Received: 21 January 2021 | Revised: 17 November 2021 | Accepted: 18 November 2021

DOI: 10.1112/jlms.12588

RESEARCH ARTICLE

Journal of the London Mathematical Society

Dyadic product BMO in the Bloom setting

Spyridon Kakaroumpas | **Odí Soler i Gibert**

Institut für Mathematik, Julius-Maximilians-Universität Würzburg, Würzburg, Germany

Correspondence
Odí Soler i Gibert, Institut für Mathematik, Julius-Maximilians-Universität Würzburg, 97074 Würzburg, Germany.
Email: odi.solerigibert@mathematik.uni-wuerzburg.de

Funding information
Alexander von Humboldt Stiftung; ERC, Grant/Award Number: DLV-682402

Abstract

Ó. Blasco and S. Pott showed that the supremum of operator norms over L^2 of all bicommutators (with the same symbol) of one-parameter Haar multipliers dominates the biparameter dyadic product BMO norm of the symbol itself. In the present work we extend this result to the Bloom setting, and to any exponent $1 < p < \infty$. The main tool is a new characterization in terms of paraproducts and two-weight John–Nirenberg inequalities for dyadic product BMO in the Bloom setting. We also extend our results to the whole scale of indexed spaces between little bmo and product BMO in the general multiparameter setting, with the appropriate iterated commutator in each case.

MSC 2020
42B35 (primary)

Contents

1. INTRODUCTION AND MAIN RESULTS . . . 900
2. BACKGROUND AND NOTATION . . . 906
3. EQUIVALENCES FOR DYADIC BLOOM PRODUCT BMO . . . 915
4. ESTIMATES FOR ITERATED COMMUTATORS OF HAAR MULTIPLIERS . . . 922
5. BOUNDS FOR GENERAL COMMUTATORS OF HAAR MULTIPLIERS . . . 929
ACKNOWLEDGEMENT . . . 934
REFERENCES . . . 934

Notation

$\mathbf{1}_E$	characteristic function of a set E;
$\mathrm{d}x$	integration with respect to Lebesgue measure;
$\|E\|$	d-dimensional Lebesgue measure of a measurable set $E \subseteq \mathbb{R}^d$;
$\langle f\rangle_E$	average with respect to Lebesgue measure, $\langle f\rangle_E := \frac{1}{\|E\|}\int_E f(x)\mathrm{d}x$;
L^∞_{c}	space of compactly supported L^∞ functions;
$L^p(w)$	weighted Lebesgue space, $\\|f\\|^p_{L^p(w)} := \int_{\mathbb{R}^d} \|f(x)\|^p w(x)\mathrm{d}x$;
$\langle f, g\rangle$	usual L^2-pairing, $\langle f, g\rangle := \int f(x)\overline{g(x)}\mathrm{d}x$;
$w(E)$	Lebesgue integral of a weight w over a set E, $w(E) := \int_E w(x)\mathrm{d}x$;
p'	Hölder conjugate exponent to p, $1/p + 1/p' = 1$;
$\mathcal{D}$	family of all dyadic intervals in $\mathbb{R}$;
I_-, I_+	respectively, left and right half of an interval $I \in \mathcal{D}$;
$\boldsymbol{\mathcal{D}}$	family of all dyadic rectangles in the product space $\mathbb{R}\times\mathbb{R}$;
$\boldsymbol{\mathcal{D}}(\Omega)$	family of all dyadic rectangles R in the product space $\mathbb{R}\times\mathbb{R}$ that are contained in the set Ω;
$\mathrm{sh}(\mathcal{U})$	'shadow' of a family $\mathcal{U}$ of dyadic rectangles, $\mathrm{sh}(\mathcal{U}) := \bigcup_{R\in\mathcal{U}} R$;
$h^{(0)}_I$, $h^{(1)}_I$	L^2-normalized *cancellative* and *non-cancellative* respectively Haar functions for an interval $I \in \mathcal{D}$, $h^{(0)}_I := \frac{\mathbf{1}_{I_+}-\mathbf{1}_{I_-}}{\sqrt{\|I\|}}$, $h^{(1)}_I := \frac{\mathbf{1}_I}{\sqrt{\|I\|}}$; for simplicity we denote $h_I := h^{(0)}_I$;
b_I	usual Haar coefficient of a function $b \in L^1_{\mathrm{loc}}(\mathbb{R})$, $b_I := \langle b, h_I\rangle$, $I \in \mathcal{D}$;
$h^{(\varepsilon_1\varepsilon_2)}_R$	any of the four L^2-normalized Haar functions for a rectangle $R \in \boldsymbol{\mathcal{D}}$, $h^{(\varepsilon_1\varepsilon_2)}_R := h^{(\varepsilon_1)}_I \otimes h^{(\varepsilon_2)}_J$, where $R = I\times J$ and $\varepsilon_1, \varepsilon_2 \in \{0,1\}$; for simplicity we denote $h_R := h^{(00)}_R$;
b_R	(00) Haar coefficient of a function $b \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, $b_R := \langle b, h_R\rangle$, $R \in \boldsymbol{\mathcal{D}}$;
T^*	formal L^2-adjoint operator to the operator T, $\langle Tf, g\rangle = \langle f, T^*g\rangle$.

The notation $x \lesssim_{a,b,\dots} y$ means $x \leqslant Cy$ with a constant $0 < C < \infty$ depending *only* on the quantities $a, b, \dots$; the notation $x \gtrsim_{a,b,\dots} y$ means $y \lesssim_{a,b,\dots} x$. We use $x \sim_{a,b,\dots} y$ if *both* $x \lesssim_{a,b,\dots} y$ and $x \gtrsim_{a,b,\dots} y$ hold. Sometimes we might omit some of these quantities $a, b, \dots$ from the notation. The context will always make clear when this happens.

1 INTRODUCTION AND MAIN RESULTS

In 1957, Z. Nehari [28] showed that Hankel operators H_b are bounded from the Hardy space $\mathcal{H}^2(\partial\mathbb{D})$ into itself if and only if the symbol b belongs to the space of analytic functions with bounded mean oscillation, or simply $b \in \mathrm{BMOA}$. In fact, Nehari [28] shows an equivalence between the norm of the operator H_b and the BMOA norm of the symbol b. His proof relies on the fact that a function f in the Hardy space $\mathcal{H}^1(\partial\mathbb{D})$ can be factored as $f = g_1g_2$, where both $g_1, g_2 \in \mathcal{H}^2(\partial\mathbb{D})$. Note here that the Hankel operator H_b is essentially equivalent to the commutator $[H, b]$, where H denotes the Hilbert transform on $\partial\mathbb{D}$ and (abusing notation) b stands for multiplication by this function. This allows one to consider not only the real variable version of Nehari's result, but also analogues in $\mathbb{R}^d$ for any d by studying commutators of the form $[R, b]$, where R denotes one of the Riesz transforms in $\mathbb{R}^d$. Nonetheless, observe that the factorization for $\mathcal{H}^1(\partial\mathbb{D})$ has no counterpart in this setting (see also [9]).

In this direction, R. R. Coifman, R. Rochberg and G. Weiss [9] proved in 1976 their celebrated commutator theorem. Namely, for a function b with bounded mean oscillation on $\mathbb{R}^d$, denoted as $b \in \mathrm{BMO}(\mathbb{R}^d)$, they show an equivalence between the BMO norm of b and the sum of the norms of the commutators $[R^{(j)}, b]$, $j = 1, \ldots, d$ with $R^{(j)}$ denoting the jth Riesz transform in $\mathbb{R}^d$, as operators from L^p into itself, for any $1 < p < \infty$. They also show that for a Calderón–Zygmund operator T, the norm of $[T, b]$ as an operator from L^p into itself, $1 < p < \infty$, is bounded above by $\|b\|_{\mathrm{BMO}}$. The argument used by Coifman–Rochberg–Weiss [9] to show the lower bound for the sum of commutators with the Riesz transforms is based on a decomposition of the identity as a linear combination of products of Riesz transforms using spherical harmonics in $\mathbb{R}^d$. This allows one to bound the oscillation of a BMO function by the sum of the norms of $[R^{(j)}, b]$, $j = 1, \ldots, d$.

In a different direction, S. Bloom [5] proved in 1985 an analogue of Nehari's result for weighted spaces. Bloom [5] showed a norm equivalence between a certain weighted BMO space and the operator norm of the commutator $[H, b]$, where H is again the Hilbert transform. To be more precise, for a function $b \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, and for a weight ν on $\mathbb{R}^d$ (that is a locally integrable, almost everywhere (a.e.) positive function on $\mathbb{R}^d$), define the one-weight BMO norm

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

where the supremum ranges over all cubes $Q \subseteq \mathbb{R}^d$ and $\langle b \rangle_Q = \frac{1}{|Q|} \int_Q b(x) \, \mathrm{d}x$ is the unweighted average of b on Q. This weighted BMO space had already been investigated by B. Muckenhoupt and R. L. Wheeden [26]. Bloom [5] showed that for any $1 < p < \infty$, for any A_p weights μ and λ on $\mathbb{R}$ (see Section 2 for the definition of A_p weights), and for the weight $\nu := \mu^{1/p} \lambda^{-1/p}$, which is easily seen to be an A_2 weight, one has the equivalence

$$\|[H, b]\|_{L^p(\mu) \to L^p(\lambda)} \sim \|b\|_{\mathrm{BMO}(\nu)}, \tag{1.1}$$

where H denotes the Hilbert transform and the implied constants depend only on p and the A_p characteristics of μ and λ. Bloom's [5] proof of the estimate $\|[H, b]\|_{L^p(\mu) \to L^p(\lambda)} \gtrsim \|b\|_{\mathrm{BMO}(\nu)}$ uses a different argument to that of Coifman–Rochberg–Weiss, and is based on a careful analysis of the set where $|b - \langle b \rangle_I|$ is not too large with respect to the average oscillation on the interval I, and the set where the former quantity is large using the adjoint commutator.

Much more recently, I. Holmes, M. T. Lacey and B. D. Wick [18] considerably extended Bloom's result, proving that for any Calderón–Zygmund operator T on $\mathbb{R}^d$ and for any A_p weights μ, λ on $\mathbb{R}^d$, $1 < p < \infty$, there holds

$$\|[T, b]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim_{T,d,p} \|b\|_{\mathrm{BMO}(\nu)}, \tag{1.2}$$

and that for the Riesz transforms $R^{(1)}, \ldots, R^{(d)}$ on $\mathbb{R}^d$ there holds in addition

$$\sum_{i=1}^{d} \|[R^{(i)}, b]\|_{L^p(\mu) \to L^p(\lambda)} \gtrsim_{p,d} \|b\|_{\mathrm{BMO}(\nu)}, \tag{1.3}$$

where in both (1.2) and (1.3), $\nu := \mu^{1/p} \lambda^{-1/p}$, and all implied constants depend on p and the A_p characteristics of μ, λ as well. Holmes–Lacey–Wick [18] proved and used in an essential way several new characterizations of the weighted BMO space $\mathrm{BMO}(\nu)$ in the form of two-weight John–Nirenberg inequalities. More precisely, Muckenhoupt–Wheeden [26] had already showed that if

ν is an A_2 weight on $\mathbb{R}^d$, then one has the equivalence

$$\|b\|_{\mathrm{BMO}(\nu)} \sim \sup_Q \left(\frac{1}{\nu(Q)} \int_Q |b(x) - \langle b \rangle_Q|^2 \, \nu^{-1}(x)\mathrm{d}x \right)^{1/2},$$

where the implied constants depend only on d and the A_2 characteristic of ν. Holmes–Lacey–Wick [18] complemented this result, by showing that if μ, λ are A_p weights on $\mathbb{R}^d$, $1 < p < \infty$, and $\nu := \mu^{1/p}\lambda^{-1/p}$, then one has the equivalences

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

and

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

where $\mu' := \mu^{-1/(p-1)}, \lambda' := \lambda^{-1/(p-1)}$, and all implied constants depend only on d, p and the A_p characteristics of μ and λ. Their proof of these equivalences employed a duality result between *dyadic* $\mathrm{BMO}(\nu)$ and a certain dyadic weighted H^1 space that they established in the same work, as well as characterizations of two-weight BMO spaces in terms of two-weight boundedness of certain paraproducts. It should be noted that the results of [18] were very recently extended to the matrix-valued setting by J. Isralowitz, S. Pott and S. Treil [22]. In fact, the authors of [22] proved there several results for the case of completely arbitrary (not necessarily A_p) matrix-valued weights, that are new even if one specializes to the fully scalar setting.

All results mentioned above concern one-parameter spaces. On the other hand, multiparameter (unweighted) BMO spaces were investigated extensively in the seminal papers by S.-Y. A. Chang [8] and R. Fefferman [14] in the late 1970s. These works concern mainly the biparameter product BMO space $\mathrm{BMO}(\mathbb{R} \times \mathbb{R})$ on the product space $\mathbb{R} \times \mathbb{R}$, defined by

$$\|b\|_{\mathrm{BMO}(\mathbb{R}\times\mathbb{R})} := \sup_\Omega \left(\frac{1}{|\Omega|} \sum_{\substack{R \in \boldsymbol{\mathcal{D}} \\ R \subseteq \Omega}} |\langle b, w_R \rangle|^2 \right)^{1/2}, \tag{1.4}$$

where the supremum ranges over all non-empty open sets Ω of finite measure, $\boldsymbol{\mathcal{D}}$ is the family of all dyadic rectangles in the product space $\mathbb{R} \times \mathbb{R}$ (with sides parallel to the coordinate axes), and $(w_R)_{R \in \boldsymbol{\mathcal{D}}}$ is some (regular enough) wavelet system adapted to dyadic rectangles. Note that an analogous definition of multiparameter product BMO can be given in any product space $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$. Works [8] and [14] provide equivalent descriptions of biparameter product BMO spaces in terms of Carleson measures, extending the one-parameter classical ones. It is important to note that in definition (1.4), one *cannot* restrict the supremum to rectangles. This follows from a famous counterexample due to L. Carleson [6], recounted in Fefferman's article [14] (see also [3] or [30]).

The first breakthrough in the study of the relation between norms of commutators and the BMO norm of their symbol in the multiparameter setting was achieved by S. H. Ferguson and

C. Sadosky [15]. The authors of [15] proved there that

$$\|[H\otimes H, b]\|_{L^2(\mathbb{R}^2)\to L^2(\mathbb{R}^2)} \sim \|b\|_{\mathrm{bmo}(\mathbb{R}\times\mathbb{R})},$$

where H is the Hilbert transform, $H\otimes H$ is a tensor product of Hilbert transforms (each acting on one of the two variables) and $\mathrm{bmo}(\mathbb{R}\times\mathbb{R})$ is the so-called *little bmo* space,

$$\|b\|_{\mathrm{bmo}(\mathbb{R}\times\mathbb{R})} := \sup_R \frac{1}{|R|}\int_R |b(x)-\langle b\rangle_R|\mathrm{d}x,$$

where the supremum is taken over all rectangles R in $\mathbb{R}\times\mathbb{R}$ (with sides parallel to the coordinate axes). In [15], an upper bound for iterated commutators is also established. Namely, if H_1 and H_2 denote, respectively, the Hilbert transforms acting on the first and second variable, then one has the upper bound

$$\|[H_1,[H_2,b]]\|_{L^2(\mathbb{R}^2)\to L^2(\mathbb{R}^2)} \lesssim \|b\|_{\mathrm{BMO}(\mathbb{R}\times\mathbb{R})}. \tag{1.5}$$

Later, L. Dalenc and Y. Ou [11] proved that if T_1, T_2 are (usual one-parameter) Calderón–Zygmund operators acting on the first and second, respectively, variables of $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$ then

$$\|[T_1,[T_2,b]]\|_{L^2(\mathbb{R}^{\vec{d}})\to L^2(\mathbb{R}^{\vec{d}})} \lesssim_{T_1,T_2,\vec{d}} \|b\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \tag{1.6}$$

(in fact, Dalenc–Ou [11] established an analogous result in any number of parameters).

The result of Ferguson–Sadosky [15] was generalized and also extended to the weighted setting by I. Holmes, S. Petermichl and B. D. Wick [19]. There, the authors proved that if T is any biparameter Calderón–Zygmund operator (aka Journé operator) on $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$, then

$$\|[T,b]\|_{L^p(\mu)\to L^p(\lambda)} \lesssim_{T,\vec{d},p} \|b\|_{\mathrm{bmo}(\nu,\mathbb{R}^{\vec{d}})}, \tag{1.7}$$

and that if moreover $R^{(1)},\dots,R^{(d)}$ are the Riesz transforms on $\mathbb{R}^d$, then

$$\sum_{k,l=1}^{d} \|[R^{(k)}\otimes R^{(l)},b]\|_{L^p(\mu)\to L^p(\lambda)} \gtrsim_{p,d} \|b\|_{\mathrm{bmo}(\nu,\mathbb{R}^d\times\mathbb{R}^d)}, \tag{1.8}$$

where in both results, μ and λ are biparameter A_p weights on $\mathbb{R}^{\vec{d}}$ (see Section 2 for the definition), all implied constants depend on the biparameter A_p characteristics of μ,λ as well, $\nu := \mu^{1/p}\lambda^{-1/p}$, $1<p<\infty$, and

$$\|b\|_{\mathrm{bmo}(\nu,\mathbb{R}^{\vec{d}})} := \sup_R \frac{1}{\nu(R)}\int_R |b(x)-\langle b\rangle_R|\mathrm{d}x,$$

where the supremum is again taken over all rectangles R in $\mathbb{R}^{\vec{d}}$ (with sides parallel to the coordinate axes). Holmes–Petermichl–Wick [19] proved and used in an essential way two-weight John–Nirenberg inequalities for the weighted space $\mathrm{bmo}(\nu,\mathbb{R}^{\vec{d}})$ that are analogous to the ones in the one-parameter setting in [18], in order to prove the lower bound (1.8). For the upper bound (1.7), Holmes–Petermichl–Wick [19] defined and used the *dyadic* product Bloom space $\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu, \mathbb{R}^{\vec{d}})$,

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu,\mathbb{R}^{\vec{d}})} := \sup_{\mathcal{U}} \left(\frac{1}{\nu(\mathrm{sh}(\mathcal{U}))} \sum_{R\in\mathcal{U}} |b_R|^2 \langle \nu^{-1} \rangle_R \right)^{1/2},$$

where the supremum ranges over all non-empty collections $\mathcal{U}$ of *dyadic* rectangles in the biparameter product space $\mathbb{R}^{\vec{d}}$,

$$\mathrm{sh}(\mathcal{U}) := \bigcup_{R\in\mathcal{U}} R,$$

and $b_R := \langle b, h_R \rangle$, where h_R is the cancellative in both-variables L^2-normalized Haar function over the dyadic rectangle R. In [19] a duality result between $\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu, \mathbb{R}^{\vec{d}})$ and a weighted H^1 space is established, which is essential for the proof of the upper bound (1.7). The authors of [19] also extended the upper BMO bound (1.6) due to Dalenc–Ou in [11] to the case of multiparameter indexed unweighted BMO spaces.

Since then, weighted product BMO and multiparameter indexed weighted BMO upper bounds have been investigated and established in full generality by E. Airta, K. Li, H. Martikainen and E. Vuorinen [1], [2], [23], [24].

It is important to note that in all of the aforementioned works, the main tools for establishing upper bounds for norms of commutators with Calderón–Zygmund operators in terms of the BMO norm of their symbol are the decomposition theorem of Calderón–Zygmund operators in terms of Haar shifts and paraproducts that T. Hytönen established and used to prove his A_2 theorem [20], as well as its extension to the multiparameter setting due to Martikainen [25].

While upper bounds for norms of commutators in terms of multiparameter BMO norms of the symbol are by now well understood, in the fully general two-weight setting, the picture for lower bounds remains incomplete even in the unweighted case. The study of such lower bounds was addressed by S. H. Ferguson and M. T. Lacey [16], who gave a converse to (1.5), namely

$$\|[H_1, [H_2, b]]\|_{L^2(\mathbb{R}^2)\to L^2(\mathbb{R}^2)} \gtrsim \|b\|_{\mathrm{BMO}(\mathbb{R}\times\mathbb{R})}. \tag{1.9}$$

The proof of this fact was based on new and beautiful arguments on this matter. More recently, Dalenc–Petermichl [12] proved similar results for iterated commutators of Riesz transforms. Moreover, Y. Ou, S. Petermichl and E. Strouse [29] extended the result in Dalenc–Petermichl [12] to the case of multiparameter-indexed BMO spaces, which are between little bmo and multiparameter Chang–Fefferman product BMO. Both works [12] and [29] rely on the result of Ferguson–Lacey [16].

There are as well some lower bounds that do not rely on (1.9), like [18, 19], [22]. However, all these employ variants of the original argument by Coifman–Rochberg–Weiss [9]. Arguments of this type rely on the availability of explicit ‘oscillatory’ expressions for BMO norms. While such expressions are indeed available in the one-parameter setting, and also in the case of the little bmo space, they are not at all available in the case of product BMO (in the sense of Chang–Fefferman), making investigating such lower bounds significantly harder.

In another direction, Ó. Blasco and S. Pott [4] related dyadic biparameter product BMO norms to iterated commutators of Haar multipliers. More precisely, consider the set Σ of all finitely supported maps $\sigma : \mathcal{D} \to \{-1, 0, 1\}$, and for each $\sigma \in \Sigma$ consider its Haar multiplier T_σ on $L^2(\mathbb{R})$ (see Section 4 for precise definitions). Furthermore, consider Haar multipliers $T^1_{\sigma_1}$ and $T^2_{\sigma_2}$ acting on $L^2(\mathbb{R}^2)$ separately on each variable. Blasco–Pott [4] show that

$$\sup_{\sigma_1,\sigma_2\in\Sigma} \|[T^1_{\sigma_1}, [T^2_{\sigma_2}, b]]\|_{L^2(\mathbb{R}^2)\to L^2(\mathbb{R}^2)} \sim \|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}}.$$

It should be noted that the supremum over all signs enables Blasco–Pott [4] to eliminate error terms by taking average over all signs and then use orthogonality arguments in order to conclude their result.

The main goal of the present paper is to extend the aforementioned result by Blasco–Pott [4] to the weighted setting, and to the full range of exponents $1 < p < \infty$. More precisely, we show the following.

Theorem 1.1. *Let $1 < p < \infty$. Consider a function $b \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, dyadic biparameter A_p weights μ, λ and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\sup_{\sigma_1,\sigma_2\in\Sigma} \|[T^1_{\sigma_1}, [T^2_{\sigma_2}, b]]\|_{L^p(\mu)\to L^p(\lambda)} \sim \|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)}, \tag{1.10}$$

where the implied constants depend only on p, $[\mu]_{A_p,\mathcal{D}}$ and $[\lambda]_{A_p,\mathcal{D}}$.

The proof of this theorem will follow similar steps to that of the result due to Blasco–Pott. Namely, firstly we show that the supremum in the left-hand side of (1.10) is equivalent to the norm of a certain operator defined in terms of paraproducts. Then we show the equivalence between the previous operator norm and the BMO norm of the symbol b. Nonetheless, here we must use additional techniques to overcome the lack of orthogonality for $p \neq 2$. In particular, we make use of a multiparameter extension of the classical Khintchine's inequality together with vector-valued estimates. Another possibility that also allows one to circumvent this difficulty would be to use duality coupled with Hölder's inequality, together with vector-valued estimates. Moreover, to be able to handle the weighted spaces appearing in Theorem 1.1, we establish equivalent characterizations of dyadic product Bloom BMO in the spirit of [18, 19]. More precisely, fix $1 < p < \infty$ and two dyadic biparameter A_p weights μ and λ on $\mathbb{R}^2$. Given $b \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, define the dyadic two-weight Bloom product BMO norm

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} := \sup_{\mathcal{U}} \frac{1}{(\mu(\mathrm{sh}(\mathcal{U})))^{1/p}} \|S_{\mathcal{U}}(b)\|_{L^p(\lambda)},$$

where the supremum ranges again over all non-empty collections $\mathcal{U}$ of dyadic rectangles in $\mathbb{R} \times \mathbb{R}$, and $S_{\mathcal{U}}(b)$ is the biparameter dyadic square function of b restricted to the collection $\mathcal{U}$ (see Section 2 for precise definitions). We show the following two-weight John–Nirenberg inequalities.

Theorem 1.2. *Let $1 < p < \infty$. Consider dyadic biparameter A_p weights μ, λ and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)} \sim \|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)},$$

where the implied constants depend only on p, $[\mu]_{A_p,\mathcal{D}}$ and $[\lambda]_{A_p,\mathcal{D}}$.

While the formulation of Theorem 1.2 reflects that of the two-weight John–Nirenberg inequalities in the one-parameter setting [18] and in the case of little bmo [19], its proof addresses several new difficulties not present in [18, 19] and requires new ideas. We split this proof in several steps. For $1 < p \leqslant 2$ we show that the two-weight norm $\|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)}$ is equivalent to a certain paraproduct norm, as an operator from $L^p(\mu)$ to $L^p(\lambda)$. Then, we prove the equivalence between the norm of this paraproduct and the one-weight norm $\|b\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)}$ for any $1 < p < \infty$. Although the remaining equivalence, that is for $2 < p < \infty$, is an immediate consequence of Hölder's inequality in the unweighted case, in our setting it requires the use of the so-called (biparameter) Triebel–Lizorkin square function (see Section 2 for the definition and basic properties). In addition, it is essential for this step of the proof to make use of an equivalence between one-weight and unweighted product BMO due to E. Airta, K. Li, H. Martikainen and E. Vuorinen [2]. In particular, the equivalence from [2] that we use corresponds to the particular case of our Theorem 1.2 when $p > 2$ and $\mu = \lambda$. These two ingredients are necessary to overcome the lack of a 'two-weight Hölder's inequality'.

Note that John–Nirenberg inequalities hold for weighted little bmo, for all $1 < p < \infty$ (see [19]). Moreover, it was already known that they also hold in unweighted product BMO as well, for all $1 < p < \infty$ (see [30] for a proof using atomic decompositions), and in fact for all $0 < p < \infty$ (see, for example, [2]). Finally, for rectangular BMO, John–Nirenberg inequalities do not hold even in the unweighted case, and for any $1 < p < \infty$ (see [4]).

It should also be noted that all of our results hold for functions defined on any multiparameter product space $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$, with identical or similar proofs. Although for simplicity we restrict the main part of this work to the case of functions on $\mathbb{R} \times \mathbb{R}$, we also indicate how to modify the arguments for the general multiparameter setting when appropriate. Moreover, we explain briefly how to extend our results to the whole scale of indexed spaces between little bmo and product BMO in the general multiparameter setting, with the appropriate iterated commutator in each case.

Plan of the paper. The article is structured as follows. In Section 2 the reader can find the notations and definitions that will be used in the rest of the paper. In Section 3 we prove the two-weight John–Nirenberg inequalities of Theorem 1.2. In Section 4 we prove Theorem 1.1. Finally, in Section 5 we extend Theorem 1.1 to the whole scale of indexed spaces between little bmo and product BMO in the general multiparameter setting, with the appropriate iterated commutator in each case.

2 | BACKGROUND AND NOTATION

We collect here some notation, definitions and a few basic facts that will be used repeatedly in the sequel. While all of them are also valid in any multiparameter product space $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$, with the obvious modifications, for simplicity we restrict ourselves to the case of the product space $\mathbb{R} \times \mathbb{R}$.

2.1 | Dyadic intervals and dyadic rectangles

We denote by $\mathcal{D}$ the set of all dyadic intervals in $\mathbb{R}$,

$$\mathcal{D} := \{[m2^k, (m+1)2^k) : k, m \in \mathbb{Z}\}.$$

We also denote by $\boldsymbol{\mathcal{D}}$ the set of all dyadic rectangles in $\mathbb{R}^2$,

$$\boldsymbol{\mathcal{D}} := \{I \times J \; : \; I, J \in \mathcal{D}\}.$$

For any $E \subseteq \mathbb{R} \times \mathbb{R}$ we denote

$$\boldsymbol{\mathcal{D}}(E) := \{R \in \boldsymbol{\mathcal{D}} \; : \; R \subseteq E\}.$$

Note that if $E \in \boldsymbol{\mathcal{D}}$, then $E \in \boldsymbol{\mathcal{D}}(E)$.

2.2 | Haar systems

2.2.1 | Haar system on $\mathbb{R}$

For any $I \in \mathcal{D}$, $h_I^{(0)}, h_I^{(1)}$ will denote, respectively, the L^2-normalized *cancellative* and *non-cancellative* Haar functions over the interval $I \in \mathcal{D}$, that is

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

(so $h_I^{(0)}$ has mean 0). For simplicity we denote $h_I := h_I^{(0)}$. For any function $f \in L^1_{\text{loc}}(\mathbb{R})$, we denote $f_I := \langle f, h_I \rangle$, $I \in \mathcal{D}$. We will also denote by Q_I the projection on the one-dimensional subspace spanned by h_I,

$$Q_I f := f_I h_I, \qquad f \in L^1_{\text{loc}}(\mathbb{R}).$$

It is well known that one has the expansion

$$f = \sum_{I \in \mathcal{D}} f_I h_I, \qquad \forall f \in L^2(\mathbb{R})$$

in the $L^2(\mathbb{R})$-sense, and that the system $\{h_I\}_{I \in \mathcal{D}}$ forms an orthonormal basis for $L^2(\mathbb{R})$. Recall that for a function $f \in L^1_{\text{loc}}(\mathbb{R})$ we denote by $\langle f \rangle_I$ its (unweighted) average on interval I, that is,

$$\langle f \rangle_I = \frac{1}{|I|} \int_I f(x) \mathrm{d}x.$$

We will also make this notation extensive later on to averages on rectangles of locally integrable functions on $\mathbb{R}^n$. It is then easy to see that for any $I \in \mathcal{D}$ there holds

$$\mathbf{1}_I (f - \langle f \rangle_I) = \sum_{\substack{J \in \mathcal{D} \\ J \subseteq I}} f_J h_J.$$

2.2.2 | Haar system on the product space $\mathbb{R} \times \mathbb{R}$

If $R = I \times J$ is a dyadic rectangle in $\mathbb{R}^2$, we denote by $h_R^{(\varepsilon_1, \varepsilon_2)}$ any of the four L^2-normalized Haar functions over R,

$$h_R^{(\varepsilon_1, \varepsilon_2)} := h_I^{(\varepsilon_1)} \otimes h_J^{(\varepsilon_2)}, \qquad \varepsilon_1, \varepsilon_2 \in \{0, 1\},$$

that is

$$h_R^{(\varepsilon_1,\varepsilon_2)}(t,s) = h_I^{(\varepsilon_1)}(t)h_J^{(\varepsilon_2)}(s), \qquad (t,s) \in \mathbb{R}^2.$$

For simplicity we denote $h_R := h_R^{(0,0)}$. For any function $f \in L^1_{\text{loc}}(\mathbb{R}^2)$, we denote

$$f_R^{(\varepsilon_1,\varepsilon_2)} := \langle f, h_R^{(\varepsilon_1,\varepsilon_2)}\rangle, \qquad R \in \boldsymbol{\mathcal{D}}, \qquad \varepsilon_1,\varepsilon_2 \in \{0,1\},$$

and we will often use the simplification $f_R := \langle f, h_R\rangle$. We will also denote by Q_R the projection on the one-dimensional subspace spanned by h_R,

$$Q_R f := f_R h_R, \qquad f \in L^1_{\text{loc}}(\mathbb{R}^2).$$

From the corresponding one-dimensional facts we immediately deduce the expansion

$$f = \sum_{R\in\boldsymbol{\mathcal{D}}} f_R h_R, \qquad \forall f \in L^2(\mathbb{R}^2)$$

in the $L^2(\mathbb{R}^2)$-sense, and that the system $\{h_R\}_{R\in\boldsymbol{\mathcal{D}}}$ forms an orthonormal basis for $L^2(\mathbb{R}^2)$. It is then easy to see by direct computation, following a reasoning similar to that of the inclusion–exclusion principle, that for any $R = I \times J \in \boldsymbol{\mathcal{D}}$ there holds

$$\sum_{R'\in\boldsymbol{\mathcal{D}}(R)} f_{R'}h_{R'}(t,s) = \mathbf{1}_R(t,s)(f(t,s) - \langle f(\cdot,s)\rangle_I - \langle f(t,\,\cdot\,)\rangle_J + \langle f\rangle_R).$$

Finally, for $I, J \in \mathcal{D}$ we denote by Q_I^1, Q_J^2 the operators acting on functions $f \in L^1_{\text{loc}}(\mathbb{R}^2)$ by

$$Q_I^1 f(t,s) = Q_I(f(\,\cdot\,,s))(t), \qquad Q_J^2 f(t,s) = Q_J(f(t,\,\cdot\,))(s), \qquad \text{for a.e. } (t,s) \in \mathbb{R}\times\mathbb{R}.$$

Thus, if $R = I \times J$, then $Q_R = Q_I^1 Q_J^2 = Q_J^2 Q_I^1$. Note that $(Q_I^1)^2 = Q_I^1$ and $(Q_J^2)^2 = Q_J^2$.

Observe that the Haar system that we have just defined using dyadic rectangles in $\mathbb{R}\times\mathbb{R}$ differs from that defined using dyadic squares on $\mathbb{R}^2$ when considered as a one-parameter space instead of a product space. In the latter case, for a given dyadic square $Q = I \times J$, $I, J \in \mathcal{D}$ and $|I| = |J|$, one defines the Haar function $h_Q^{(\varepsilon_1,\varepsilon_2)} = h_I^{(\varepsilon_1)} \otimes h_J^{(\varepsilon_2)}$ with $(\varepsilon_1,\varepsilon_2) \neq (1,1)$. Both the biparameter system $\{h_I \otimes h_J\}_{I,J}$ with $I, J \in \mathcal{D}$ and the one-parameter system $\{h_I^{(\varepsilon_1)} \otimes h_J^{(\varepsilon_2)}\}_{I,J}$ with $I, J \in \mathcal{D}$, $|I| = |J|$ and $(\varepsilon_1,\varepsilon_2) \neq (1,1)$ are orthonormal bases of $L^2(\mathbb{R}^2) = L^2(\mathbb{R}\times\mathbb{R})$. However, the system defined using dyadic rectangles is more suitable to study problems that are invariant under different rescalings on each variable (as it is the case in the product spaces that we treat here), since it considers all possible combinations of scales on each variable. On the other hand, the system defined using dyadic cubes is more suitable for problems that are invariant only under uniform rescalings on all variables, since all of its functions take the same scale on each variable.

2.3 | A_p weights

By weight we always mean a locally integrable, a.e. positive function. We fix in what follows $1 < p < \infty$. Given a weight w, we consider the weighted Lebesgue space $L^p(w)$, which is the space of p-integrable functions with respect to the measure $w(x)\,\mathrm{d}x$. In other words, we say that a function

f belongs to $L^p(w)$ if

$$\|f\|_{L^p(w)} := \left(\int |f(x)|^p w(x)\,\mathrm{d}x\right)^{1/p} < \infty.$$

The dual of $L^p(w)$ is the weighted space $L^{p'}(w')$ under the usual L^2-pairing $\langle f, g\rangle = \int f(x)\overline{g(x)}\,\mathrm{d}x$, where p' is the Hölder conjugate of p (that is, $\frac{1}{p} + \frac{1}{p'} = 1$) and $w'(x) = (w(x))^{1-p'}$ is the conjugate weight to w. Note here that, in the particular case $p = 2$, one has that the dual of $L^2(w)$ is $L^2(w^{-1})$.

2.3.1 | A_p weights on $\mathbb{R}$

Consider a weight w on $\mathbb{R}$. It should be noted that all that follows would still hold if we considered weights on $\mathbb{R}^d$, for any d, just by substituting intervals of $\mathbb{R}$ by cubes of $\mathbb{R}^d$. We define the *Muckenhoupt A_p characteristic* of w, denoted by $[w]_{A_p}$, as

$$[w]_{A_p} := \sup_I \langle w\rangle_I \langle w^{-1/(p-1)}\rangle_I^{p-1} = \sup_I \langle w\rangle_I \langle w'\rangle_I^{p-1},$$

where the supremum is taken over all intervals I in $\mathbb{R}$. We define a dyadic version of this by

$$[w]_{A_p,\mathcal{D}} := \sup_{I\in\mathcal{D}} \langle w\rangle_I \langle w^{-1/(p-1)}\rangle_I^{p-1}.$$

We say that w is an A_p weight, respectively a dyadic A_p weight, if $[w]_{A_p} < \infty$, respectively, $[w]_{A_p,\mathcal{D}} < \infty$. It is a very well-known fact that $[w]_{A_p,\mathcal{D}} \geqslant 1$, in fact an immediate application of Hölder's inequality gives

$$\langle w\rangle_I \langle w^{-1/(p-1)}\rangle_I^{p-1} = \left(\langle w\rangle_I^{1/p} \langle w^{-1/(p-1)}\rangle_I^{1/p'}\right)^p \geqslant \langle w^{1/p} w^{-1/(p'(p-1))}\rangle_I^p = \langle w^{1/p} w^{-1/p}\rangle_I^p = 1,$$

for any interval I, where $p' = p/(p-1)$. Observe as well that w is an A_p weight if and only if w' is an $A_{p'}$ weight, and in this case $[w']_{A_{p'}} = [w]_{A_p}^{p'-1}$. The analogous fact is also true in the dyadic case.

It is a classical result that $[w]_{A_p} < \infty$ if and only if the Hardy–Littlewood maximal function M given by

$$Mf := \sup_I \langle |f|\rangle_I \mathbf{1}_I,$$

where supremum is taken over all intervals I in $\mathbb{R}$, is bounded as an operator from $L^p(w)$ into itself and that in fact one has the estimate

$$\|M\|_{L^p(w)\to L^p(w)} \lesssim_p [w]_{A_p}^{1/(p-1)}. \tag{2.1}$$

A dyadic version of this is also true for the dyadic Hardy–Littlewood maximal function $M_{\mathcal{D}}$ given by

$$M_{\mathcal{D}}f := \sup_{I\in\mathcal{D}} \langle |f|\rangle_I \mathbf{1}_I.$$

2.3.2 | Biparameter A_p weights on $\mathbb{R}^2$

Consider now a weight w on $\mathbb{R} \times \mathbb{R}$. As before, all that follows would be equally valid for weights on any $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$, with the obvious modifications. We define the *biparameter Muckenhoupt A_p characteristic* $[w]_{A_p}$ of w by

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

where supremum is taken over *all* rectangles in $\mathbb{R} \times \mathbb{R}$ (with sides parallel to the coordinate axis). We define a dyadic version of this by

$$[w]_{A_p,\boldsymbol{\mathcal{D}}} := \sup_{R \in \boldsymbol{\mathcal{D}}} \langle w \rangle_R \langle w^{-1/(p-1)} \rangle_R^{p-1}.$$

We say that w is a biparameter A_p weight, respectively, a dyadic biparameter A_p weight, if $[w]_{A_p} < \infty$, respectively, $[w]_{A_p,\boldsymbol{\mathcal{D}}} < \infty$. Note that similarly to the one-parameter case we have $[w]_{A_p} \geqslant 1$ and $[w']_{A_{p'}} = [w]_{A_p}^{p'-1}$, as well as the analogous facts for dyadic A_p and dyadic $A_{p'}$ weights.

Consider the *strong* maximal function M_{S} given by

$$M_{\mathrm{S}} f := \sup_R \langle |f| \rangle_R \mathbf{1}_R,$$

where as previously the supremum is taken over *all* rectangles in $\mathbb{R} \times \mathbb{R}$ (with sides parallel to the coordinate axis). Consider also the Hardy–Littlewood maximal functions acting in each variable separately,

$$M^1 f(t,s) := M(f(\cdot, s))(t), \qquad M^2 f(t,s) := M(f(t, \cdot))(s), \qquad f \in L^1_{\mathrm{loc}}(\mathbb{R}^2).$$

Using the Lebesgue differentiation theorem it is easy to see that

$$[w]_{A_p} \geqslant \max(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}} [w(x_1, \cdot)]_{A_p}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}} [w(\cdot, x_2)]_{A_p}).$$

It is also easy to see that $M_{\mathrm{S}} f \leqslant M^1(M^2 f)$, which coupled with (2.1) implies immediately

$$\|M_{\mathrm{S}}\|_{L^p(w) \to L^p(w)} \lesssim_p \operatorname*{ess\,sup}_{x_1 \in \mathbb{R}} [w(x_1, \cdot)]_{A_p}^{1/(p-1)} \cdot \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}} [w(\cdot, x_2)]_{A_p}^{1/(p-1)}.$$

On the other hand, it is also immediate to see that

$$\|M_{\mathrm{S}}\|_{L^p(w) \to L^p(w)} \geqslant [w]_{A_p}^{1/p}.$$

Therefore, w is a biparameter A_p weight if and only if w is an A_p weight in each variable separately and uniformly. Dyadic versions of these facts are similarly true for the *dyadic strong* maximal function $M_{\boldsymbol{\mathcal{D}}}$ given by

$$M_{\boldsymbol{\mathcal{D}}} f := \sup_{R \in \boldsymbol{\mathcal{D}}} \langle |f| \rangle_R \mathbf{1}_R.$$

Here it is worth mentioning the analogous property for A_p weights defined on $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$. We firstly introduce some convenient notation for the general setting. Given $x \in \mathbb{R}^{\vec{d}}$ and $k \in \{1, 2, \dots, t\}$, let us denote $x_{\overline{k}} := (x_1, \dots, x_{k-1}, \cdot, x_{k+1}, \dots, x_t)$, so that a function $f(x_{\overline{k}})$ depends only on the variable x_k. Similarly, we denote $\mathbb{R}^{\vec{d}_{\overline{k}}} := \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_{k-1}} \times \mathbb{R}^{d_{k+1}} \times \cdots \times \mathbb{R}^{d_t}$. In general, we will also extend this notation to any number of parameters, so that if $k_1, \dots, k_s \in \{1, 2, \dots, t\}$, then we will denote by $x_{\overline{k_1,\dots,k_s}} := (x_1, \dots, x_{k_1-1}, \cdot, x_{k_1+1}, \dots, x_{k_s-1}, \cdot, x_{k_s+1}, \dots, x_t)$, and similarly for $\mathbb{R}^{\vec{d}_{\overline{k_1,\dots,k_s}}}$. Then, the same reasoning as before shows that a weight w is a multiparameter A_p weight on $\mathbb{R}^{\vec{d}}$ if and only if, for any subsequence $k = (k_1, \dots, k_s)$ of $(1, \dots, t)$, the weight $w(x_{\overline{k}})$ is an A_p weight on $\mathbb{R}^{\vec{d}_{\overline{k}}}$ with A_p characteristic uniformly bounded on $x \in \mathbb{R}^{\vec{d}}$. In particular, it is easy to see that

$$[w]_{A_p} \geqslant \max_k \left(\operatorname*{ess\,sup}_{x_{\overline{k}} \in \mathbb{R}^{\vec{d}_{\overline{k}}}} [w(x_{\overline{k}})]_{A_p} \right),$$

where the maximum ranges over all subsequences k of $(1, \dots, t)$.

2.3.3 | Averages of A_p weights

We recall a few standard facts about averages of A_p weights. Let μ, λ be biparameter A_p weights on $\mathbb{R} \times \mathbb{R}$. Using several times Jensen's inequality, Hölder's inequality and the A_p condition for the weights μ and λ it is easy to see that for all rectangles R one has the estimates

$$\langle \mu^{1/p} \rangle_R \sim_{[\mu]_{A_p}} \langle \mu \rangle_R^{1/p} \sim_{[\mu]_{A_p}} \langle \mu^{-1/(p-1)} \rangle_R^{-(p-1)/p} \sim_{[\mu]_{A_p}} \langle \mu^{-1/p} \rangle_R^{-1}, \tag{2.2}$$

and

$$\langle \mu^{1/p} \lambda^{-1/p} \rangle_R \sim_{[\mu]_{A_p}, [\lambda]_{A_p}} \langle \mu \rangle_R^{1/p} \langle \lambda \rangle_R^{-1/p}, \tag{2.3}$$

see [22, p. 2] for a sketch of the argument (only the one-parameter setting is treated there, but it is obvious that the same argument works in the multiparameter setting without any changes at all). Moreover, using Hölder's inequality it is easy to see that

$$1 \leqslant [\mu^{1/p} \lambda^{-1/p}]_{A_2} \leqslant [\mu]_{A_p}^{1/p} [\lambda]_{A_p}^{1/p},$$

see [18, Lemma 2.7] for a full proof (again, while only the one-parameter setting is treated there, the multiparameter result follows from the same arguments).

Note that dyadic versions of all the above facts are similarly true.

2.4 | Dyadic square functions and Littlewood–Paley estimates

We denote by $S_{\mathcal{D}}$ the dyadic square function in $\mathbb{R}$,

$$S_{\mathcal{D}} f := \left(\sum_{I \in \mathcal{D}} |Q_I f|^2 \right)^{1/2} = \left(\sum_{I \in \mathcal{D}} |f_I|^2 \frac{\mathbf{1}_I}{|I|} \right)^{1/2}, \qquad f \in L^1_{\mathrm{loc}}(\mathbb{R}).$$

We also denote by $S_{\mathcal{D}}$ the dyadic biparameter square function in $\mathbb{R}\times\mathbb{R}$,

$$S_{\boldsymbol{\mathcal{D}}} f := \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |Q_R f|^2\right)^{1/2} = \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |f_R|^2 \frac{\mathbf{1}_R}{|R|}\right)^{1/2}, \qquad f\in L^1_{\mathrm{loc}}(\mathbb{R}^2).$$

It is well known that if w is a dyadic A_p weight on $\mathbb{R}$, $1<p<\infty$, then

$$\|S_{\mathcal{D}} f\|_{L^p(w)} \sim_{p,[w]_{A_p,\mathcal{D}}} \|f\|_{L^p(w)},$$

for all (suitable) functions f on $\mathbb{R}$. Iterating this and using well-known results about vector-valued extensions of linear operators (see, for example, [17, Chapter 5]) we deduce, as remarked in [19], that if w is a dyadic biparameter A_p weight on $\mathbb{R}\times\mathbb{R}$, then

$$\|S_{\boldsymbol{\mathcal{D}}} f\|_{L^p(w)} \sim_{p,[w]_{A_p,\boldsymbol{\mathcal{D}}}} \|f\|_{L^p(w)},$$

for all (suitable) functions f on $\mathbb{R}^2$ (for example, $f\in L^\infty_{\mathrm{c}}(\mathbb{R}^2)$ suffices, and then using approximation arguments one can extend it to more general functions f). In particular, the set of all finite linear combinations of (bi-cancellative) Haar functions in $\mathbb{R}^2$ is dense in $L^p(w)$.

Note that for $p=2$ we simply have

$$\|S_{\boldsymbol{\mathcal{D}}} f\|_{L^2(w)} = \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |f_R|^2 \langle w\rangle_R\right)^{1/2}.$$

The above results also hold in the multiparameter setting, with the usual modifications for general product spaces.

For the general multiparameter case, we will also need to consider *indexed square functions* for functions defined on $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_t}$. For $i\in\{1,\ldots,t\}$ and $I_i\in\mathcal{D}(\mathbb{R}^{d_i})$, we denote here by $Q^i_{I_i}$ the operator Q_{I_i} acting on the ith variable, that is, $Q^i_{I_i} f(x) = Q_{I_i}(f(x_{\bar{i}}))(x_i)$ for $f\in L^1_{\mathrm{loc}}(\mathbb{R}^{\vec{d}})$. Similarly, for a subsequence $k=(k_1,\ldots,k_s)$ of $(1,\ldots,t)$, let $R = I_{k_1}\times\cdots\times I_{k_s}\in\boldsymbol{\mathcal{D}}(\mathbb{R}^{\vec{d}_{\bar{k}}})$ and denote $Q^k_R = Q^{k_1}_{I_1}\ldots Q^{k_s}_{I_s}$. Fix a subsequence $k=(k_1,\ldots,k_s)$ of $(1,\ldots,t)$. Then, we define the k*-indexed square function* $S^k_{\boldsymbol{\mathcal{D}}} f$ of f by

$$S^k_{\boldsymbol{\mathcal{D}}} f := \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |Q^k_R f|^2\right)^{1/2} = \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |f^k_R|^2 \frac{\mathbf{1}^k_R}{|R|}\right)^{1/2}, \qquad f\in L^1_{\mathrm{loc}}(\mathbb{R}^{\vec{d}})$$

(here we make an abuse of the notation $\boldsymbol{\mathcal{D}}$, relying on the context for understanding to which space it refers). Then, using an A_∞-extrapolation result due to D. Cruz-Uribe, J. M. Martell and C. Pérez [10, Theorem 2.1] one can see that for any subsequence k of $(1,\ldots,t)$, any $1<p<\infty$ and any dyadic t-parameter A_p weight w it holds that

$$\|S^k_{\boldsymbol{\mathcal{D}}} f\|_{L^p(w)} \sim_{\vec{d},p,[w]_{A_p,\boldsymbol{\mathcal{D}}}} \|f\|_{L^p(w)}$$

(see also [1, Lemma 2.2]).

2.4.1 | Dyadic square function over collections of dyadic rectangles

Let $\mathcal{U}$ be any collection of dyadic rectangles in $\mathbb{R}\times\mathbb{R}$. We denote

$$P_{\mathcal{U}} f := \sum_{R\in\mathcal{U}} f_R h_R, \qquad S_{\mathcal{U}} f := \left(\sum_{R\in\mathcal{U}} |f_R|^2 \frac{\mathbf{1}_R}{|R|}\right)^{1/2}.$$

By the above we have that if w is a biparameter A_p weight on $\mathbb{R}\times\mathbb{R}$, then there holds

$$\|P_{\mathcal{U}} f\|_{L^p(w)} \sim_{p,[w]_{A_p,\boldsymbol{D}}} \|S_{\mathcal{U}} f\|_{L^p(w)} \leqslant \|S_{\boldsymbol{D}} f\|_{L^p(w)} \sim_{p,[w]_{A_p,\boldsymbol{D}}} \|f\|_{L^p(w)}.$$

In particular

$$\|Q_R f\|_{L^p(w)} \lesssim_{[w]_{A_p,\boldsymbol{D}}} \|f\|_{L^p(w)}, \qquad \forall R\in\boldsymbol{D}.$$

If Ω is any subset of $\mathbb{R}^2$, we denote

$$P_\Omega f := P_{\boldsymbol{D}(\Omega)} f.$$

Again, all these statements hold in the multiparameter setting as well.

2.4.2 | Incorporating the weight in the square function

It is sometimes convenient to define square functions in an alternative way that directly incorporates the weight in the operator. This type of square functions and their properties will be useful later for the proof of Proposition 3.3. Namely, given $1<p<\infty$ and any dyadic biparameter A_p weight w on $\mathbb{R}^2$, define

$$S_w f := \left(\sum_{R\in\boldsymbol{D}} |f_R|^2 \langle w\rangle_R^{2/p} \frac{\mathbf{1}_R}{|R|}\right)^{1/2}, \qquad f\in L^1_{\mathrm{loc}}(\mathbb{R}^2).$$

This type of square functions appears naturally in the theory of matrix-valued weights (see, for example, [21] for estimates in the matrix-weighted one-parameter setting), and is sometimes referred to as the Triebel–Lizorkin square function associated to w, because L^p bounds for it allow one to identify $L^p(w)$ as a certain Trielel–Lizorkin space (see, for example, [13] for the scalar one-parameter case). Here we prove an estimate in the scalar biparameter setting. In fact, the proofs of Lemma 2.1 and Corollary 2.2 below readily extend to the matrix-valued setting. This will be part of forthcoming work of the authors.

Lemma 2.1. *Let $f\in L^1_{\mathrm{loc}}(\mathbb{R}^2)$. Then*

$$\|S_w f\|_{L^p} \lesssim_{p,[w]_{A_p,\boldsymbol{D}}} \|Sf\|_{L^p(w)}.$$

In particular

$$\|S_w f\|_{L^p} \lesssim_{p,[w]_{A_p,\mathcal{D}}} \|f\|_{L^p(w)}, \qquad \forall f \in L^\infty_{\mathrm{c}}(\mathbb{R}^2).$$

Proof. First of all, we note that by the Monotone Convergence Theorem we can assume without loss of generality that f has only finitely-many non-zero Haar coefficients. Now, we notice that by (2.2) we have

$$\begin{aligned}\|S_w f\|^p_{L^p} &= \int_{\mathbb{R}^2} \left(\sum_{R\in\mathcal{D}} |f_R|^2 \langle w\rangle_R^{2/p} \frac{\mathbf{1}_R(x)}{|R|} \right)^{p/2} \mathrm{d}x \\ &\lesssim_{p,[w]_{A_p,\mathcal{D}}} \int_{\mathbb{R}^2} \left(\sum_{R\in\mathcal{D}} |f_R|^2 (\langle w^{1/p}\rangle_R)^2 \frac{\mathbf{1}_R(x)}{|R|} \right)^{p/2} \mathrm{d}x.\end{aligned}$$

Thus, by standard (unweighted) dyadic Littlewood–Paley theory we only have to prove that

$$\|F\|_{L^p} := \left\| \sum_{R\in\mathcal{D}} |f_R| \langle w^{1/p}\rangle_R h_R \right\|_{L^p} \lesssim_p \|Sf\|_{L^p(w)}.$$

We use duality for that. Let $g \in L^{p'}$. Then, we have

$$\begin{aligned}|\langle F, g\rangle| &\leqslant \sum_{R\in\mathcal{D}} |f_R| \langle w^{1/p}\rangle_R \cdot |g_R| = \int_{\mathbb{R}^2} \sum_{R\in\mathcal{D}} |f_R| w(x)^{1/p} \cdot h_R(x) \cdot |g_R| \cdot h_R(x)\mathrm{d}x \\ &\leqslant \int_{\mathbb{R}^2} (Sf)(x) w(x)^{1/p} (Sg)(x)\mathrm{d}x \leqslant \|Sf\|_{L^p(w)} \|Sg\|_{L^{p'}} \sim_p \|Sf\|_{L^p(w)} \|g\|_{L^{p'}},\end{aligned}$$

concluding the proof. □

Of course, for $p = 2$ we have just $\|S_w f\|_{L^2} = \|Sf\|_{L^2(w)}$.

Using Lemma 2.1, we can immediately deduce the corresponding lower bound.

Corollary 2.2. *There holds*

$$\|f\|_{L^p(w)} \lesssim_{p,[w]_{A_p,\mathcal{D}}} \|S_w f\|_{L^p}, \qquad \forall f \in L^\infty_{\mathrm{c}}(\mathbb{R}^2).$$

Proof. We use duality. Recall that $w' := w^{-1/(p-1)}$ is a dyadic biparameter $A_{p'}$ weight with $[w]^{1/p'}_{A_{p'}} = [w]^{1/p}_{A_p}$. Let $f, g \in L^\infty_{\mathrm{c}}(\mathbb{R}^2)$. Then, using (2.2) and applying Lemma 2.1 (for the weight w') we get

$$\begin{aligned}|\langle f, g\rangle| &\leqslant \sum_{R\in\mathcal{D}} |f_R| \cdot |g_R| = \sum_{R\in\mathcal{D}} \int_{\mathbb{R}^2} |f_R| \langle w\rangle_R^{1/p} h_R(x) \cdot |g_R| \langle w\rangle_R^{-1/p} h_R(x)\mathrm{d}x \\ &\lesssim_{p,[w]_{A_p,\mathcal{D}}} \sum_{R\in\mathcal{D}} \int_{\mathbb{R}^2} |f_R| \langle w\rangle_R^{1/p} h_R(x) \cdot |g_R| \langle w'\rangle_R^{1/p'} h_R(x)\mathrm{d}x\end{aligned}$$

$$\leqslant \int_{\mathbb{R}^2} (S_w f(x))(S_{w'} g(x))\mathrm{d}x \leqslant \|S_w f\|_{L^p} \|S_{w'} g\|_{L^{p'}}$$

$$\lesssim_{p,[w]_{A_p,\boldsymbol{D}}} \|S_w f\|_{L^p} \|g\|_{L^{p'}(w')},$$

so $\|f\|_{L^p(w)} \lesssim_{p,[w]_{A_p,\boldsymbol{D}}} \|S_w f\|_{L^p}$, concluding the proof. □

It is clear that the above results remain true in the general multiparameter setting.

3 | EQUIVALENCES FOR DYADIC BLOOM PRODUCT BMO

We prove in this section that the dyadic Bloom product BMO admits several equivalent descriptions. We will focus on dyadic biparameter A_p weights, with $1 < p < \infty$, on $\mathbb{R} \times \mathbb{R}$. It should be noted that the results presented here also hold in the general multiparameter setting of functions and weights defined on $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$, with identical proofs. However, for simplicity we will restrict ourselves to the case of $\mathbb{R} \times \mathbb{R}$.

In the sequel we fix $1 < p < \infty$ and dyadic biparameter A_p weights μ, λ on $\mathbb{R}^2$, and we set $\nu := \mu^{1/p}\lambda^{-1/p}$. Note that we will be systematically suppressing from the notation dependence of implied constants on the value of p and the Muckenhoupt characteristics $[\mu]_{A_p,\boldsymbol{D}}$ and $[\lambda]_{A_p,\boldsymbol{D}}$. Recall that ν is a dyadic biparameter A_2 weight with $1 \leqslant [\nu]_{A_2} \leqslant [\mu]_{A_p}^{1/p}[\lambda]_{A_p}^{1/p}$.

Given any sequence $a = \{a_R\}_{R\in\boldsymbol{D}}$ of complex numbers, we define the *dyadic two-weight Bloom product BMO p-norm* $\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)}$ by

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)} := \sup_{\mathcal{U}} \frac{1}{(\mu(\mathrm{sh}(\mathcal{U})))^{1/p}} \left\| \left(\sum_{R\in\mathcal{U}} |a_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p(\lambda)},$$

where the supremum ranges over all non-empty collections $\mathcal{U}$ of dyadic rectangles in $\mathbb{R} \times \mathbb{R}$, and

$$\mathrm{sh}(\mathcal{U}) := \bigcup_{R\in\mathcal{U}} R.$$

By the Monotone Convergence Theorem, it is clear that one can restrict the supremum in the definition of $\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)}$ to just non-empty finite sucollections $\mathcal{U}$ of $\boldsymbol{D}$. Observe that if $\mathcal{U}$ is a non-empty finite collection of dyadic rectangles in $\mathbb{R} \times \mathbb{R}$, then $\Omega := \mathrm{sh}(\mathcal{U})$ is a bounded measurable subset of $\mathbb{R}^2$ of non-zero measure, and moreover, since the measure $\mu(x)\mathrm{d}x$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^2$, it follows that there exists a sequence $(\Omega_n)_{n=1}^{\infty}$ of bounded open subsets in $\mathbb{R}^2$ such that $\Omega \subseteq \Omega_n$, for all $n = 1, 2, \ldots$ and $\lim_{n\to\infty} \mu(\Omega_n) = \mu(\Omega)$. This shows that we have in fact

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)} = \sup_{\Omega} \frac{1}{(\mu(\Omega))^{1/p}} \left\| \left(\sum_{R\in\boldsymbol{D}(\Omega)} |a_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p(\lambda)},$$

where the supremum is taken over all non-empty bounded open sets Ω in $\mathbb{R}^2$, or even over all measurable subsets Ω of $\mathbb{R}^2$ of non-zero finite measure.

For any dyadic biparameter A_2 weight ν on $\mathbb{R}^2$, define the *dyadic one-weight Bloom product BMO norm* $\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)}$ by

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)} := \|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu,\nu^{-1},2)}.$$

We also define the *unweighted product BMO norm*

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}} := \|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(1)}.$$

If $b \in L^1_{\mathrm{loc}}(\mathbb{R})$, then we define

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)} := \|\{b_R\}_{R\in\boldsymbol{D}}\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)}.$$

The main goal of this section is to prove Theorem 1.2, which we state again for the reader's convenience, in a slightly more general (but in view of the Monotone Convergence Theorem, equivalent) form.

Theorem 1.2. *Let $1 < p < \infty$. Consider dyadic biparameter A_p weights μ, λ and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)} \sim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)},$$

for any sequence of complex numbers $a = \{a_R\}_{R\in\boldsymbol{D}}$, where the implied constants depend only on p, $[\mu]_{A_p,\boldsymbol{D}}$ and $[\lambda]_{A_p,\boldsymbol{D}}$.

The one-parameter analogue of Theorem 1.2 was shown by I. Holmes, M. T. Lacey and B. D. Wick [18]. Moreover, the little bmo analogue of Theorem 1.2 was established by I. Holmes, S. Petermichl and B. D. Wick [19] by iterating the one-parameter result of [18]. Also, in the special case $\mu = \lambda$, Theorem 1.2 was proved by E. Airta, K. Li, H. Martikainen and E. Vuorinen [2] (in fact under the weaker assumption that $\mu = \lambda$ is just a biparameter A_∞ weight).

The proof of Theorem 1.2 will be done in several steps. Note that by the Monotone Convergence Theorem, we can without loss of generality assume that the sequence $a = \{a_R\}_{R\in\boldsymbol{D}}$ is finitely supported.

Now, consider the 'purely non-cancellative' biparameter paraproduct with symbol a,

$$\Pi_a^{(1,1)} f := \sum_{R\in\boldsymbol{D}} a_R \langle f\rangle_R h_R.$$

Note that the superscript $(1,1)$ indicates that in the expression of the paraproduct f is integrated only against Haar functions of the form $h_R^{(1,1)}$. This operator is defined and used by Holmes–Lacey–Wick [18] in the Bloom one-parameter setting, as well as by Blasco–Pott [4] in the unweighted biparameter setting (for $p = 2$) and Holmes–Petermichl–Wick [19] in the Bloom biparameter setting.

First note that just by testing $\Pi_a^{(1,1)}$ on $\mathbf{1}_{\mathrm{sh}(\mathcal{U})}$ and then using the weighted Littlewood–Paley estimates, for any $\mathcal{U} \subseteq \boldsymbol{D}$, we trivially deduce

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\mu,\lambda,p)} \lesssim \|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)}. \tag{3.1}$$

We first show that the previous two quantities are actually equivalent in the regime $1 < p \leqslant 2$. To this end, we follow a scheme similar to that of one of the standard proofs of the dyadic Carleson's embedding theorem (see, for instance, [31] for a very general one-parameter version), but with the one-parameter maximal function replaced by the dyadic strong maximal function.

Proposition 3.1. *Assume* $1 < p \leqslant 2$. *Then, there holds*

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\mu,\lambda,p)} \sim \|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)},$$

where the implied constants depend only on p, $[\mu]_{A_p,\boldsymbol{\mathcal{D}}}$ *and* $[\lambda]_{A_p,\boldsymbol{\mathcal{D}}}$.

Proof. We only need to show the $\gtrsim$ direction, that is, the inequality

$$\|\Pi_a^{(1,1)}(f)\|_{L^p(\lambda)} \lesssim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\mu,\lambda,p)}\|f\|_{L^p(\mu)}.$$

By the weighted Littlewood–Paley estimates, we have

$$\|\Pi_a^{(1,1)}(f)\|_{L^p(\lambda)}^p \sim \int_{\mathbb{R}\times\mathbb{R}} \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |a_R|^2 |\langle f\rangle_R|^2 \frac{\mathbf{1}_R(x)}{|R|}\right)^{p/2} \lambda(x)\mathrm{d}x.$$

Thus, without loss of generality we may assume that $f \geqslant 0$.

Fix $x \in \mathbb{R}\times\mathbb{R}$. Consider the measure m on the countable set $\boldsymbol{\mathcal{D}}$ given by

$$m(R) := |a_R|^2 \frac{\mathbf{1}_R(x)}{|R|}, \qquad \forall R \in \boldsymbol{\mathcal{D}}.$$

Consider also the function $g : \boldsymbol{\mathcal{D}} \to [0,\infty)$ given by $g(R) := (\langle f\rangle_R)^p$, for all $R \in \boldsymbol{\mathcal{D}}$. Then, we have

$$\left(\sum_{R\in\boldsymbol{\mathcal{D}}} |a_R|^2 (\langle f\rangle_R)^2 \frac{\mathbf{1}_R(x)}{|R|}\right)^{p/2} = \|g\|_{L^{2/p}(\boldsymbol{\mathcal{D}},m)}.$$

We emphasize here that $2/p \geqslant 1$. Thus, in the scale of Lorentz spaces, we have (see, for example, [17, Proposition 1.4.10])

$$L^{2/p,1} \subseteq L^{2/p,2/p} = L^{2/p}.$$

Therefore

$$\left(\sum_{R\in\boldsymbol{\mathcal{D}}} |a_R|^2 (\langle f\rangle_R)^2 \frac{\mathbf{1}_R(x)}{|R|}\right)^{p/2} = \|g\|_{L^{2/p}(\boldsymbol{\mathcal{D}},m)} \lesssim_p \|g\|_{L^{2/p,1}(\boldsymbol{\mathcal{D}},m)}$$

$$\sim_p \int_0^\infty (m(\{g > t\}))^{p/2}\mathrm{d}t = \int_0^\infty \left(\sum_{R\in\boldsymbol{\mathcal{D}}\left(\langle f\rangle_R\right)^p > t} |a_R|^2 \frac{\mathbf{1}_R(x)}{|R|}\right)^{p/2} \mathrm{d}t.$$

Therefore

$$\begin{aligned}
\|\Pi_a^{(1,1)}(f)\|_{L^p(\lambda)}^p &\lesssim_p \int_{\mathbb{R}\times\mathbb{R}} \lambda(x) \int_0^\infty \left(\sum_{R\in\mathcal{D}\left(\langle f\rangle_R\right)^p>t} |a_R|^2 \frac{\mathbf{1}_R(x)}{|R|} \right)^{p/2} \mathrm{d}t\mathrm{d}x \\
&\leqslant \int_0^\infty \int_{\mathbb{R}\times\mathbb{R}} \left(\sum_{\substack{R\in\mathcal{D} \\ R\subseteq\{(M_{\mathcal{D}}f)^p>t\}}} |a_R|^2 \frac{\mathbf{1}_R(x)}{|R|} \right)^{p/2} \lambda(x)\mathrm{d}x\mathrm{d}t \\
&\leqslant \|a\|^p_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} \int_0^\infty \mu(\{(M_{\mathcal{D}}f)^p > t\})\mathrm{d}t \\
&= \|a\|^p_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} \|(M_{\mathcal{D}}f)^p\|_{L^1(\mu)} = \|a\|^p_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} \|M_{\mathcal{D}}f\|^p_{L^p(\mu)} \\
&\lesssim \|a\|^p_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} \|f\|^p_{L^p(\mu)},
\end{aligned}$$

concluding the proof. □

Note that by applying the Monotone Convergence Theorem coupled with the weighted Littlewood–Paley estimates, Proposition 3.1 extends to all (not necessarily finitely supported) sequences.

Holmes–Petermichl–Wick [19, Proposition 6.1] prove that

$$\|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)} \lesssim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)},$$

where $\nu := \mu^{1/p}\lambda^{-1/p}$ (which is an A_2 weight), $1 < p < \infty$ (in the general multiparameter case, the analogous result is due to Airta [1]). From this and (3.1) it follows that for all $1 < p < \infty$ we have

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} \lesssim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)}.$$

Since also $\nu = (\lambda')^{1/p'}(\mu')^{-1/p'}$, where $\mu' := \mu^{-1/(p-1)}$, $\lambda' := \lambda^{-1/(p-1)}$, we deduce as well

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\lambda',\mu',p')} \lesssim \|\Pi_a^{(1,1)}\|_{L^{p'}(\lambda')\to L^{p'}(\mu')} \lesssim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)}.$$

We show now that $\|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)}$ and $\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)}$ are actually equivalent, for all $1 < p < \infty$.

Proposition 3.2. *Let $1 < p < \infty$ and dyadic biparameter A_p weights μ and λ on $\mathbb{R}\times\mathbb{R}$. Define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then, there holds*

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)} \sim \|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)} \sim \|\Pi_a^{(1,1)}\|_{L^{p'}(\lambda')\to L^{p'}(\mu')},$$

where the implied constants depend only on p, $[\mu]_{A_p,\mathcal{D}}$ and $[\lambda]_{A_p,\mathcal{D}}$.

Proof. Recall that since μ is a dyadic biparameter A_p weight, $\mu' := \mu^{-1/(p-1)}$ is a dyadic biparameter $A_{p'}$ weight with $[\mu']_{A_{p'},\boldsymbol{\mathcal{D}}} = [\mu]_{A_p,\boldsymbol{\mathcal{D}}}^{p'-1}$. Similarly, $\lambda' := \lambda^{-1/(p-1)}$ is a dyadic biparameter $A_{p'}$ weight with $[\lambda']_{A_{p'},\boldsymbol{\mathcal{D}}} = [\lambda]_{A_p,\boldsymbol{\mathcal{D}}}^{p'-1}$, and ν^{-1} is a dyadic biparameter A_2 weight with $[\nu^{-1}]_{A_2,\boldsymbol{\mathcal{D}}} = [\nu]_{A_2,\boldsymbol{\mathcal{D}}}$. In particular, μ', λ', ν^{-1} are also locally integrable. Using these observations and since the sequence a is finitely supported, it is easy to see that

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu)},\ \|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)},\ \|\Pi_a^{(1,1)}\|_{L^{p'}(\lambda')\to L^{p'}(\mu')} < \infty.$$

For brevity we set $C(a) := \|a\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu)}$. By the weighted Littlewood–Paley estimates we have that

$$\|\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)} \sim C_1(a), \qquad \|\Pi_a^{(1,1)}\|_{L^{p'}(\lambda')\to L^{p'}(\mu')} \sim C_2(a),$$

where $C_1(a), C_2(a)$ are the best finite non-negative constants such that

$$\left\|\left(\sum_{R\in\boldsymbol{\mathcal{D}}} |a_R|^2 \langle |f|\rangle_R^2 \frac{\mathbf{1}_R}{|R|}\right)^{1/2}\right\|_{L^p(\lambda)} \leqslant C_1(a)\|f\|_{L^p(\mu)},$$

$$\left\|\left(\sum_{R\in\boldsymbol{\mathcal{D}}} |a_R|^2 \langle |f|\rangle_R^2 \frac{\mathbf{1}_R}{|R|}\right)^{1/2}\right\|_{L^{p'}(\mu')} \leqslant C_2(a)\|f\|_{L^{p'}(\lambda')},$$

for all measurable functions f on $\mathbb{R}\times\mathbb{R}$, that is

$$C_1(a) = \|S_{\boldsymbol{\mathcal{D}}}\Pi_a^{(1,1)}\|_{L^p(\mu)\to L^p(\lambda)}, \qquad C_2(a) = \|S_{\boldsymbol{\mathcal{D}}}\Pi_a^{(1,1)}\|_{L^{p'}(\lambda')\to L^{p'}(\mu')},$$

where $S_{\boldsymbol{\mathcal{D}}}$ is the dyadic biparameter square function, see Subsection 2.4 above. Thus, it suffices to prove that

$$C(a) \sim C_1(a) \sim C_2(a), \tag{3.2}$$

where all implied constants depend on $p, [\mu]_{A_p,\boldsymbol{\mathcal{D}}}, [\lambda]_{A_p,\boldsymbol{\mathcal{D}}}$. We will use *bilinear* estimates. Fix any $\mathcal{U} \subseteq \boldsymbol{\mathcal{D}}$. Let f, g be any two measurable functions on $\mathbb{R}\times\mathbb{R}$ taking non-negative values, and consider the bilinear form

$$B(f,g) := \sum_{R\in\mathcal{U}} |a_R|^2 \langle f\rangle_R \langle g\rangle_R \langle \nu^{-1}\rangle_R.$$

Clearly $B(f,g) = \int_{\mathbb{R}\times\mathbb{R}} F(x)\nu^{-1}(x)\mathrm{d}x$, where

$$F := \sum_{R\in\mathcal{U}} |a_R|^2 \langle f\rangle_R \langle g\rangle_R \frac{\mathbf{1}_R}{|R|}.$$

By Cauchy–Schwarz and Hölder's inequality we obtain

$$
\begin{aligned}
B(f,g) &= \int_{\mathbb{R}\times\mathbb{R}} F(x)\nu^{-1}(x)\mathrm{d}x \\
&\leqslant \int_{\mathbb{R}\times\mathbb{R}} \left(\sum_{R\in\mathcal{U}} |a_R|^2(\langle f\rangle_R)^2\frac{\mathbf{1}_R(x)}{|R|}\right)^{1/2}\left(\sum_{R\in\mathcal{U}_n} |a_R|^2(\langle g\rangle_R)^2\frac{\mathbf{1}_R(x)}{|R|}\right)^{1/2}\nu^{-1}(x)\mathrm{d}x \\
&= \int_{\mathbb{R}\times\mathbb{R}} \left(\sum_{R\in\mathcal{U}} |a_R|^2(\langle f\rangle_R)^2\frac{\mathbf{1}_R(x)}{|R|}\right)^{1/2}\lambda^{1/p}(x)\left(\sum_{R\in\mathcal{U}} |a_R|^2(\langle g\rangle_R)^2\frac{\mathbf{1}_R(x)}{|R|}\right)^{1/2}\mu^{-1/p}(x)\mathrm{d}x \\
&\leqslant \left(\int_{\mathbb{R}\times\mathbb{R}} \left(\sum_{R\in\mathcal{U}} |a_R|^2(\langle f\rangle_R)^2\frac{\mathbf{1}_R(x)}{|R|}\right)^{p/2}\lambda(x)\mathrm{d}x\right)^{1/p} \\
&\cdot \left(\int_{\mathbb{R}\times\mathbb{R}} \left(\sum_{R\in\mathcal{U}} |a_R|^2(\langle g\rangle_R)^2\frac{\mathbf{1}_R(x)}{|R|}\right)^{p'/2}\mu^{-p'/p}(x)\mathrm{d}x\right)^{1/p'} \\
&= \|S_{\boldsymbol{\mathcal{D}}}\Pi_a^{(1,1)}f\|_{L^p(\lambda)}\|S_{\boldsymbol{\mathcal{D}}}\Pi_a^{(1,1)}g\|_{L^{p'}(\mu')} \\
&\leqslant C_1(a)C_2(a)\|f\|_{L^p(\mu)}\|g\|_{L^{p'}(\lambda')}.
\end{aligned}
$$

We now pick

$$f := \mu^{-1/p}\nu^{1/p}\mathbf{1}_{\mathrm{sh}(\mathcal{U})}, \qquad g := \lambda^{1/p}\nu^{1/p'}\mathbf{1}_{\mathrm{sh}(\mathcal{U})}.$$

We have

$$\|f\|_{L^p(\mu)}^p = \int_{\mathrm{sh}(\mathcal{U})} \mu^{-1}(x)\nu(x)\mu(x)\mathrm{d}x = \nu(\mathrm{sh}(\mathcal{U})),$$

$$\|g\|_{L^{p'}(\lambda')}^{p'} = \int_{\mathrm{sh}(\mathcal{U})} \lambda^{p'/p}(x)\nu(x)\lambda^{-1/(p-1)}(x)\mathrm{d}x = \nu(\mathrm{sh}(\mathcal{U})),$$

therefore

$$\|f\|_{L^p(\mu)}\|g\|_{L^{p'}(\lambda')} = \nu(\mathrm{sh}(\mathcal{U})).$$

So we have

$$\sum_{R\in\mathcal{U}} |a_R|^2\langle\mu^{-1/p}\nu^{1/p}\rangle_R\langle\lambda^{1/p}\nu^{1/p'}\rangle_R\langle\nu^{-1}\rangle_R \leqslant C_1(a)C_2(a)\nu(\mathrm{sh}(\mathcal{U})).$$

Set $w := \mu^{-1/p}\nu^{1/p}$. Then clearly

$$\mu^{-1/p}\nu^{1/p}\lambda^{1/p}\nu^{1/p'} = 1,$$

so an immediate application of Jensen's inequality (with exponent $-1 < 0$) gives

$$\langle \mu^{-1/p} \nu^{1/p} \rangle_R \langle \lambda^{1/p} \nu^{1/p'} \rangle_R = \langle w \rangle_R \langle w^{-1} \rangle_R \geqslant 1.$$

Thus

$$\sum_{R \in \mathcal{U}} |a_R|^2 \langle \nu^{-1} \rangle_R \leqslant C_1(a) C_2(a) \nu(\mathrm{sh}(\mathcal{U})).$$

It follows that $C(a)^2 \leqslant C_1(a)C_2(a)$. Since we already know that $C(a) \gtrsim C_i(a),\ i = 1, 2$, we deduce $C(a) \sim C_1(a) \sim C_2(a)$. □

Note that by applying the Monotone Convergence Theorem coupled with the weighted Littlewood–Paley estimates, Proposition 3.2 extends to all (not necessarily finitely supported) sequences.

Combining Proposition 3.1 and 3.2, we already deduce that if $1 < p \leqslant 2$, then

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)} \sim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)} \gtrsim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\lambda',\mu',p')},$$

and that if $2 \leqslant p < \infty$, then

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)} \sim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\lambda',\mu',p')} \gtrsim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)}.$$

The next proposition will complete the proof of Theorem 1.2. Note that this remaining equivalence is an immediate consequence of Hölder's inequality in the unweighted case. In order to circumvent the lack of such a tool, we will make use of the (biparameter) Triebel–Lizorkin square function (see Subsection 2.4.2). In addition, it is also essential to use an equivalence between one-weight and unweighted product BMO from [2]. In particular, the equivalence from [2] that we use corresponds to the particular case of our Theorem 1.2 when $p > 2$ and $\mu = \lambda$.

Proposition 3.3. *Let a, p, μ, λ, ν be as above. Assume $p > 2$. Then, there holds*

$$\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)} \lesssim \|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\mu,\lambda,p)},$$

where the implied constant depends only on p, $[\mu]_{A_p,\mathcal{D}}$ and $[\lambda]_{A_p,\mathcal{D}}$.

Proof. First of all, note that we have already proved that

$$\|c\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu,1,2)} \lesssim \|c\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(1,\nu^{-1},2)},$$

for *any* sequence $c = \{c_R\}_{R \in \mathcal{D}}$. Applying this for the sequence $c = \{c_R := a_R \langle \nu^{-1} \rangle_R^{1/2}\}_{R \in \mathcal{D}}$, we deduce

$$\begin{aligned}\|a\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu)} &= \|\{a_R \langle \nu^{-1} \rangle_R^{1/2}\}_{R \in \mathcal{D}}\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu,1,2)} \lesssim \|\{a_R \langle \nu^{-1} \rangle_R^{1/2}\}_{R \in \mathcal{D}}\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(1,\nu^{-1},2)} \\ &= \|\{a_R \langle \nu^{-1} \rangle_R\}_{R \in \mathcal{D}}\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}}.\end{aligned}$$

Since μ is a dyadic biparameter A_p weight and $p > 2$, using the equivalence between one-weight and unweighted product BMO from the second half of the proof of [2, Theorem 3.2], we have

$$\|c\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}} \lesssim \sup_{\mathcal{U}\subseteq\mathcal{D}} \frac{1}{(\mu(\mathrm{sh}(\mathcal{U})))^{1/2}} \left\| \left(\sum_{R\in\mathcal{U}} |c_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^2(\mu)},$$

where the implied constant depends only on p and $[\mu]_{A_p,\mathcal{D}}$, for any sequence $c = \{c_R\}_{R\in\mathcal{D}}$, so since $p > 2$, by Hölder's inequality we deduce

$$\|c\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}} \lesssim \sup_{\mathcal{U}\subseteq\mathcal{D}} \frac{1}{(\mu(\mathrm{sh}(\mathcal{U})))^{1/p}} \left\| \left(\sum_{R\in\mathcal{U}} |c_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p(\mu)}.$$

Thus

$$\|\{a_R\langle\nu^{-1}\rangle_R\}_{R\in\mathcal{D}}\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}} \lesssim \sup_{\mathcal{U}\subseteq\mathcal{D}} \frac{1}{(\mu(\mathrm{sh}(\mathcal{U})))^{1/p}} \left\| \left(\sum_{R\in\mathcal{U}} |a_R\langle\nu^{-1}\rangle_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p(\mu)}.$$

For all $\mathcal{U} \subseteq \mathcal{D}$, using Lemma 2.1 and Corollary 2.2 (since a is finitely supported) and the fact that $\langle\nu^{-1}\rangle_R \sim \langle\mu\rangle_R^{-1/p}\langle\lambda\rangle_R^{1/p}$, for all $R \in \mathcal{D}$, we get

$$\begin{aligned} &\left\| \left(\sum_{R\in\mathcal{U}} |a_R\langle\nu^{-1}\rangle_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p(\mu)} \\ &\quad\sim \left\| \left(\sum_{R\in\mathcal{U}} |a_R|^2 \langle\nu^{-1}\rangle_R^2 \langle\mu\rangle_R^{2/p} \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p} \\ &\quad\sim \left\| \left(\sum_{R\in\mathcal{U}} |a_R|^2 \langle\lambda\rangle_R^{2/p} \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p} \sim \left\| \left(\sum_{R\in\mathcal{U}} |a_R|^2 \frac{\mathbf{1}_R}{|R|} \right)^{1/2} \right\|_{L^p(\lambda)}, \end{aligned}$$

concluding the proof. □

4 | ESTIMATES FOR ITERATED COMMUTATORS OF HAAR MULTIPLIERS

Let Σ be the set of all finitely supported maps $\sigma : \mathcal{D} \to \{-1, 0, 1\}$. In the sequel, the elements of Σ similar spaces will be called *sign choices*, and will always be considered to be finitely supported. For each $\sigma \in \Sigma$, we consider the constant coefficients *Haar multiplier* T_σ, which we might also call *martingale transform*, on the real line given by

$$T_\sigma f := \sum_{I\in\mathcal{D}} \sigma(I) f_I h_I, \qquad f \in L^1_{\mathrm{loc}}(\mathbb{R}).$$

Moreover, we consider Haar multipliers T^1_σ, T^2_σ acting on functions $f \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$ separately in each of the two variables, namely

$$T^1_\sigma f(t,s) := T_\sigma(f(\,\cdot\,,s))(t), \qquad T^2_\sigma f(t,s) := T_\sigma(f(t,\,\cdot\,))(s),$$

for a.e. $(t,s) \in \mathbb{R}^2$. It is clear that

$$T^1_\sigma f = \sum_{I,J\in D} \sigma(I) f_{I\times J} h_{I\times J} = \sum_{I\in D} \sigma(I) Q^1_I f,$$

$$T^2_\sigma f = \sum_{I,J\in D} \sigma(J) f_{I\times J} h_{I\times J} = \sum_{J\in D} \sigma(J) Q^2_J f.$$

The main result of this section is Theorem 1.1, which we recall here.

Theorem 1.1. *Let $1 < p < \infty$. Consider a function $b \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, dyadic biparameter A_p weights μ, λ and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\sup_{\sigma_1,\sigma_2\in\Sigma} \|[T^1_{\sigma_1},[T^2_{\sigma_2},b]]\|_{L^p(\mu)\to L^p(\lambda)} \sim \|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)},$$

where the implied constants depend only on p, $[\mu]_{A_p,\boldsymbol{D}}$ and $[\lambda]_{A_p,\boldsymbol{D}}$.

In the unweighted case with $p = 2$, Theorem 1.1 is proved by Blasco–Pott [4]. They get this result by averaging over the set of sign choices and then using orthogonality in Hilbert spaces. While we still rely on averaging over sign choices in our proof, we use a multiparameter extension of Khintchine's inequalities as well as vector-valued estimates to compensate for the lack of orthogonality for $p \neq 2$.

In the sequel we fix $1 < p < \infty$, $b \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, dyadic biparameter A_p weights μ, λ on $\mathbb{R}^2$, and we set $\nu := \mu^{1/p}\lambda^{-1/p}$. Note that we will be systematically suppressing from the notation dependence of implied constants on the value of p and the Muckenhoupt characteristics $[\mu]_{A_p,\boldsymbol{D}}$ and $[\lambda]_{A_p,\boldsymbol{D}}$.

4.1 | Relating Haar multipliers to a 'symmetrized' paraproduct

Consider the 'purely non-cancellative' and 'purely cancellative', respectively, biparameter paraproducts

$$\Pi^{(1,1)}_b f := \sum_{R\in\boldsymbol{D}} b_R \langle f\rangle_R h_R, \qquad \Pi^{(0,0)}_b f := \sum_{R\in\boldsymbol{D}} b_R f_R \frac{\mathbf{1}_R}{|R|},$$

and the 'mixed non-cancellative–cancellative' biparameter paraproducts

$$\Pi^{(1,0)}_b f := \sum_{R\in\boldsymbol{D}} b_R f^{(1,0)}_R h^{(0,1)}_R, \qquad \Pi^{(0,1)}_b f := \sum_{R\in\boldsymbol{D}} b_R f^{(0,1)}_R h^{(1,0)}_R.$$

In the notation of each paraproduct, the superscript indicates the type of Haar functions against which the argument of the paraproduct is integrated, while the 'complementary' pair of indices indicates the type of Haar functions appearing directly in the paraproduct.

Blasco–Pott [4] consider the 'symmetrized' paraproduct in the biparameter setting

$$\Lambda_b := \Pi_b^{(1,1)} + \Pi_b^{(0,0)} + \Pi_b^{(1,0)} + \Pi_b^{(0,1)}.$$

Blasco–Pott [4] prove via direct computation that Λ_b deserves to be called 'symmetrized' paraproduct in the sense that

$$\Lambda_b f = \sum_{R \in \mathcal{D}} (P_R b) f_R h_R, \tag{4.1}$$

where we recall that

$$P_R b = \sum_{R' \in \mathcal{D}(R)} b_{R'} h_{R'}.$$

Note that in the general multiparameter setting of $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$ one can also define an analogous operator $\boldsymbol{\Lambda}_b$ as a sum of generalized paraproducts in a way that one still has $\boldsymbol{\Lambda}_b f = \sum_{R \in \boldsymbol{\mathcal{D}}} (\boldsymbol{P}_R b) f_R h_R$, where $\boldsymbol{P}_R$ denotes the multiparameter analogue to P_R (and, abusing the notation, $\boldsymbol{\mathcal{D}}$ denotes the set of all dyadic rectangles in the product space $\mathbb{R}^{\vec{d}}$).

It is important to note that for any $R = I \times J \in \boldsymbol{\mathcal{D}}$, one has

$$P_R b(t,s) = (b - \langle b(\,\cdot\,, s)\rangle_I - \langle b(t, \,\cdot\,)\rangle_J + \langle b \rangle_{I \times J}) \mathbf{1}_R(t,s), \tag{4.2}$$

for a.e. $(t,s) \in \mathbb{R}^2$. Also, the same computation along the lines of the inclusion–exclusion principle that leads to (4.2) extends to yield a multiparameter analogue for $\boldsymbol{P}_R$. Using expressions (4.1) and (4.2), it is easy to see via direct computation, as remarked by Blasco–Pott [4], that

$$[Q_I^1, [Q_J^2, b]] = [Q_I^1, [Q_J^2, \Lambda_b]], \qquad \forall I, J \in \mathcal{D}, \tag{4.3}$$

and

$$Q_I^1 \Lambda_b Q_I^1 = 0, \qquad Q_J^2 \Lambda_b Q_J^2 = 0, \qquad \forall I, J \in \mathcal{D} \tag{4.4}$$

(and, as before, the analogues of these two expressions also hold in the multiparameter setting). In fact, the weighted Littlewood–Paley estimates imply that the family of all finite linear combinations of Haar functions h_R, $R \in \boldsymbol{\mathcal{D}}$ is dense in the weighted space $L^p(w)$, for any dyadic biparameter A_p weight w, $1 < p < \infty$, so one needs to check (4.3) and (4.4) only on functions of this type. Note that (4.3) immediately implies

$$[T_{\sigma_1}^1, [T_{\sigma_2}^2, b]] = [T_{\sigma_1}^1, [T_{\sigma_2}^2, \Lambda_b]], \qquad \forall \sigma_1, \sigma_2 \in \Sigma. \tag{4.5}$$

The following lemma contains one of the most important steps towards the proof of Theorem 1.1.

Lemma 4.1. *There holds*

$$\|\Lambda_b\|_{L^p(\mu)\to L^p(\lambda)} \sim \sup_{\sigma_1,\sigma_2\in\Sigma} \|[T^1_{\sigma_1},[T^2_{\sigma_2},b]]\|_{L^p(\mu)\to L^p(\lambda)}.$$

For the proof of Lemma 4.1 we will rely on a straightforward extension of Khintchine's inequalities to the multiparameter setting, which is of independent interest.

Lemma 4.2. *Let $(\mathbb{X}_i,\mathbb{P}_i)$, $i=1,2$ be probability spaces. For $j=1,2$, let $(X^i_j)_{j=1}^{N_i}$ be a Rademacher sequence on $(\mathbb{X}_i,\mathbb{P}_i)$, that is, X^i_j, $j=1,\dots,N_j$ are independent with*

$$\mathbb{P}_i(X^i_j=1)=\mathbb{P}_i(X^i_j=-1)=1/2,\qquad j=1,\dots,N_j.$$

Let A be an $N_1\times N_2$ (complex) matrix. Then, there holds

$$\left\|\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2} A(j_1,j_2)X^1_{j_1}\otimes X^2_{j_2}\right\|_{L^q(\mathbb{X}_1\times\mathbb{X}_2)} \sim_{q,r} \left\|\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2} A(j_1,j_2)X^1_{j_1}\otimes X^2_{j_2}\right\|_{L^r(\mathbb{X}_1\times\mathbb{X}_2)}$$

for all $0<q,r<\infty$. In particular

$$\left\|\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2} A(j_1,j_2)X^1_{j_1}\otimes X^2_{j_2}\right\|_{L^q(\mathbb{X}_1\times\mathbb{X}_2)} \sim_q \left(\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2}|A(j_1,j_2)|^2\right)^{1/2},\qquad \forall 0<q<\infty.$$

Proof. Let $0<q,r<\infty$ be arbitrary. Without loss of generality, we may assume $q<r$. Then, by Hölder's inequality, it suffices only to prove that

$$\left\|\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2} A(j_1,j_2)X^1_{j_1}\otimes X^2_{j_2}\right\|_{L^r(\mathbb{X}_1\times\mathbb{X}_2)} \lesssim_{q,r} \left\|\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2} A(j_1,j_2)X^1_{j_1}\otimes X^2_{j_2}\right\|_{L^q(\mathbb{X}_1\times\mathbb{X}_2)}.$$

Set

$$Y_{j_1}:=\sum_{j_2=1}^{N_2}A(j_1,j_2)X^2_{j_2},\qquad j_1=1,\dots,N_1,$$

$$Z_{j_2}:=\sum_{j_1=1}^{N_1}A(j_1,j_2)X^1_{j_1},\qquad j_2=1,\dots,N_2.$$

Then, using first Khintchine's inequalities, then Minkowski's inequality (in view of the fact that $r/q\geqslant 1$), and finally again Khintchine's inequalities, we get

$$\left\|\sum_{j_1=1}^{N_1}\sum_{j_2=1}^{N_2} A(j_1,j_2)X^1_{j_1}\otimes X^2_{j_2}\right\|^r_{L^r(\mathbb{X}_1\times\mathbb{X}_2)} = \int_{\mathbb{X}_2}\left(\int_{\mathbb{X}_1}\left|\sum_{j_1=1}^{N_1}Y_{j_1}(\omega_2)X^1_{j_1}(\omega_1)\right|^r \mathrm{d}\mathbb{P}_1(\omega_1)\right)\mathrm{d}\mathbb{P}_2(\omega_2)$$

$$\sim_{q,r}\int_{\mathbb{X}_2}\left(\int_{\mathbb{X}_1}\left|\sum_{j_1=1}^{N_1}Y_{j_1}(\omega_2)X^1_{j_1}(\omega_1)\right|^q \mathrm{d}\mathbb{P}_1(\omega_1)\right)^{r/q}\mathrm{d}\mathbb{P}_2(\omega_2)$$

$$\leqslant \left(\int_{\mathbb{X}_1} \left(\int_{\mathbb{X}_2} \left| \sum_{j_1=1}^{N_1} Y_{j_1}(\omega_2) X^1_{j_1}(\omega_1) \right|^{q \cdot \frac{r}{q}} \mathrm{d}\mathbb{P}_2(\omega_2) \right)^{q/r} \mathrm{d}\mathbb{P}_1(\omega_1) \right)^{r/q}$$

$$= \left(\int_{\mathbb{X}_1} \left(\int_{\mathbb{X}_2} \left| \sum_{j_2=1}^{N_2} Z_{j_2}(\omega_1) X^2_{j_2}(\omega_2) \right|^{r} \mathrm{d}\mathbb{P}_2(\omega_2) \right)^{q/r} \mathrm{d}\mathbb{P}_1(\omega_1) \right)^{r/q}$$

$$\sim_{q,r} \left(\int_{\mathbb{X}_1} \left(\int_{\mathbb{X}_2} \left| \sum_{j_2=1}^{N_2} Z_{j_2}(\omega_1) X^2_{j_2}(\omega_2) \right|^{q} \mathrm{d}\mathbb{P}_2(\omega_2) \right) \mathrm{d}\mathbb{P}_1(\omega_1) \right)^{r/q}$$

$$= \left\| \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} A(j_1, j_2) X^1_{j_1} \otimes X^2_{j_2} \right\|^r_{L^q(\mathbb{X}_1 \times \mathbb{X}_2)},$$

concluding the proof.

The second claim follows immediately from the first by just noting that an iteration of independence gives

$$\left\| \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} A(j_1, j_2) X^1_{j_1} \otimes X^2_{j_2} \right\|_{L^2(\mathbb{X}_1 \times \mathbb{X}_2)} = \left(\sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} |A(j_1, j_2)|^2 \right)^{1/2}. \qquad \square$$

Clearly, one can use induction to prove similarly a multiparameter version of Khintchine's inequalities, for any $0 < q, r < \infty$, in any number of parameters (one has just to replace any of the two uses of Khintchine's inequalities in the proof above by use of the inductive hypothesis).

Proof of Lemma 4.1. We first consider the direction $\gtrsim$. It is well known that Haar multipliers on $\mathbb{R}$ are bounded from $L^p(w)$ into $L^p(w)$ for dyadic A_p weights w on $\mathbb{R}$, within constants depending only on $[w]_{A_p,\mathcal{D}}$ (and not the sign choice in the definition of the Haar multiplier). It follows immediately that for any biparameter dyadic A_p weight w on $\mathbb{R} \times \mathbb{R}$ there holds

$$\|T^i_\sigma\|_{L^p(w) \to L^p(w)} \lesssim_{[w]_{A_p,\mathcal{D}}} 1, \qquad \forall i = 1, 2, \qquad \forall \sigma \in \Sigma,$$

therefore

$$\|[T^1_{\sigma_1}, [T^2_{\sigma_2}, b]]\|_{L^p(\mu) \to L^p(\lambda)} = \|[T^1_{\sigma_1}, [T^2_{\sigma_2}, \Lambda_b]]\|_{L^p(\mu) \to L^p(\lambda)}$$
$$= \|T^1_{\sigma_1} T^2_{\sigma_2} \Lambda_b - T^1_{\sigma_1} \Lambda_b T^2_{\sigma_2} - T^2_{\sigma_2} \Lambda_b T^1_{\sigma_1} + \Lambda_b T^2_{\sigma_2} T^1_{\sigma_1}\|_{L^p(\mu) \to L^p(\lambda)} \lesssim 4 \|\Lambda_b\|_{L^p(\mu) \to L^p(\lambda)}$$

for all $\sigma_1, \sigma_2 \in \Sigma$.

We now turn to the other direction. Let $(\mathcal{F}_n)_{n=1}^\infty$ be an increasing sequence of subsets of $\mathcal{D}$ exhausting $\mathcal{D}$. For each $n = 1, 2, \ldots$, let Σ_n be the set of all maps $\sigma : \mathcal{D} \to \{-1, 0, 1\}$ that vanish outside of $\mathcal{F}_n$ and that take values only $-1, 1$ on $\mathcal{F}_n$, and consider the natural probability measure $\mathbb{P}_n$ on Σ_n that to each coordinate $I \in \mathcal{F}_n$ assigns each of the values 1 and -1 with probability $1/2$, independently of all the other coordinates. Clearly, it suffices to prove that

$$\sup_{n=1,2,\ldots}\int_{\Sigma_n\times\Sigma_n}\|[T^1_{\sigma_1},[T^2_{\sigma_2},b]](f)\|^p_{L^p(\lambda)}\mathrm{d}(\mathbb{P}_n\otimes\mathbb{P}_n)(\sigma_1,\sigma_2)\gtrsim\|\Lambda_b(f)\|^p_{L^p(\lambda)}, \tag{4.6}$$

for all (suitable) functions f on $\mathbb{R}^2$. For brevity we set $C_{I\times J}:=[Q^1_I,[Q^2_J,b]]$. Applying Lemma 4.2, we have

$$\begin{aligned}
&\sup_{n=1,2,\ldots}\int_{\Sigma_n\times\Sigma_n}\|[T^1_{\sigma_1},[T^2_{\sigma_2},b]](f)\|^p_{L^p(\lambda)}\mathrm{d}(\mathbb{P}_n\otimes\mathbb{P}_n)(\sigma_1,\sigma_2)\\
&=\sup_{n=1,2,\ldots}\int_{\Sigma_n\times\Sigma_n}\Bigg\|\sum_{I\times J\in\boldsymbol{\mathcal{D}}}\sigma_1(I)\sigma_2(J)C_{I\times J}(f)\Bigg\|^p_{L^p(\lambda)}\mathrm{d}(\mathbb{P}_n\otimes\mathbb{P}_n)(\sigma_1,\sigma_2)\\
&=\sup_{n=1,2,\ldots}\int_{\mathbb{R}\times\mathbb{R}}\left(\int_{\Sigma_n\times\Sigma_n}\Bigg|\sum_{I\times J\in\boldsymbol{\mathcal{D}}}\sigma_1(I)\sigma_2(J)C_{I\times J}(f)(x)\Bigg|^p\mathrm{d}(\mathbb{P}_n\otimes\mathbb{P}_n)(\sigma_1,\sigma_2)\right)\lambda(x)\mathrm{d}x\\
&\sim_p\sup_{n=1,2,\ldots}\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I,J\in\mathcal{F}_n}|C_{I\times J}(f)(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x=\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|C_{I\times J}(f)(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x,
\end{aligned}$$

where in the last equality we applied the Monotone Convergence Theorem. Observe that

$$|Q_R g|\leqslant\langle|g|\rangle_R\mathbf{1}_R\leqslant M_{\boldsymbol{\mathcal{D}}}g,\qquad\forall R\in\boldsymbol{\mathcal{D}}.$$

Moreover, O. N. Capri and C. Gutiérrez [7] establish the following one-weight vector-valued estimate for the dyadic strong maximal function (in any number of parameters):

$$\left\|\left(\sum_{n=1}^{\infty}|M_{\boldsymbol{\mathcal{D}}}g_n|^2\right)^{1/2}\right\|_{L^p(w)}\lesssim\left\|\left(\sum_{n=1}^{\infty}|g_n|^2\right)^{1/2}\right\|_{L^p(w)},$$

where the implied constants depend only on p and $[w]_{A_p,\boldsymbol{\mathcal{D}}}$ (their proof is for the case of the strong maximal function M_S and multiparameter A_p weights w, but it works without any changes for the case of the dyadic strong maximal function $M_{\boldsymbol{\mathcal{D}}}$ and dyadic multiparameter A_p weights w). Thus, we have

$$\begin{aligned}
&\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|C_{I\times J}(f)(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x\gtrsim\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|M_{\boldsymbol{\mathcal{D}}}(C_{I\times J}(f))(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x\\
&\geqslant\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|Q^1_IQ^2_J(C_{I\times J}(f))(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x\\
&=\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|Q^1_IQ^2_J([Q^1_I,[Q^2_J,\Lambda_b]](f))(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x,
\end{aligned}$$

where in the last equality we have used (4.3). Notice that using (4.4) we obtain

$$\begin{aligned}&Q^1_I Q^2_J([Q^1_I,[Q^2_J,\Lambda_b]](f))\\&=Q^1_I Q^2_J(Q^1_I Q^2_J \Lambda_b(f)-Q^1_I\Lambda_b Q^2_J(f)-Q^2_J\Lambda_b Q^1_I(f)+\Lambda_b Q^1_I Q^2_J(f))\\&=Q^1_I Q^2_J \Lambda_b(f),\end{aligned}$$

for all $I,J\in\mathcal{D}$. It follows that

$$\begin{aligned}&\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|Q^1_I Q^2_J([Q^1_I,[Q^2_J,\Lambda_b]](f))(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x\\&=\int_{\mathbb{R}\times\mathbb{R}}\left(\sum_{I\times J\in\boldsymbol{\mathcal{D}}}|Q^1_I Q^2_J\Lambda_b(f)(x)|^2\right)^{p/2}\lambda(x)\mathrm{d}x\sim\|\Lambda_b(f)\|^p_{L^p(\lambda)},\end{aligned}$$

concluding the proof. □

4.2 | Bounds for the 'symmetrized' paraproduct and conclusion of the proof

In this section we complete the proof of Theorem 1.1. Blasco–Pott [4] show that

$$P_\Omega(b)=P_\Omega(\Lambda_b(\mathbf{1}_\Omega)).$$

This can be readily checked by direct computation using the definition of the operator Λ_b and how paraproducts act on characteristic functions. From this, it follows that

$$\|P_\Omega(b)\|_{L^p(\lambda)}=\|P_\Omega(\Lambda_b(\mathbf{1}_\Omega))\|_{L^p(\lambda)}\lesssim\|\Lambda_b(\mathbf{1}_\Omega)\|_{L^p(\lambda)}\leqslant\|\Lambda_b\|_{L^p(\mu)\to L^p(\lambda)}(\mu(\Omega))^{1/p}.$$

In the $\lesssim$ we used the weighted Littlewood–Paley estimates. The analogous expressions are also valid for the multiparameter operators $\boldsymbol{P}_\Omega$ and $\boldsymbol{\Lambda}_b$ and measurable sets $\Omega\subseteq\mathbb{R}^{\vec{d}}$, and their proofs use the same idea of checking the action of the various paraproducts on characteristic functions. It follows that

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\mu,\lambda,p)}\lesssim\|\Lambda_b\|_{L^p(\mu)\to L^p(\lambda)}.$$

Combining this with Theorem 1.2 and Lemma 4.1 we deduce

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu)}\lesssim\sup_{\sigma_1,\sigma_2\in\Sigma}\|[T^1_{\sigma_1},[T^2_{\sigma_2},b]]\|_{L^p(\mu)\to L^p(\lambda)}.\tag{4.7}$$

Moreover, Holmes–Petermichl–Wick [19] prove that

$$\|P_b\|_{L^p(\mu)\to L^p(\lambda)}\lesssim\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(\nu)},$$

where P_b is any of the four paraproducts $\Pi_b^{(\varepsilon_1,\varepsilon_2)}$, $\varepsilon_1, \varepsilon_2 \in \{0,1\}$ (the same fact for all relevant multiparameter paraproducts is shown by Airta in [1]). It follows that

$$\|\Lambda_b\|_{L^p(\mu)\to L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)}. \tag{4.8}$$

Thus, combining Lemma 4.1, (4.7) and (4.8) we deduce

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)} \sim \sup_{\sigma_1,\sigma_2\in\Sigma} \|[T^1_{\sigma_1},[T^2_{\sigma_2},b]]\|_{L^p(\mu)\to L^p(\lambda)},$$

concluding the proof.

Note that since ν is a dyadic biparameter A_2 weight on $\mathbb{R}^2$, $\nu = \nu^{1/2}(\nu^{-1})^{-1/2}$ and $[\nu]_{A_2,\boldsymbol{D}} = [\nu^{-1}]_{A_2,\boldsymbol{D}}$, we also get

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)} \sim \sup_{\sigma_1,\sigma_2\in\Sigma} \|[T^1_{\sigma_1},[T^2_{\sigma_2},b]]\|_{L^2(\nu)\to L^2(\nu^{-1})},$$

where all implied constants depend only on $[\nu]_{A_2,\boldsymbol{D}}$.

4.3 | General multiparameter result

The ideas presented in this section can also be applied, with only minor modifications, to any multiparameter setting. That is, Theorem 1.1 can be stated and proved for iterated commutators on functions defined on $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_t}$, $\vec{d} := (d_1,\dots,d_t)$. We have already been commenting along the proof which steps have to be modified in this context. Abusing slightly the notation, here we use $\boldsymbol{D}$ to denote the set of dyadic rectangles in the product space $\mathbb{R}^{\vec{d}}$. The statement of the result in full generality is the following.

Theorem 4.3. *Let $1 < p < \infty$. Consider a function $b \in L^1_{\mathrm{loc}}(\mathbb{R}^{\vec{d}})$, dyadic multiparameter A_p weights μ, λ on $\mathbb{R}^{\vec{d}}$ and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\sup_{\sigma_1,\dots,\sigma_t\in\Sigma} \|[T^1_{\sigma_1},[\dots[T^t_{\sigma_t},b]\dots]]\|_{L^p(\mu)\to L^p(\lambda)} \sim \|b\|_{\mathrm{BMO}_{\mathrm{prod},\boldsymbol{D}}(\nu)},$$

where the implied constants depend only on $\vec{d}$, p, $[\mu]_{A_p,\boldsymbol{D}}$ and $[\lambda]_{A_p,\boldsymbol{D}}$.

5 | BOUNDS FOR GENERAL COMMUTATORS OF HAAR MULTIPLIERS

In this section, we show bounds analogous to those of Section 4 for iterated commutators of general martingale transforms, not necessarily of tensor type. As usual, consider dyadic multiparameter A_p weights μ and λ on $\mathbb{R}^d$, with $1 < p < \infty$, and let $\nu := \mu^{1/p}\lambda^{-1/p}$. We define the *dyadic two-weight little BMO p-norm* $\|b\|_{\mathrm{bmo}_{\boldsymbol{D}}(\mu,\lambda,p)}$ by

$$\|b\|_{\mathrm{bmo}_{\boldsymbol{D}}(\mu,\lambda,p)} := \sup_{R\in\boldsymbol{D}} \frac{1}{(\mu(R))^{1/p}} \|(b-\langle b\rangle_R)\mathbf{1}_R\|_{L^p(\lambda)}.$$

We also say, for a dyadic multiparameter A_2 weight w on $\mathbb{R}^{\vec{d}}$, that a function b is in the *dyadic one-weight little BMO* space if

$$\sup_{R\in\mathcal{D}} \frac{1}{w(R)} \|(b-\langle b\rangle_R)\mathbf{1}_R\|_{L^1(\mathbb{R}^{\vec{d}})} < \infty.$$

In this case, we assign to function b the norm

$$\|b\|_{\mathrm{bmo}_{\mathcal{D}}(w)} := \|b\|_{\mathrm{bmo}_{\mathcal{D}}(w,w^{-1},2)},$$

which is equivalent to the previous supremum. This equivalence is shown by Holmes–Petermichl–Wick [19] in the biparameter case using an iteration of the one-parameter argument due to Holmes–Lacey–Wick [18], but the same argument can be iterated to any number of parameters. Furthermore, the authors of [19] also prove that, for μ, λ and ν as before, there holds

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

(it is actually shown there for the continuous biparameter setting, but their argument holds equally well in the dyadic case and can be iterated to any number of parameters as well).

Recall that if $x \in \mathbb{R}^{\vec{d}}$ and $k \in \{1,2,\dots,t\}$, we denote $x_{\overline{k}} := (x_1,\dots,x_{k-1},\cdot,x_{k+1},\dots,x_t)$, so that a function $f(x_{\overline{k}})$ depends only on the variable x_k. Similarly, we also denote $\mathbb{R}^{\vec{d}_{\overline{k}}} := \mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_{k-1}}\times\mathbb{R}^{d_{k+1}}\times\cdots\times\mathbb{R}^{d_t}$. In general, we extend this notation to any number of parameters, so that if $k_1,\dots,k_s \in \{1,2,\dots,t\}$, then we take $x_{\overline{k_1,\dots,k_s}} := (x_1,\dots,x_{k_1-1},\cdot,x_{k_1+1},\dots,x_{k_s-1},\cdot,x_{k_s+1},\dots,x_t)$, and similarly for $\mathbb{R}^{\vec{d}_{\overline{k_1,\dots,k_s}}}$. Using this notation, Holmes–Petermichl–Wick [19] show that

$$\|b\|_{\mathrm{bmo}_{\mathcal{D}}(\nu)} \sim \max\left\{ \operatorname*{ess\,sup}_{x_{\overline{k}}\in\mathbb{R}^{\vec{d}_{\overline{k}}}} \|b(x_{\overline{k}})\|_{\mathrm{BMO}_{\mathcal{D}}(\nu(x_{\overline{k}}))} :\, k\in\{1,2,\dots,t\}\right\},$$

where $\|\cdot\|_{\mathrm{BMO}_{\mathcal{D}}(\nu)}$ denotes the usual dyadic weighted one-parameter BMO norm (again, the result is stated and proved there only in the continuous biparameter setting, but the argument holds as well for the dyadic spaces and in any number of parameters). In other words, a function b is in dyadic weighted little bmo if and only if it is uniformly in dyadic-weighted one-parameter BMO in each variable. Moreover, they also observe that

$$\mathrm{bmo}_{\mathcal{D}}(\nu) \subseteq \mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu).$$

More generally, let $\mathcal{I} = \{I_1,\dots,I_l\}$ be any partition of $\{1,\dots,t\}$. Let μ,λ be t-parameter A_p weights on $\mathbb{R}^{\vec{d}}$, $1<p<\infty$. Define

$$\|b\|_{\mathrm{bmo}^{\mathcal{I}}_{\mathcal{D}}(\mu,\lambda,p)} := \max_{\iota\in I_1\times\cdots\times I_l} \left(\operatorname*{ess\,sup}_{x_{\bar{\iota}}\in\mathbb{R}^{\vec{d}_{\bar{\iota}}}} \|b(x_{\bar{\iota}})\|_{\mathrm{BMO}_{\mathrm{prod}_{\mathcal{D}}}(\mu(x_{\bar{\iota}}),\lambda(x_{\bar{\iota}}),p)}\right),$$

and as before if w is any t-parameter A_2 weight on $\mathbb{R}^{\vec{d}}$, then we define

$$\|b\|_{\mathrm{bmo}^{\mathcal{I}}_{\mathcal{D}}(w)} := \|b\|_{\mathrm{bmo}^{\mathcal{I}}_{\mathcal{D}}(w,w^{-1},2)}.$$

If $\mathcal{I}$ consists only of singletons, then one recovers product BMO in t parameters, while if $\mathcal{I}$ has just one element, then one recovers little BMO. In general, for any partition $\mathcal{I}$ we have that

$$\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(w) \subseteq \mathrm{bmo}^{\mathcal{I}}_{\boldsymbol{\mathcal{D}}}(w) \subseteq \mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}(w).$$

Observe that our results in Section 3 immediately imply that if $\nu := \mu^{1/p}\lambda^{-1/p}$, then

$$\|b\|_{\mathrm{bmo}^{\mathcal{I}}_{\boldsymbol{\mathcal{D}}}(\nu)} \sim \|b\|_{\mathrm{bmo}^{\mathcal{I}}_{\boldsymbol{\mathcal{D}}}(\mu,\lambda,p)},$$

where all implied constants depend only on p, $[\mu]_{A_p,\boldsymbol{\mathcal{D}}}$ and $[\lambda]_{A_p,\boldsymbol{\mathcal{D}}}$. Notice that here we are making use of the fact that for any $\iota \in I_1 \times \cdots \times I_l$, and for almost every $x_{\bar{\iota}} \in \mathbb{R}^{\vec{d}_{\bar{\iota}}}$, the weight $\mu(x_{\bar{\iota}})$ is a dyadic l-parameter A_p weight on $\mathbb{R}^{d_{i_1}} \times \cdots \times \mathbb{R}^{d_{i_l}}$ with $[\mu(x_{\bar{\iota}})]_{A_p,\boldsymbol{\mathcal{D}}} \leqslant [\mu]_{A_p,\boldsymbol{\mathcal{D}}}$, and similarly for λ and ν.

Consider now the set $\boldsymbol{\Sigma}$ of sign choices $\sigma = \{\sigma(R)\}_{R\in\boldsymbol{\mathcal{D}}}$. For each $\sigma \in \boldsymbol{\Sigma}$ we consider the Haar multiplier T_σ defined by

$$T_\sigma f := \sum_{R\in\boldsymbol{\mathcal{D}}} \sigma(R) f_R h_R, \qquad f \in L^2(\mathbb{R}^{\vec{d}}).$$

Observe that here we consider all possible choices of signs over the dyadic rectangles in $\boldsymbol{\mathcal{D}}$, not only those that are of tensor type, as opposed to Section 4. We will study bounds for $\|[T_\sigma, b]\|_{L^p(\mu)\to L^p(\lambda)}$ in terms of $\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}$. We adapt the arguments presented in Section 4 to the biparameter case.

Proposition 5.1. *Let $1 < p < \infty$. Consider a function $b \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, dyadic biparameter A_p weights μ, λ on $\mathbb{R}^2$ and define $\nu := \mu^{1/p}\lambda^{-1/p}$. Then*

$$\sup_{\sigma\in\boldsymbol{\Sigma}} \|[T_\sigma, b]\|_{L^p(\mu)\to L^p(\lambda)} \sim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)},$$

where the implied constants depend only on p, $[\mu]_{A_p,\boldsymbol{\mathcal{D}}}$ and $[\lambda]_{A_p,\boldsymbol{\mathcal{D}}}$.

Proof. Airta [1, Theorem 4.12] shows general upper bounds for iterated commutators of multiparameter Haar shifts in terms of the symbol norm in the appropriate indexed BMO space. In particular, for a single commutator, this includes the upper bound $\|[T_\sigma, b]\|_{L^p(\mu)\to L^p(\lambda)} \lesssim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}$. Thus, we only need to show the corresponding lower bound.

Define the operator Θ_b by

$$\Theta_b f = \sum_{R\in\boldsymbol{\mathcal{D}}} (\omega_R b) f_R h_R, \qquad f \in L^2(\mathbb{R}^2),$$

where $\omega_R b := (b(x) - \langle b\rangle_R)\mathbf{1}_R(x)$. This operator satisfies the relations

$$[Q_R, b] = [Q_R, \Theta_b], \qquad \forall R \in \boldsymbol{\mathcal{D}},$$

and

$$Q_R \Theta_b Q_R = 0, \qquad \forall R \in \boldsymbol{\mathcal{D}},$$

where Q_R denotes the orthogonal projection from $L^2(\mathbb{R}^2)$ onto the one-dimensional space spanned by h_R. Observe that a simple computation using the definition of Θ_b shows that they hold for any Haar function h_R, from which the general result follows by a density argument.

Now we see that

$$\sup_{\sigma\in\boldsymbol{\Sigma}} \|[T_\sigma, b]\|_{L^p(\mu)\to L^p(\lambda)} \gtrsim \|\Theta_b\|_{L^p(\mu)\to L^p(\lambda)}. \tag{5.1}$$

As before, the lower bound (5.1) will follow from that of the average of the commutator norms over $\boldsymbol{\Sigma}$. In this case we have

$$\begin{aligned}
&\int_{\boldsymbol{\Sigma}} \|[T_\sigma, b](f)\|^p_{L^p(\lambda)} \mathrm{d}\mathbb{P}(\sigma) \\
&= \int_{\mathbb{R}^2} \left(\int_{\boldsymbol{\Sigma}} \left| \sum_{R\in\boldsymbol{\mathcal{D}}} \sigma(R)[Q_R, b](f)(x) \right|^p \mathrm{d}\mathbb{P}(\sigma) \right) \lambda(x)\mathrm{d}x \\
&\sim \int_{\mathbb{R}^2} \left(\int_{\boldsymbol{\Sigma}} \left| \sum_{R\in\boldsymbol{\mathcal{D}}} \sigma(R)[Q_R, b](f)(x) \right|^2 \mathrm{d}\mathbb{P}(\sigma) \right)^{p/2} \lambda(x)\mathrm{d}x,
\end{aligned}$$

where we have used Khintchine's inequalities in the last step. This last quantity is equal to

$$\begin{aligned}
&= \int_{\mathbb{R}^2} \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |[Q_R, b](f)(x)|^2 \right)^{p/2} \lambda(x)\mathrm{d}x \\
&\gtrsim \int_{\mathbb{R}^2} \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |M_{\boldsymbol{\mathcal{D}}}[Q_R, b](f)(x)|^2 \right)^{p/2} \lambda(x)\mathrm{d}x \\
&\geqslant \int_{\mathbb{R}^2} \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |Q_R[Q_R, \Theta_b](f)(x)|^2 \right)^{p/2} \lambda(x)\mathrm{d}x \\
&= \int_{\mathbb{R}^2} \left(\sum_{R\in\boldsymbol{\mathcal{D}}} |Q_R\Theta_b(f)(x)|^2 \right)^{p/2} \lambda(x)\mathrm{d}x \sim \|\Theta_b(f)\|^p_{L^p(\lambda)}.
\end{aligned}$$

This shows (5.1).

Now we are only left with checking that

$$\|\Theta_b\|_{L^p(\mu)\to L^p(\lambda)} \gtrsim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}.$$

Observe that testing the operator on Haar functions we immediately get

$$\|\Theta_b\|_{L^p(\mu)\to L^p(\lambda)} \|h_R\|_{L^p(\mu)} \geqslant \|\Theta_b h_R\|_{L^p(\lambda)} = \|(b - \langle b\rangle_R)\mathbf{1}_R\|_{L^p(\lambda)} |R|^{-1/2}.$$

We also have that $\|h_R\|_{L^p(\mu)} = (\mu(R))^{1/p}|R|^{-1/2}$, so that

$$\|\Theta_b\|_{L^p(\mu)\to L^p(\lambda)} \geqslant \frac{1}{(\mu(R))^{1/p}} \|(b - \langle b\rangle_R)\mathbf{1}_R\|_{L^p(\lambda)}.$$

But the supremum of the right-hand side is precisely $\|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\mu,\lambda,p)} \sim \|b\|_{\mathrm{bmo}_{\boldsymbol{\mathcal{D}}}(\nu)}$. □

This method can also be adapted to general indexed BMO spaces, and thus to the general multiparameter little bmo case. We explain how to do it taking as an example the product space $\mathbb{R}^3 = \mathbb{R}\times\mathbb{R}\times\mathbb{R}$, and the partition $\mathcal{I} := \{I_1 := \{1,3\},\ I_2 := \{2\}\}$ of $\{1,2,3\}$. In this case, we consider the set $\boldsymbol{\Sigma}_{1,3}$ of sign choices $\sigma = \{\sigma(I\times J)\}_{I\times J\in\boldsymbol{\mathcal{D}}}$, and the set $\boldsymbol{\Sigma}_2$ of sign choices $\sigma = \{\sigma(I)\}_{I\in\boldsymbol{\mathcal{D}}}$. Given $\sigma_{1,3}\in\boldsymbol{\Sigma}_{1,3}$ and $\sigma_2\in\boldsymbol{\Sigma}_2$ consider the martingale transform $T^{1,3}_{\sigma_{1,3}}$ acting on variables 1 and 3 and the martingale transform $T^2_{\sigma_2}$ acting on variable 2. The previous method can be adapted to show that

$$\sup_{\sigma_{1,3}\in\boldsymbol{\Sigma}_{1,3},\ \sigma_2\in\Sigma_2} \|[T^{1,3}_{\sigma_{1,3}}, [T^2_{\sigma_2}, b]]\|_{L^p(\mu)\to L^p(\lambda)} \gtrsim \|b\|_{\mathrm{bmo}^{\mathcal{I}}_{\boldsymbol{\mathcal{D}}}(\nu)},$$

while the corresponding upper bound was already proved by Airta (see [1, Theorem 4.12]). To this end, define the operator Ξ_b by

$$\Xi_b f = \sum_{I,J,K\in D} (\xi_{I\times J\times K} b) f_{I\times J\times K} h_{I\times J\times K}, \qquad f\in L^2(\mathbb{R}^3),$$

where

$$\xi_{I\times J\times K} b = (b - \langle b\rangle^{1,3}_{I\times K} - \langle b\rangle^2_J + \langle b\rangle^{1,2,3}_{I\times J\times K})\mathbf{1}_{I\times J\times K}$$

and where we use $\langle b\rangle^{1,3}_{I\times K}$ to denote the average taken on variables 1 and 3 taken over $I\times K$ (similarly for $\langle b\rangle^2_J$). Then one can repeat the arguments in Lemma 4.1 to show that

$$\sup_{\sigma_{1,3}\in\boldsymbol{\Sigma}_{1,3},\ \sigma_2\in\Sigma_2} \|[T^{1,3}_{\sigma_{1,3}}, [T^2_{\sigma_2}, b]]\|_{L^p(\mu)\to L^p(\lambda)} \gtrsim \|\Xi_b\|_{L^p(\mu)\to L^p(\lambda)}.$$

The lower bound for the operator norm of Ξ_b follows a similar argument to that of Λ_b, with the difference that in this case one needs to check the $\mathrm{BMO}_{\mathrm{prod},\boldsymbol{\mathcal{D}}}$ norm in variables 1, 2 and in variables 2, 3. We focus on how to get the bound for variables 1 and 2, as the other case is done in the same way. Consider an arbitrary open set $\Omega\subseteq\mathbb{R}\times\mathbb{R}$ of non-zero finite measure and its characteristic function $\mathbf{1}_\Omega(x_1,x_2)$ in variables 1 and 2. Fix x_3 and take $K\in\mathcal{D}$ such that $x_3\in K$. Note that for $x\in\mathbb{R}\times\mathbb{R}\times\{x_3\}$ we trivially have $\Xi_{b(x_{\overline{1,2}})}(\mathbf{1}_\Omega\otimes h_K) = (\Lambda_{b(x_{\overline{1,2}})}\mathbf{1}_\Omega)\otimes h_K$. By testing the operator Ξ_b on $\mathbf{1}_\Omega\otimes h_K$ one gets

$$\begin{aligned}
&\frac{1}{|K|^{1-p/2}}\|\Xi_b(\mathbf{1}_\Omega\otimes h_K)\|^p_{L^p(\lambda)}\\
&= \int_{\mathbb{R}\times\mathbb{R}} \frac{1}{|K|}\int_K \Bigg|\sum_{I,J\in D}(b - \langle b\rangle^{1,3}_{I\times K} - \langle b\rangle^2_J + \langle b\rangle^{1,2,3}_{I\times J\times K})(\mathbf{1}_\Omega)_{I\times J}\Bigg|^p \lambda(x_1,x_2,x_3)\mathrm{d}x\\
&\leqslant \|\Xi_b\|^p_{L^p(\mu)\to L^p(\lambda)}\langle\mu(\Omega,x_3)\rangle^3_K.
\end{aligned}$$

Applying twice Lebesgue Differentiation Theorem and Fatou's lemma we get

$$\begin{aligned}
&\|\Lambda_{b(x_{\overline{1,2}})}\mathbf{1}_\Omega\|^p_{L^p(\lambda(x_{\overline{1,2}}))}\\
&= \int_{\mathbb{R}\times\mathbb{R}} \Bigg|\sum_{I,J\in D}(b(x_{\overline{1,2}}) - \langle b(x_{\overline{1,2}})\rangle^1_I - \langle b(x_{\overline{1,2}})\rangle^2_J + \langle b(x_{\overline{1,2}})\rangle^{1,2}_{I\times J})(\mathbf{1}_\Omega)_{I\times J}\Bigg|^p \lambda(x_{\overline{1,2}})\,\mathrm{d}x_1\,\mathrm{d}x_2\\
&\leqslant \lim_{K\to x_3}\|\Xi_b\|^p_{L^p(\mu)\to L^p(\lambda)}\langle\mu(\Omega,x_3)\rangle^3_K = \|\Xi_b\|^p_{L^p(\mu)\to L^p(\lambda)}\mu(\Omega,x_3),
\end{aligned}$$

where $K \to x_3$ denotes that the limit is taken through a sequence of intervals containing x_3 with side length tending to 0, and where the last equality holds at almost every x_3. Thus, by the bound (4.8) for operator $\Lambda_{b(x_{\overline{1,2}})}$ in terms of the dyadic one-weight biparameter product BMO norm, we get

$$\|\Xi_b\|_{L^p(\mu)\to L^p(\lambda)} \geqslant \operatorname*{ess\,sup}_{x_3} \|b(x_{\overline{1,2}})\|_{\mathrm{BMO}_{\mathrm{prod},\mathcal{D}}(\nu(x_{\overline{1,2}}))}.$$

The case of general commutators and indexed spaces can be worked out in a similar way, considering the appropriate multiparameter analogues of the operator Ξ_b, Lemma 4.1 and Equation (4.8).

ACKNOWLEDGEMENT

The authors would like to thank Professor Stefanie Petermichl for suggesting the problem of this paper, and for the valuable feedback and guidance in the process of writing it. First author is supported by the Alexander von Humboldt Stiftung. Second author is supported by the ERC project CHRiSHarMa no. DLV-682402.

JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES

1. E. Airta, *Two-weight commutator estimates: general multiparameter framework*, Publ. Mat. **64** (2020), no. 2, 681–729.
2. E. Airta, K. Li, H. Martikainen, and E. Vuorinen, *Some new weighted estimates on product spaces*, Preprint (accepted/in press), available at arXiv:1910.12546.
3. Ó. Blasco and S. Pott, *Carleson's counterexample and a scale of Lorentz-BMO spaces on the bitorus*, Ark. Mat. **43** (2005), no. 2, 289–305.
4. Ó. Blasco and S. Pott, *Dyadic BMO on the bidisk*, Rev. Mat. Iberoamericana **21** (2005), no. 2, 483–510.
5. S. Bloom, *A commutator theorem and weighted BMO*, Trans. Amer. Math. Soc. **292** (1985), no. 1, 103–122.
6. L. Carleson, *A Counterexample for Measures Bounded on H^p spaces for the Bidisk*, Mittag-Leffler Report no. 7, 1974.
7. O. N. Capri and C. E. Gutiérrez, *Weighted inequalities for a vector-valued strong maximal function*, Rocky Mountain J. Math. **18** (1988), no. 3, 565–570.
8. S.-Y. A. Chang, *Carleson measure on the Bi-Disc*, Ann. of Math. (2) **109** (1979), no. 3, 613–620.
9. R. R. Coifman, R. Rochberg and G. Weiss, *Factorization theorems for Hardy spaces in several variables*, Ann. of Math. (2) **103** (1976), no. 3, 611–635.
10. D. Cruz-Uribe, J. M. Martell, and C. Pérez, *Extrapolation from A_∞ weights and applications*, J. Funct. Anal. **213** (2004), no. 2, 412–439.
11. L. Dalenc and Y. Ou, *Upper bound for multi-parameter iterated commutators*, Publ. Mat. **60** (2016), no. 1, 191–220.
12. L. Dalenc and S. Petermichl, *A lower bound criterion for iterated commutators*, J. Funct. Anal. **266** (2014), no. 8, 5300–5320.
13. M. Frazier, B. Jawerth and G. Weiss, *Littlewood–Paley theory and the study of function spaces*, CBMS Regional Conference Series in Mathematics, vol. 79, American Mathematical Society, Providence, RI, 1991.
14. R. Fefferman, *Bounded mean oscillation on the polydisk*, Ann. of Math. (2) **110** (1979), no. 2, 395–406.

15. S. H. Ferguson and C. Sadosky, *Characterizations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures*, J. Anal. Math. **81** (2000), 239–267
16. S. H. Ferguson and M. T. Lacey, *A characterisation of product BMO by commutators*, Acta Math. **189** (2002), no. 2, 143–160.
17. L. Grafakos, *Classical Fourier analysis*, 3rd ed., Graduate Texts in Mathematics, Springer, New York, 2014.
18. I. Holmes, M. T. Lacey, and B. D. Wick, *Commutators in the two-weight setting*, Math. Ann. **367** (2017), no. 1-2, 51–80.
19. I. Holmes, S. Petermichl, and B. D. Wick, *Weighted little bmo and two-weight inequalities for Journé commutators*, Anal. PDE **11** (2018), no. 7, 1693–1740.
20. T. Hytönen, *The sharp weighted bound for general Calderón–Zygmund operators*, Ann. of Math. (2) **175** (2012), no. 3, 1473–1506.
21. J. Isralowitz, *Matrix weighted Triebel-Lizorkin bounds: a simple stopping time proof*, Preprint (submitted), available at arXiv:1507.06700.
22. J. Isralowitz, S. Pott, and S. Treil, *Commutators in the two scalar and matrix weighted setting*, Preprint (submitted), available at arXiv:2001.11182.
23. K. Li, H. Martikainen, and E. Vuorinen, *Bloom type upper bounds in the product BMO setting*, J. Geom. Anal. **30** (2020), no. 3, 3181–3203
24. K. Li, H. Martikainen, and E. Vuorinen, *Bloom type inequality for bi-parameter singular integrals: efficient proof and iterated commutators*, International Mathematics Research Notices, rnz072, 2019.
25. H. Martikainen, *Representation of bi-parameter singular integrals by dyadic operators*, Adv. Math. **229** (2012), no. 3, 1734–1761.
26. B. Muckenhoupt and R. L. Wheeden, *Weighted bounded mean oscillation and the Hilbert transform*, Studia Math. **54** (1975/76), no. 3, 221–237.
27. F. Nazarov, S. Petermichl, S. Treil, and A. Volberg, *Convex body domination and weighted estimates with matrix weights*, Adv. Math. **318** (2017), 279–306
28. Z. Nehari, *On bounded bilinear forms*, Ann. of Math. (2) **65** (1957), 153–162.
29. Y. Ou, S. Petermichl, and E. Strouse, *Higher order Journé commutators and characterizations of multi-parameter BMO*, Adv. Math. **291** (2016), 24–58.
30. T. Tao, *Dyadic product H^1, BMO, and Carleson's counterexample*, Unpublished manuscript, available at https://www.math.ucla.edu/~tao/preprints/Expository/product.dvi.
31. S. Treil, *A remark on two weight estimates for positive dyadic operators*, Operator-related function theory and time-frequency analysis, K. Gröchenig, Y. Lyubarskii, and K. Seip (eds.), Abel Symposia, vol. 9, Springer, Cham, 2015.