{(i)}, Q \in R_{(j)}} a_{PQR}^{\epsilon\eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)^k \langle b \rangle_P h_Q^{\eta}.$$

Since $\Theta_b^{k+1}(\mathbb{S}^{ij})f = \Pi_{\Theta_b^k(\mathbb{S}^{ij})f}b - \Theta_b^k(\mathbb{S}^{ij})\Pi_f b$, we see that (6.3) holds for $k+1$. Similarly,

$$U_{(j)}\Theta_b^k(\mathbb{S}^{ij})f = \sum_{\substack{R \in \mathcal{D}; \epsilon, \eta \neq 1 \\ P \in R_{(i)}, Q \in R_{(j)}}} a_{PQR}^{\epsilon\eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)^k (\langle b \rangle_Q - \langle b \rangle_R) h_Q^{\eta}, \quad \text{and}$$

$$\Theta_b^k(\mathbb{S}^{ij})U_{(i)}f = \sum_{\substack{R \in \mathcal{D}; \epsilon, \eta \neq 1 \\ P \in R_{(i)}, Q \in R_{(j)}}} a_{PQR}^{\epsilon\eta} \widehat{f}(P, \epsilon)(\langle b \rangle_Q - \langle b \rangle_P)^k (\langle b \rangle_P - \langle b \rangle_R) h_Q^{\eta},$$

from which (6.4) with $k+1$ follows.

For the case $i = j = 0$,

$$\Theta_b^k(\mathbb{S}^{00}) = \Theta_b^k(\Pi_a) + \Theta_b^k(\Pi_d^*),$$

with $\|a\|_{\text{BMO}_{\mathcal{D}}^2} \lesssim 1$ and $\|d\|_{\text{BMO}_{\mathcal{D}}^2} \lesssim 1$. Proposition 4.2 for $i = j = 0$ therefore follows trivially from the next result. ∎

PROPOSITION 6.2. *Under the same assumptions as Theorem 3.2, let P_a denote either one of the operators Π_a and Π_a^*, and Λ denote either one of the operators $\Lambda_{a,b}$ or $\Lambda_{b,a}$. Then for any integer $k \geqslant 1$,*

$$(6.5) \qquad \|\Theta_b^k(P_a)\|_{L^p(\mu;\lambda)} \leqslant c\|a\|_{\text{BMO}_{\mathcal{D}}^2} \|b\|_{\text{BMO}_{\mathcal{D}}^2}^{k-1} \|b\|_{\text{BMO}_{\mathcal{D}}^2(\nu)},$$

$$(6.6) \qquad \|\Theta_b^k(\Lambda)\|_{L^p(\mu;\lambda)} \leqslant c\|a\|_{\text{BMO}_{\mathcal{D}}^2} \|b\|_{\text{BMO}_{\mathcal{D}}^2}^{k} \|b\|_{\text{BMO}_{\mathcal{D}}^2(\nu)},$$

where c is a constant depending on n, p, μ, λ, and k. In particular, if $\mu = \lambda = w \in A_2$,

$$(6.7) \qquad \|\Theta_b^k(P_a)\|_{L^2(w)} \leqslant c\|a\|_{\text{BMO}_{\mathcal{D}}^2} \|b\|_{\text{BMO}_{\mathcal{D}}^2}^{k} [w]_{A_2}^{k+1},$$

$$(6.8) \qquad \|\Theta_b^k(\Lambda)\|_{L^2(w)} \leqslant c\|a\|_{\text{BMO}_{\mathcal{D}}^2} \|b\|_{\text{BMO}_{\mathcal{D}}^2}^{k+1} [w]_{A_2}^{k+2},$$

where c is a constant depending on n and k.

Proof. This result with $k = 1$ was proved in Lemma 4.3 for $\Theta_b(P_a)$, and in Subsection 5.2 for $\Theta_b(\Lambda)$. We proceed by (strong) induction. Fix $m \geqslant 1$ and suppose Proposition 6.2 holds for all $1 \leqslant k \leqslant m$. We show that it then holds for $k = m + 1$.

Let us look at the case $P_a = \Pi_a$:

$$\Theta_b^{m+1}(\Pi_a) = \Theta_b^m(\Theta_b(\Pi_a)) = \Theta_b^m(\Pi_a(\Pi_b + \Gamma_b)) + \Theta_b^m(\Lambda_{a,b}).$$

Remark that the last term is already controlled by the induction assumption. To analyze the first term, we use the binomial formula

$$\Theta_b^m(ST) = \sum_{k=0}^{m} \binom{m}{k} \Theta_b^{m-k}(S)\Theta_b^k(T), \tag{6.9}$$

which follows from (5.1) by a simple induction argument. Then

$$\|\Theta_b^m(\Pi_a(\Pi_b + \Gamma_b))\|_{L^p(\mu;\lambda)}$$
$$\leqslant \sum_{k=0}^{m} \binom{m}{k} \|\Theta_b^{m-k}(\Pi_a)\|_{L^p(\lambda)} \|\Theta_b^k(\Pi_b + \Gamma_b)\|_{L^p(\mu;\lambda)}. \tag{6.10}$$

The statement

$$\|\Theta_b^{m-k}(\Pi_a)\|_{L^2(w)} \lesssim C(m-k)\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{m-k}[w]_{A_2}^{m-k+1} \quad 0 \leqslant k \leqslant m, \tag{6.11}$$

for all $w \in A_2$, follows from the induction assumption on P_a for $0 \leqslant k \leqslant m-1$, and from (2.11) for $k = m$. Then, by A_p extrapolation,

$$\|\Theta_b^{m-k}(\Pi_a)\|_{L^p(\lambda)} \lesssim C(\lambda, p, m-k)\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{m-k} \quad 0 \leqslant k \leqslant m. \tag{6.12}$$

On the other hand, noting that $\Theta_b^0(\Pi_b + \Gamma_b) = \Pi_b + \Gamma_b$ and $\Theta_b^k(\Pi_b + \Gamma_b) = \Theta_b^k(\Pi_b)$ for $1 \leqslant k \leqslant m$, we have

$$\|\Theta_b^k(\Pi_b + \Gamma_b)\|_{L^p(\mu;\lambda)} \lesssim C(\mu, \lambda, p, k)\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^k \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} \quad 0 \leqslant k \leqslant m,$$

which follows from Theorem 2.2 for $k = 0$, and from the induction assumption on P_a with $a = b$ for $1 \leqslant k \leqslant m$. Similarly,

$$\|\Theta_b^k(\Pi_b + \Gamma_b)\|_{L^2(w)} \lesssim C(k)\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{k+1}[w]_{A_2}^{k+1} \quad 0 \leqslant k \leqslant m.$$

From these estimates and (6.10), we obtain

$$\|\Theta_b^m(\Pi_a(\Pi_b + \Gamma_b))\|_{L^p(\mu;\lambda)} \lesssim C(\mu, \lambda, p, m)\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^m \|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)}, \quad \text{and}$$

$$\|\Theta_b^m(\Pi_a(\Pi_b + \Gamma_b))\|_{L^2(w)} \lesssim C(m)\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{m+1}[w]_{A_2}^{m+2},$$

which proves the result for $k = m + 1$ and $P_a = \Pi_a$. The case $P_a = \Pi_a^*$ follows similarly.

Now suppose $\Lambda = \Lambda_{a,b}$. Then, by (5.7),

$$\Theta_b^{m+1}(\Lambda_{a,b}) = \Theta_b^m(\Theta_b(\Lambda_{a,b})) = \Theta_b^m(\Pi_a \Lambda_{b,b}).$$

Using the binomial formula again,

$$\|\Theta_b^{m+1}(\Lambda_{a,b})\|_{L^p(\mu;\lambda)} \leqslant \sum_{k=0}^{m} \binom{m}{k} \|\Theta_b^{m-k}(\Pi_a)\|_{L^p(\lambda)} \|\Theta_b^k(\Lambda_{b,b})\|_{L^p(\mu;\lambda)}. \tag{6.13}$$

The statements

$$\|\Theta_b^k(\Lambda_{b,b})\|_{L^p(\mu;\lambda)} \lesssim C(\mu,\lambda,p,k)\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{k+1}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)} \quad 0 \leqslant k \leqslant m,$$

$$\|\Theta_b^k(\Lambda_{b,b})\|_{L^2(w)} \lesssim C(k)\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{k+2}[w]_{\mathrm{BMO}_{\mathcal{D}}^2}^{k+2} \quad 0 \leqslant k \leqslant m,$$

follow from the induction assumption on Λ with $a = b$ for $1 \leqslant k \leqslant m$, and from Lemma 4.3 for $k = 0$. Combining these with (6.11) and (6.12), we have from (6.13),

$$\|\Theta_b^{m+1}(\Lambda_{a,b})\|_{L^p(\mu;\lambda)} \lesssim C(\mu,\lambda,p,m+1)\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{m+1}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2(\nu)},$$

$$\|\Theta_b^{m+1}(\Lambda_{a,b})\|_{L^2(w)} \lesssim C(m+1)\|a\|_{\mathrm{BMO}_{\mathcal{D}}^2}\|b\|_{\mathrm{BMO}_{\mathcal{D}}^2}^{m+2}[w]_{A_2}^{m+3},$$

which proves the result for $k = m + 1$ and $\Lambda = \Lambda_{a,b}$. Since the case $\Lambda = \Lambda_{b,a}$ follows similarly, Proposition 6.2 is proved. ∎

We now have all the tools needed to prove Theorem 6.1.

Proof of Theorem 6.1. We prove the result for $k = 1$, so we look at

$$\Theta_b^M(C_b^1(\mathbb{S}^{ij})) = \Theta_b^M([\mathfrak{P}_b, \mathbb{S}^{ij}] + \Theta_b(\mathbb{S}^{ij})),$$

for some integer $M \geqslant 0$. At this point, we use another easily deduced binomial formula:

$$(6.14) \qquad \Theta_b^M([S, T]) = \sum_{m=0}^{M} \binom{M}{m} [\Theta_b^{M-m}(S), \Theta_b^m(T)].$$

Then

$$\|\Theta_b^M(C_b^1(\mathbb{S}^{ij}))\|_{L^p(\mu;\lambda)}$$

$$\leqslant \sum_{m=0}^{M} \binom{M}{m} \|\Theta_b^{M-m}(\mathfrak{P}_b)\|_{L^p(\mu;\lambda)}(\|\Theta_b^m(\mathbb{S}^{ij})\|_{L^p(\mu)} + \|\Theta_b^m(\mathbb{S}^{ij})\|_{L^p(\lambda)})$$

$$+ \|\Theta_b^{M+1}(\mathbb{S}^{ij})\|_{L^p(\mu;\lambda)}.$$

Applying Proposition 6.2 with $a = b$ and Proposition 4.2, we obtain the result for $k = 1$. Finally, suppose Theorem 6.1 holds for some $k \geqslant 1$ and let an integer $M \geqslant 0$. Then

$$\Theta_b^M(C_b^{k+1}(\mathbb{S}^{ij})) = \Theta_b^M([\mathfrak{P}_b, C_b^k(\mathbb{S}^{ij})] + \Theta_b(C_b^k(\mathbb{S}^{ij}))),$$

and the result for $k + 1$ again follows from Propositions 6.2 and 4.2. ∎

REFERENCES

[1] O. BEZNOSOVA, Linear bound for the dyadic paraproduct on weighted Lebesgue space $L_2(w)$, *J. Funct. Anal.* **225**(2008), 994–1007.

[2] S. BLOOM, A commutator theorem and weighted BMO, *Trans. Amer. Math. Soc.* **292**(1985), 103–122.

[3] S.M. BUCKLEY, Estimates for operator norms and reverse Jensen's inequalities, *Trans. Amer. Math. Soc.* **340**(1993), 253–272.

[4] M. CHRIST, R. FEFFERMAN, A note on weighted norm inequalities for the Hardy–Littlewood maximal operator, *Proc. Amer. Math. Soc.* **87**(1983), 447–448.

[5] D. CHUNG, Weighted inequalities for multivariable dyadic paraproducts, *Publ. Mat.* **55**(2011), 475–499.

[6] D. CHUNG, M.C. PEREYRA, C. PEREZ, Sharp bounds for general commutators on weighted Lebesgue spaces, *Trans. Amer. Math. Soc.* **364**(2012), 1163–1177.

[7] R.R. COIFMAN, R. ROCHBERG, G. WEISS, Factorization theorems for Hardy spaces in several variables, *Ann. of Math. (2)* **103**(1976), 611–635.

[8] D. CRUZ-URIBE, J.M. MARTELL, C. PÉREZ, Sharp weighted estimates for classical operators, *Adv. in Math.* **229**(2012), 408–441.

[9] L. DALENC, Y. OU, Upper bound for multi-parameter ierated commutators, *Publ. Mat.* **60**(2016), 191–220.

[10] O. DRAGIČEVIĆ, L. GRAFAKOS, M.C. PEREYRA, S. PETERMICHL, Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces, *Publ. Mat.* **49**(2005), 73–91.

[11] I. HOLMES, M.T. LACEY, B.D. WICK, Bloom's inequality: commutators in a two-weight setting, *Arch. Math.* **106**(2016), 53–63.

[12] I. HOLMES, M.T. LACEY, B.D. WICK, Commutators in the two-weight setting, *Math. Ann.* **367**(2017), 51–80.

[13] T. HYTÖNEN, The sharp weighted bound for general Calderón–Zygmund operators, *Ann. of Math. (2)* **175**(2012), 1473–1506.

[14] T. HYTÖNEN, Representation of singular integrals by dyadic operators, and the A_2 theorem, *Expo. Math.* **35**(2017), 166–205.

[15] T. HYTÖNEN, M.T. LACEY, H. MARTIKAINEN, T. ORPONEN, M. REGUERA, E.T. SAWYER, I. URIARTE-TUERO, Weak and strong type estimates for maximal truncations of Calderón–Zygmund operators on A_p weighted spaces, *J. Anal. Math.* **118**(2012), 177–220.

[16] T. HYTÖNEN, C. PÉREZ, S. TREIL, A. VOLBERG, Sharp weighted estimates for dyadic shifts and the A_2 conjecture, *J. Reine Angew. Math.* **687**(2014), 43–86.

[17] M.T. LACEY, On the A_2 inequality for Calderón–Zygmund operators, in *Recent Advances in Harmonic Analysis and Applications*, Springer, New York 2013.

[18] B. MUCKENHOUPT, Weighted norm inequalities for the Hardy maximal function, *Trans. Amer. Math. Soc.* **165**(1972), 207–226.

[19] B. MUCKENHOUPT, R.L. WHEEDEN, Weighted bounded mean oscillation and the Hilbert transform, *Studia Math.* **54**(1975), 221–237.

[20] S. TREIL, Sharp A_2 estimates of Haar shifts via Bellman function, in *Recent Trends in Analysis. Proceedings of the Conference in Honor of Nikolai Nikolski on the Occasion of his 70th Birthday, Bordeaux, France, August 31 – September 2, 2011*, Theta Ser. Adv. Math., vol. 16, The Theta Foundation, Bucharest 2013, pp. 187–208.

[21] J. WITTWER, A sharp estimate on the norm of the martingale transform, *Math. Res. Lett.* **7**(2000), 1–12.

IRINA HOLMES, DEPARTMENT OF MATHEMATICS, MICHIGAN STATE UNIVERSITY, EAST LANSING, 48824, U.S.A.
E-mail address: holmesir@math.msu.edu

BRETT D. WICK, DEPARTMENT OF MATHEMATICS, WASHINGTON UNIVERSITY IN ST. LOUIS, ST. LOUIS, 6310, U.S.A.
E-mail address: wick@math.wustl.edu