

BLOOM TYPE UPPER BOUNDS IN THE PRODUCT BMO SETTING

KANGWEI LI, HENRI MARTIKAINEN, AND EMIL VUORINEN

ABSTRACT. We prove some Bloom type estimates in the product BMO setting. More specifically, for a bounded singular integral T_n in \mathbb{R}^n and a bounded singular integral T_m in \mathbb{R}^m we prove that

$$\|[T_n^1, [b, T_m^2]]\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)},$$

where $p \in (1, \infty)$, $\mu, \lambda \in A_p$ and $\nu := \mu^{1/p} \lambda^{-1/p}$ is the Bloom weight. Here T_n^1 is T_n acting on the first variable, T_m^2 is T_m acting on the second variable, A_p stands for the bi-parameter weights of $\mathbb{R}^n \times \mathbb{R}^m$ and $\text{BMO}_{\text{prod}}(\nu)$ is a weighted product BMO space.

1. INTRODUCTION

Let μ and λ be two general Radon measures in \mathbb{R}^n . A two-weight problem asks for a characterisation of the boundedness $T: L^p(\mu) \rightarrow L^p(\lambda)$, where T is, for instance, a singular integral operator (SIO). Singular integral operators take the form

$$Tf(x) = \int_{\mathbb{R}^n} K(x, y) f(y) dy,$$

where different assumptions on the *kernel* K lead to important classes of linear transformations arising across pure and applied analysis. There exists a two-weight characterisation for the Hilbert transform $T = H$, where $K(x, y) = 1/(x - y)$, by Lacey [25] and Lacey, Sawyer, Uriarte-Tuero and Shen [26] (see also Hytönen [20]).

In a Bloom type variant of the two-weight question we require that μ and λ are Muckenhoupt A_p weights and that the problem involves a function b that can be taken to be in some appropriate weighted BMO space $\text{BMO}(\nu)$ formed using the Bloom weight $\nu := \mu^{1/p} \lambda^{-1/p} \in A_2$. The presence of the function b lands us naturally to the commutator setting. Coifman–Rochberg–Weiss [3] showed that the commutator $[b, T]f := bTf - T(bf)$ satisfies

$$\|b\|_{\text{BMO}} \lesssim \|[b, T]\|_{L^p \rightarrow L^p} \lesssim \|b\|_{\text{BMO}}, \quad p \in (1, \infty),$$

for a wide class of singular integrals T . These estimates are called the commutator lower bound and the commutator upper bound, respectively. Both estimates are non-trivial and proved quite differently. The upper bound should hold for all bounded SIOs, while the lower bound requires some suitable non-degeneracy. Given some operator A^b , the definition of which depends naturally on some function b , the Bloom type questions concerns the validity of the estimate

$$\|A^b\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}(\nu)}.$$

2010 *Mathematics Subject Classification.* 42B20.

Key words and phrases. Iterated commutators, Bloom's inequality, product BMO, weighted BMO.

When studied in the natural commutator setting we may also, for suitable singular integrals T , hope to prove the appropriate lower bound, such as

$$\|b\|_{\text{BMO}(\nu)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \| [b, T] \|_{L^p(\mu) \rightarrow L^p(\lambda)}.$$

In the Hilbert transform case $T = H$ Bloom [2] proved such a two-sided estimate.

A renewed interest on the Bloom type estimates started from the recent works of Holmes–Lacey–Wick [14, 15]. They proved Bloom’s upper bound for general bounded SIOs in \mathbb{R}^n using modern proof techniques, and considered the lower bound in the Riesz case. Lerner–Ombrosi–Rivera-Ríos [30] further refined these results using sparse domination. An iterated commutator of the form $[b, [b, T]]$ is studied by Holmes–Wick [17], when $b \in \text{BMO} \cap \text{BMO}(\nu)$. It was also possible to prove this iterated case from the first order case via the so-called Cauchy integral trick of Coifman–Rochberg–Weiss [3], see Hytönen [21]. However, this works precisely because it is also assumed that $b \in \text{BMO}$. It turns out that $\text{BMO} \cap \text{BMO}(\nu)$ is not optimal. A fundamentally improved iterated case is by Lerner–Ombrosi–Rivera-Ríos [31]. There $b \in \text{BMO}(\nu^{1/2}) \supseteq \text{BMO} \cap \text{BMO}(\nu)$, and this is a characterisation as lower bounds also hold by [31]. In the recent paper [22] by Hytönen lower bounds with weak non-degeneracy assumptions on T (or its kernel K) are shown. On the other hand, multilinear Bloom type inequalities were initiated by Kunwar–Ou [24].

We are interested in Bloom type inequalities in the bi-parameter setting. Classical *one-parameter* kernels are “singular” (involve “division by zero”) exactly when $x = y$. In contrast, the *multi-parameter* theory is concerned with kernels whose singularity is spread over the union of all hyperplanes of the form $x_i = y_i$, where $x, y \in \mathbb{R}^d$ are written as $x = (x_i)_{i=1}^t \in \mathbb{R}^{d_1} \times \dots \times \mathbb{R}^{d_t}$ for a fixed partition $d = d_1 + \dots + d_t$. The bi-parameter case $d = d_1 + d_2 = n + m$ is already representative of many of the challenges arising in this context. The prototype example is $1/[(x_1 - y_1)(x_2 - y_2)]$, the product of Hilbert kernels in both coordinate directions of \mathbb{R}^2 , but general two-parameter kernels are neither assumed to be of the product nor of the convolution form. A general bi-parameter SIO is a linear operator with various kernel representations depending whether we have separation in both \mathbb{R}^n and \mathbb{R}^m (a full kernel representation), or just in \mathbb{R}^n or \mathbb{R}^m (a partial kernel representation). See the bi-parameter representation theorem [34] by one of us. This is the modern viewpoint but deals with the same class of SIOs as originally introduced by Journé [23] (see Grau de la Herrán [13]).

In general, there is a vast difference between the techniques that are required in the bi-parameter (or more generally multi-parameter) setting and in the one-parameter theory. For example, there is only one sparse domination paper in this setting, namely Barron–Pipher [1], but the sparse domination is quite a bit more restricted than in the one-parameter setting.

Bloom theory in the bi-parameter setting is very recent. Bloom type upper bounds for commutators of the form $[b, T]$ are valid for bi-parameter Calderón–Zygmund operators T and they are established using the bi-parameter dyadic representation theorem [34] and carefully estimating the resulting commutators of various dyadic model operators (DMOs). Here we define a bi-parameter Calderón–Zygmund operator (CZO) to be a bi-parameter SIO such that $T1, T^*1, T_1(1), T_1^*(1) \in \text{BMO}_{\text{prod}}$, and certain additional weak testing conditions hold. Here T_1 is a *partial adjoint* of T in the first slot, $\langle T_1(f_1 \otimes f_2), g_1 \otimes g_2 \rangle = \langle T(g_1 \otimes f_2), f_1 \otimes g_2 \rangle$, and BMO_{prod} is the product BMO of Chang and Fefferman

[6, 7]. Holmes–Petermichl–Wick [16] proved the first bi-parameter Bloom type estimate

$$\|[b, T]\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{bmo}(\nu)}.$$

Here A_p stands for bi-parameter weights (replace cubes by rectangles in the usual definition) and $\text{bmo}(\nu)$ is the weighted little BMO space defined using the norm

$$\|b\|_{\text{bmo}(\nu)} := \sup_R \frac{1}{\nu(R)} \int_R |b - \langle b \rangle_R|, \text{ where } \langle b \rangle_R = \frac{1}{|R|} \int_R b, \nu(R) = \int_R \nu,$$

and the supremum is over all rectangles $R = I \times J \subset \mathbb{R}^n \times \mathbb{R}^m$. We refer to these types of commutators as little BMO commutators.

In the recent paper [33] we reproved the result of [16] in an efficient way based on improved commutator decompositions from our bilinear bi-parameter theory [32]. The clear structure of our proof allowed us to also handle the iterated little BMO commutator case and to prove the upper bound

$$(1.1) \quad \|[b_k, \dots [b_2, [b_1, T]] \dots]\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \prod_{i=1}^k \|b_i\|_{\text{bmo}(\nu^{\theta_i})}, \quad \sum_{i=1}^k \theta_i = 1.$$

There is also a different type of commutator, which is equally fundamental in the bi-parameter setting, but which we have not yet discussed. *In this paper we are interested in Bloom type question for this commutator* – the main point is that now the right space is not the little bmo but the much harder product BMO. If T_n and T_m are one-parameter CZOs (bounded SIOs) in \mathbb{R}^n and \mathbb{R}^m , respectively, then for the $L^p \rightarrow L^p$ norm of the commutator

$$[T_n^1, [b, T_m^2]]f = T_n^1(bT_m^2f) - T_n^1T_m^2(bf) - bT_m^2T_n^1f + T_m^2(bT_n^1f),$$

where $T_n^1f(x) = T_n(f(\cdot, x_2))(x_1)$, the right object is $b \in \text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})$. We refer to these commutators as product BMO type commutators. Classical references for commutators of both type in the Hilbert transform case include Ferguson–Sadosky [12] and the groundbreaking paper Ferguson–Lacey [11]. In particular, [11] contains the extremely difficult lower bound estimate

$$\|b\|_{\text{BMO}_{\text{prod}}(\mathbb{R}^2)} \lesssim \|[H^1, [b, H^2]]\|_{L^2 \rightarrow L^2}.$$

See also Lacey–Petermichl–Pipher–Wick [27, 28, 29] for the higher dimensional Riesz setting and applications to div-curl lemmas. The corresponding upper bound is not easy either – even for special operators such as the Riesz transforms. However, in [9] Dalenc and Ou proved that for all CZOs

$$\|[T_n^1, [b, T_m^2]]\|_{L^p \rightarrow L^p} \lesssim \|b\|_{\text{BMO}_{\text{prod}}}$$

using the modern approach via the one-parameter representation theorem of Hytönen [18].

However, no Bloom type estimates have been considered. In this paper we prove the Dalenc–Ou [9] type bound in the Bloom setting – i.e., we prove Bloom type upper bounds for product BMO commutators. The following is our main result.

1.2. Theorem. *Let T_n and T_m be one-parameter CZOs in \mathbb{R}^n and \mathbb{R}^m , respectively, and let $p \in (1, \infty)$, $\mu, \lambda \in A_p$ and $\nu := \mu^{1/p} \lambda^{-1/p}$. We have the quantitative estimate*

$$\|[T_n^1, [b, T_m^2]]\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

Here $\text{BMO}_{\text{prod}}(\nu)$ is a weighted product BMO space as defined at least in [16, 33]. We note that $\text{bmo}(\nu) \subset \text{BMO}_{\text{prod}}(\nu)$ (a proof can be found at least in [33]).

This product BMO Bloom estimate is proved in Section 4. While we were inspired by our previous paper [33] dealing with the new approach to little BMO commutators, a quite different take on things is required by the current product BMO setting. The proof idea is outlined in the beginning of Section 4.

It is probably possible to add more singular integrals to this Bloom bound and/or consider multi-parameter CZOs here, but we content with the most fundamental case here. However, we still mention the following. If we consider multi-parameter singular integrals in the product BMO type commutators, then some kind of little product BMO assumptions concerning b are the right thing. Suppose e.g. that T_1 and T_2 are bi-parameter CZOs in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ and $\mathbb{R}^{n_3} \times \mathbb{R}^{n_4}$, respectively. Then according to Ou–Petermichl–Strouse [35] and Holmes–Petermichl–Wick [16] we have

$$\begin{aligned} \|[T_1, [b, T_2]]f\|_{L^2(\prod_{i=1}^4 \mathbb{R}^{n_i})} &\lesssim \max \left(\sup_{x_2, x_4} \|b(\cdot, x_2, \cdot, x_4)\|_{\text{BMO}_{\text{prod}}}, \sup_{x_2, x_3} \|b(\cdot, x_2, x_3, \cdot)\|_{\text{BMO}_{\text{prod}}}, \right. \\ &\quad \left. \sup_{x_1, x_4} \|b(x_1, \cdot, \cdot, x_4)\|_{\text{BMO}_{\text{prod}}}, \sup_{x_1, x_3} \|b(x_1, \cdot, x_3, \cdot)\|_{\text{BMO}_{\text{prod}}} \right) \|f\|_{L^2(\prod_{i=1}^4 \mathbb{R}^{n_i})}, \end{aligned}$$

where instead of T_1 we could also similarly as above write $T_1^{1,2}$ highlighting the fact that here it acts on the first two variables. In general, the field of multi-parameter commutator estimates is again very active – we mention e.g. that recently the commutators of multi-parameter flag singular integrals were investigated by Duong–Li–Ou–Pipher–Wick [10].

The product BMO Bloom type upper bound is the main contribution of this paper. However, we also complement our previous paper [33] and (1.1) by giving an easy little BMO iterated commutator lower bound proof using the so-called median method, previously used in the one-parameter setting in [31, 22] (see also [32]). In [33] we only recorded the following remark regarding the Estimate (1.1). Choosing $b_1 = \dots = b_k = b$ and $\theta_1 = \dots = \theta_k = 1/k$ in (1.1) we get a bi-parameter analog of [31], while choosing $\theta_1 = 1$ (and the rest zero) we get analogs of [17, 21]. However, the first is the better choice as $\text{bmo}(\nu^{1/k}) \supset \text{bmo} \cap \text{bmo}(\nu)$. Indeed, similarly as in the one-parameter case [31], this is seen by using that $\langle \nu \rangle_R^\theta \lesssim_{[\nu]_{A_2}} \langle \nu^\theta \rangle_R$ for all $\theta \in (0, 1)$ and rectangles R (this estimate follows from [4, Theorem 2.1] by iteration). In this paper we actually prove the lower bound showing the optimality of $\text{bmo}(\nu^{1/k})$. The details are given quickly in Section 5.

Acknowledgements. K. Li was supported by Juan de la Cierva - Formación 2015 FJCI-2015-24547, by the Basque Government through the BERC 2018-2021 program and by Spanish Ministry of Economy and Competitiveness MINECO through BCAM Severo Ochoa excellence accreditation SEV-2017-0718 and through project MTM2017-82160-C2-1-P funded by (AEI/FEDER, UE) and acronym “HAQMEC”.

H. Martikainen was supported by the Academy of Finland through the grants 294840 and 306901, and by the three-year research grant 75160010 of the University of Helsinki. He is a member of the Finnish Centre of Excellence in Analysis and Dynamics Research.

E. Vuorinen was supported by the Academy of Finland through the grant 306901, by the Finnish Centre of Excellence in Analysis and Dynamics Research, and by Jenny and Antti Wihuri Foundation.

2. DEFINITIONS AND PRELIMINARIES

2.1. Basic notation. We denote $A \lesssim B$ if $A \leq CB$ for some constant C that can depend on the dimension of the underlying spaces, on integration exponents, and on various other constants appearing in the assumptions. We denote $A \sim B$ if $B \lesssim A \lesssim B$.

We work in the bi-parameter setting in the product space $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$. In such a context $x = (x_1, x_2)$ with $x_1 \in \mathbb{R}^n$ and $x_2 \in \mathbb{R}^m$. We often take integral pairings with respect to one of the two variables only: If $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$ and $h: \mathbb{R}^n \rightarrow \mathbb{C}$, then $\langle f, h \rangle_1: \mathbb{R}^m \rightarrow \mathbb{C}$ is defined by $\langle f, h \rangle_1(x_2) = \int_{\mathbb{R}^n} f(y_1, x_2)h(y_1) dy_1$.

2.2. Dyadic notation, Haar functions and martingale differences. We denote a dyadic grid in \mathbb{R}^n by \mathcal{D}^n and a dyadic grid in \mathbb{R}^m by \mathcal{D}^m . If $I \in \mathcal{D}^n$, then $I^{(k)}$ denotes the unique dyadic cube $S \in \mathcal{D}^n$ so that $I \subset S$ and $\ell(S) = 2^k \ell(I)$. Here $\ell(I)$ stands for side length. Also, $\text{ch}(I)$ denotes the dyadic children of I , i.e., $I' \in \text{ch}(I)$ if $I' \in \mathcal{D}^n$, $I' \subset I$ and $\ell(I') = \ell(I)/2$. We often write $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$.

When $I \in \mathcal{D}^n$ we denote by h_I a cancellative L^2 normalised Haar function. This means the following. Writing $I = I_1 \times \dots \times I_n$ we can define the Haar function h_I^η , $\eta = (\eta_1, \dots, \eta_n) \in \{0, 1\}^n$, by setting

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

where $h_{I_i}^0 = |I_i|^{-1/2} 1_{I_i}$ and $h_{I_i}^1 = |I_i|^{-1/2} (1_{I_{i,l}} - 1_{I_{i,r}})$ for every $i = 1, \dots, n$. Here $I_{i,l}$ and $I_{i,r}$ are the left and right halves of the interval I_i respectively. The reader should carefully notice that h_I^0 is the non-cancellative Haar function for us and that in some other papers a different convention is used. If $\eta \in \{0, 1\}^n \setminus \{0\}$ the Haar function is cancellative: $\int h_I^\eta = 0$. We usually suppress the presence of η and simply write h_I for some h_I^η , $\eta \in \{0, 1\}^n \setminus \{0\}$. Then $h_I h_I$ can stand for $h_I^{\eta_1} h_I^{\eta_2}$, but we always treat such a product as a non-cancellative function i.e. use only its size.

For $I \in \mathcal{D}^n$ and a locally integrable function $f: \mathbb{R}^n \rightarrow \mathbb{C}$, we define the martingale difference

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

Here $\langle f \rangle_I = \frac{1}{|I|} \int_I f$. We also write $E_I f = \langle f \rangle_I 1_I$. Now, we have $\Delta_I f = \sum_{\eta \neq 0} \langle f, h_I^\eta \rangle h_I^\eta$, or suppressing the η summation, $\Delta_I f = \langle f, h_I \rangle h_I$, where $\langle f, h_I \rangle = \int f h_I$. A martingale block is defined by

$$\Delta_{K,i} f = \sum_{\substack{I \in \mathcal{D}^n \\ I^{(i)} = K}} \Delta_I f, \quad K \in \mathcal{D}^n, i \in \mathbb{N}.$$

Next, we define bi-parameter martingale differences. Let $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{C}$ be locally integrable. Let $I \in \mathcal{D}^n$ and $J \in \mathcal{D}^m$. We define the martingale difference

$$\Delta_I^1 f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}, \Delta_I^1 f(x) := \Delta_I(f(\cdot, x_2))(x_1).$$

Define $\Delta_J^2 f$ analogously, and also define E_I^1 and E_J^2 similarly. We set

$$\Delta_{I \times J} f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}, \Delta_{I \times J} f(x) = \Delta_I^1(\Delta_J^2 f)(x) = \Delta_J^2(\Delta_I^1 f)(x).$$

Notice that $\Delta_I^1 f = h_I \otimes \langle f, h_I \rangle_1$, $\Delta_J^2 f = \langle f, h_J \rangle_2 \otimes h_J$ and $\Delta_{I \times J} f = \langle f, h_I \otimes h_J \rangle h_I \otimes h_J$ (suppressing the finite η summations). Martingale blocks are defined in the natural way

$$\Delta_{K \times V}^{i,j} f = \sum_{I: I^{(i)}=K} \sum_{J: J^{(j)}=V} \Delta_{I \times J} f = \Delta_{K,i}^1 (\Delta_{V,j}^2 f) = \Delta_{V,j}^2 (\Delta_{K,i}^1 f).$$

2.3. Weights. A weight $w(x_1, x_2)$ (i.e. a locally integrable a.e. positive function) belongs to bi-parameter $A_p(\mathbb{R}^n \times \mathbb{R}^m)$, $1 < p < \infty$, if

$$[w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)} := \sup_R \langle w \rangle_R \langle w^{1-p'} \rangle_R^{p-1} < \infty,$$

where the supremum is taken over $R = I \times J$, where $I \subset \mathbb{R}^n$ and $J \subset \mathbb{R}^m$ are cubes with sides parallel to the axes (we simply call such R rectangles). We have

$$[w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)} < \infty \text{ iff } \max \left(\operatorname{ess\,sup}_{x_1 \in \mathbb{R}^n} [w(x_1, \cdot)]_{A_p(\mathbb{R}^m)}, \operatorname{ess\,sup}_{x_2 \in \mathbb{R}^m} [w(\cdot, x_2)]_{A_p(\mathbb{R}^n)} \right) < \infty,$$

and that $\max \left(\operatorname{ess\,sup}_{x_1 \in \mathbb{R}^n} [w(x_1, \cdot)]_{A_p(\mathbb{R}^m)}, \operatorname{ess\,sup}_{x_2 \in \mathbb{R}^m} [w(\cdot, x_2)]_{A_p(\mathbb{R}^n)} \right) \leq [w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}$, while the constant $[w]_{A_p}$ is dominated by the maximum to some power. Of course, $A_p(\mathbb{R}^n)$ is defined similarly as $A_p(\mathbb{R}^n \times \mathbb{R}^m)$ – just take the supremum over cubes Q . For the basic theory of bi-parameter weights consult e.g. [16].

2.4. Square functions and maximal functions. Given $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$ and $g: \mathbb{R}^n \rightarrow \mathbb{C}$ we denote the dyadic maximal functions by

$$M_{\mathcal{D}^n} g := \sup_{I \in \mathcal{D}^n} \frac{1_I}{|I|} \int_I |g| \text{ and } M_{\mathcal{D}} f := \sup_{R \in \mathcal{D}} \frac{1_R}{|R|} \int_R |f|.$$

We also set $M_{\mathcal{D}^n}^1 f(x_1, x_2) = M_{\mathcal{D}^n}(f(\cdot, x_2))(x_1)$. The operator $M_{\mathcal{D}^m}^2$ is defined similarly.

Define the square functions

$$S_{\mathcal{D}} f = \left(\sum_{R \in \mathcal{D}} |\Delta_R f|^2 \right)^{1/2}, \quad S_{\mathcal{D}^n}^1 f = \left(\sum_{I \in \mathcal{D}^n} |\Delta_I^1 f|^2 \right)^{1/2} \text{ and } S_{\mathcal{D}^m}^2 f = \left(\sum_{J \in \mathcal{D}^m} |\Delta_J^2 f|^2 \right)^{1/2}.$$

Define also

$$S_{\mathcal{D}, M}^1 f = \left(\sum_{I \in \mathcal{D}^n} \frac{1_I}{|I|} \otimes [M_{\mathcal{D}^m} \langle f, h_I \rangle_1]^2 \right)^{1/2} \text{ and } S_{\mathcal{D}, M}^2 f = \left(\sum_{J \in \mathcal{D}^m} [M_{\mathcal{D}^n} \langle f, h_J \rangle_2]^2 \otimes \frac{1_J}{|J|} \right)^{1/2}.$$

We record the following standard estimates, which are used repeatedly below in an implicit manner. Some similar estimates that follow from the ones below are also used.

2.1. Lemma. For $p \in (1, \infty)$ and $w \in A_p = A_p(\mathbb{R}^n \times \mathbb{R}^m)$ we have the weighted square function estimates

$$\|f\|_{L^p(w)} \sim_{[w]_{A_p}} \|S_{\mathcal{D}} f\|_{L^p(w)} \sim_{[w]_{A_p}} \|S_{\mathcal{D}^n}^1 f\|_{L^p(w)} \sim_{[w]_{A_p}} \|S_{\mathcal{D}^m}^2 f\|_{L^p(w)}.$$

Moreover, for $p, s \in (1, \infty)$ we have the Fefferman–Stein inequality

$$\left\| \left(\sum_j |M f_j|^s \right)^{1/s} \right\|_{L^p(w)} \lesssim_{[w]_{A_p}} \left\| \left(\sum_j |f_j|^s \right)^{1/s} \right\|_{L^p(w)}.$$

Here M can e.g. be $M_{\mathcal{D}^n}^1$ or $M_{\mathcal{D}}$. Finally, we have

$$\|S_{\mathcal{D}, M}^1 f\|_{L^p(w)} + \|S_{\mathcal{D}, M}^2 f\|_{L^p(w)} \lesssim_{[w]_{A_p}} \|f\|_{L^p(w)}.$$

One easy way to show such estimates is to reduce to $p = 2$ via standard extrapolation. When $p = 2$ it is especially easy to use one-parameter results iteratively. See e.g. [5, 8] for one-parameter square function results and their history.

2.5. BMO spaces. Let $b: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$ be locally integrable.

We define the weighted product BMO space. Given $\nu \in A_2(\mathbb{R}^n \times \mathbb{R}^m)$ set

$$\|b\|_{\text{BMO}_{\text{prod}}(\nu, \mathcal{D})} := \sup_{\Omega} \left(\frac{1}{\nu(\Omega)} \sum_{\substack{R \in \mathcal{D} \\ R \subset \Omega}} |\langle b, h_R \rangle|^2 \langle \nu \rangle_R^{-1} \right)^{1/2},$$

where $h_R := h_I \otimes h_J$ and the supremum is taken over those sets $\Omega \subset \mathbb{R}^{n+m}$ such that $|\Omega| < \infty$ and such that for every $x \in \Omega$ there exist $R \in \mathcal{D}$ so that $x \in R \subset \Omega$. The non-dyadic product BMO space $\text{BMO}_{\text{prod}}(\nu)$ is defined using the norm defined by the supremum over all dyadic grids of the above dyadic norms.

We also say that b belongs to the weighted dyadic little BMO space $\text{bmo}(\nu, \mathcal{D})$ if

$$\|b\|_{\text{bmo}(\nu, \mathcal{D})} := \sup_{R \in \mathcal{D}} \frac{1}{\nu(R)} \int_R |b - \langle b \rangle_R| < \infty.$$

The non-dyadic variant $\text{bmo}(\nu)$ has the obvious definition. There holds $\text{bmo}(\nu, \mathcal{D}) \subset \text{BMO}_{\text{prod}}(\nu, \mathcal{D})$ (this is proved explicitly at least in [33]). We do not need this fact in this paper, however.

For a sequence $(a_I)_{I \in \mathcal{D}^n}$ we define

$$\|(a_I)\|_{\text{BMO}(\mathcal{D}^n)} = \sup_{I_0 \in \mathcal{D}^n} \left(\frac{1}{|I_0|} \sum_{I \subset I_0} |a_I|^2 \right)^{1/2}.$$

3. MARTINGALE DIFFERENCE EXPANSIONS OF PRODUCTS

Let \mathcal{D}^n and \mathcal{D}^m be some fixed dyadic grids in \mathbb{R}^n and \mathbb{R}^m , respectively, and write $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$. In what follows we sum over $I \in \mathcal{D}^n$ and $J \in \mathcal{D}^m$. In general, in this paper always $K, I, I_1, I_2 \in \mathcal{D}^n$ and $V, J, J_1, J_2 \in \mathcal{D}^m$.

Paraproduct operators. The product BMO type paraproducts are

$$\begin{aligned} A_1(b, f) &= \sum_{I, J} \Delta_{I \times J} b \Delta_{I \times J} f, & A_2(b, f) &= \sum_{I, J} \Delta_{I \times J} b E_I^1 \Delta_J^2 f, \\ A_3(b, f) &= \sum_{I, J} \Delta_{I \times J} b \Delta_I^1 E_J^2 f, & A_4(b, f) &= \sum_{I, J} \Delta_{I \times J} b \langle f \rangle_{I \times J}. \end{aligned}$$

The little BMO type paraproducts are

$$\begin{aligned} A_5(b, f) &= \sum_{I, J} E_I^1 \Delta_J^2 b \Delta_{I \times J} f, & A_6(b, f) &= \sum_{I, J} E_I^1 \Delta_J^2 b \Delta_I^1 E_J^2 f, \\ A_7(b, f) &= \sum_{I, J} \Delta_I^1 E_J^2 b \Delta_{I \times J} f, & A_8(b, f) &= \sum_{I, J} \Delta_I^1 E_J^2 b E_I^1 \Delta_J^2 f. \end{aligned}$$

The ‘‘illegal’’ bi-parameter paraproduct is

$$W(b, f) = \sum_{I, J} \langle b \rangle_{I \times J} \Delta_{I \times J} f.$$

Things make sense at least with the a priori assumptions that b is bounded and f is bounded and compactly supported. For us $b \in \text{BMO}_{\text{prod}}(\nu)$ so that only the A_1, \dots, A_4 are good as standalone operators. If we would have $b \in \text{bmo}(\nu) \subset \text{BMO}_{\text{prod}}(\nu)$, then also A_5, \dots, A_8 would be good (but this is not the case here). So we can only rely on the following boundedness property. See [16, 33].

3.1. Lemma. *Suppose $b \in \text{BMO}_{\text{prod}}(\nu)$, where $\nu = \mu^{1/p} \lambda^{-1/p}$, $\mu, \lambda \in A_p$ and $p \in (1, \infty)$. Then for $i = 1, \dots, 4$ we have*

$$\|A_i(b, \cdot)\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

We formally have

$$bf = \sum_{i=1}^8 A_i(b, f) + W(b, f).$$

When we write like this we say that we decompose bf in the bi-parameter sense. When we employ this decomposition in practice we have to form suitable differences of the problematic operators with some other terms to get something that is bounded with the product BMO assumption. The operator W has a special role.

Next, we define operators related to one-parameter type decompositions. Define the one-parameter paraproducts

$$a_1^1(b, f) = \sum_I \Delta_I^1 b \Delta_I^1 f, \quad a_2^1(b, f) = \sum_I \Delta_I^1 b E_I^1 f.$$

Define also the “illegal” one-parameter paraproduct

$$w^1(b, f) = \sum_I E_I^1 b \Delta_I^1 f.$$

The operators $a_i^1(b, \cdot)$ would be bounded if $b \in \text{bmo}(\nu) \subset \text{BMO}_{\text{prod}}(\nu)$, but again this is not the case here. The operators $a_1^2(b, f)$, $a_2^2(b, f)$ and $w^2(b, f)$ are defined analogously. We formally have

$$bf = \sum_{i=1}^2 a_i^1(b, f) + w^1(b, f) = \sum_{i=1}^2 a_i^2(b, f) + w^2(b, f).$$

In this case we say that we decomposed bf in the one-parameter sense (either in \mathbb{R}^n or \mathbb{R}^m).

4. THE BLOOM TYPE PRODUCT BMO UPPER BOUND

4.1. Theorem. *Let T_n and T_m be one-parameter CZOs in \mathbb{R}^n and \mathbb{R}^m , respectively, and let $p \in (1, \infty)$, $\mu, \lambda \in A_p$ and $\nu := \mu^{1/p} \lambda^{-1/p}$. We have the quantitative estimate*

$$\| [T_n^1, [b, T_m^2]] \|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

4.2. Remark. Here we are proving only the right quantitative bound, and prefer to understand the inequality so that b is nice to begin with – we at least make the purely qualitative a priori assumption $b \in L^\infty(\mathbb{R}^{n+m})$.

Proof. It is enough to prove the Bloom type inequality for $[U_n^1, [b, U_m^2]]$, where U_n and U_m are DMOs (dyadic model operators) appearing in the representation theorem [18, 19]. This means that $U_n \in \{S_n, \pi_n\}$, where S_n is a dyadic shift in \mathbb{R}^n and π_n is a dyadic paraproduct in \mathbb{R}^n (we will recall what these mean as we go). We only have to maintain a polynomial dependence on the complexity of the shift.

Let $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$ be bounded and compactly supported. How we expand various products of functions (in the bi-parameter sense, in the one-parameter sense, or not at all) depends on the structure of the model operators: a cancellative Haar function in a correct position is required to trigger an expansion in the corresponding parameter. In general, the model operators have the form

$$U_n^1 f = \sum_{I_i^{(k_i)}=K}^K a_{K,(I_i)} \tilde{h}_{I_2} \otimes \langle f, \tilde{h}_{I_1} \rangle_1, \quad U_m^2 f = \sum_{J_j^{(v_j)}=V}^V a_{V,(J_j)} \langle f, \tilde{h}_{J_1} \rangle_2 \otimes \tilde{h}_{J_2},$$

where $k_1, k_2, v_1, v_2 \geq 0$, $K, I_1, I_2 \in \mathcal{D}^n$, $V, J_1, J_2 \in \mathcal{D}^m$, $a_{K,(I_i)}, a_{V,(J_j)}$ are appropriate constants and $\tilde{h}_{I_i} \in \{h_{I_i}, 1_{I_i}/|I_i|\}$, $\tilde{h}_{J_j} \in \{h_{J_j}, 1_{J_j}/|J_j|\}$. We will write

$$[U_n^1, [b, U_m^2]]f = I - II - III + IV,$$

where

$$I = U_n^1(bU_m^2 f), \quad II = U_n^1 U_m^2(bf), \quad III = bU_m^2 U_n^1 f \quad \text{and} \quad IV = U_m^2(bU_n^1 f).$$

In all of the terms I, \dots, IV the appearing product $b \cdot U_m^2 f$, $b \cdot f$, $b \cdot U_m^2 U_n^1 f$ or $b \cdot U_n^1 f$, respectively, is expanded. We now explain which of the appearing Haar functions determine the expansion strategy in each of the terms. For I the determining Haar functions are $\tilde{h}_{I_1}, \tilde{h}_{J_2}$, for II they are $\tilde{h}_{I_1}, \tilde{h}_{J_1}$, for III they are $\tilde{h}_{I_2}, \tilde{h}_{J_2}$ and for IV they are $\tilde{h}_{I_2}, \tilde{h}_{J_1}$. If both of the determining Haar functions are cancellative, we expand in the bi-parameter sense using the paraproducts $A_i(b, \cdot)$, $W(b, \cdot)$. If one of them is cancellative we expand in the corresponding parameter $k \in \{1, 2\}$ using the paraproducts $a_i^k(b, \cdot)$, $w^k(b, \cdot)$. In addition, when a non-cancellative Haar function appears we sometimes add and subtract an average to ease the upcoming expansions.

The case $[S_n^1, [b, S_m^2]]$. We study the case that

$$S_n^1 f = \sum_{I_i^{(k_i)}=K}^K a_{K,(I_i)} h_{I_2} \otimes \langle f, h_{I_1} \rangle_1, \quad S_m^2 f = \sum_{J_j^{(v_j)}=V}^V a_{V,(J_j)} \langle f, h_{J_1} \rangle_2 \otimes h_{J_2}.$$

Here $k_1, k_2, v_1, v_2 \geq 0$, $K, I_1, I_2 \in \mathcal{D}^n$ and $V, J_1, J_2 \in \mathcal{D}^m$. Only finitely many of the constants $a_{K,(I_i)}$ are non-zero and

$$|a_{K,(I_i)}| \leq \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|},$$

and similarly for $a_{V,(J_j)}$.

We denote

$$(S_n^1 S_m^2)^{b,1,1} f := \sum_{I_i^{(k_i)}=K}^K \sum_{J_j^{(v_j)}=V}^V \langle b \rangle_{I_1 \times J_1} a_{K,(I_i)} a_{V,(J_j)} \langle f, h_{I_1} \otimes h_{J_1} \rangle h_{I_2} \otimes h_{J_2}.$$

So here 1,1 refers to cubes over which we average b over, namely $\langle b \rangle_{I_1 \times J_1}$. Define $(S_n^1 S_m^2)^{b,1,2}$ etc. analogously. We expand the appearing product in the bi-parameter sense in all of the terms

$$I = S_n^1(b S_m^2 f), II = S_n^1 S_m^2(bf), III = b S_m^2 S_n^1 f \text{ and } IV = S_m^2(b S_n^1 f).$$

The term related to the paraproduct $W(b, \cdot)$ e.g. yields $(S_n^1 S_m^2)^{b,1,2}$ in I , and so we get

$$\begin{aligned} [S_n^1, [b, S_m^2]]f &= I - II - III + IV = E \\ &+ \sum_{i=1}^8 [S_n^1(A_i(b, S_m^2 f)) - S_n^1 S_m^2(A_i(b, f)) - A_i(b, S_m^2 S_n^1 f) + S_m^2(A_i(b, S_n^1 f))], \end{aligned}$$

where

$$E := (S_n^1 S_m^2)^{b,1,2} - (S_n^1 S_m^2)^{b,1,1} - (S_n^1 S_m^2)^{b,2,2} + (S_n^1 S_m^2)^{b,2,1}.$$

We start looking at the sum over i . For $i \leq 4$ even all the individual terms are bounded. This is because of Lemma 3.1 and basic weighted bounds of DMOs:

$$\|S_n^1 f\|_{L^p(w)} \lesssim_{[w]_{A_p}} \|f\|_{L^p(w)}.$$

For $i \geq 5$ the term $S_n^1(A_i(b, S_m^2 f))$ should be paired with one of the terms with a minus. For $i = 5, 6$ we pair with $-A_i(b, S_m^2 S_n^1 f)$ and for $i = 7, 8$ we pair with $-S_n^1 S_m^2(A_i(b, f))$. Let us take $i = 5$. Notice that it is enough to study $S_n^1(A_5(b, f)) - A_5(b, S_n^1 f)$, since $S_n^1 S_m^2 = S_m^2 S_n^1$ and S_m^2 is bounded.

Now, the term $S_n^1(A_5(b, f)) - A_5(b, S_n^1 f)$ equals

$$\sum_{I_i^{(k_i)}=K}^K \sum_J a_{K,(I_i)} [\langle \langle b, h_J \rangle_2 \rangle_{I_1} - \langle \langle b, h_J \rangle_2 \rangle_{I_2}] \langle f, h_{I_1} \otimes h_J \rangle h_{I_2} \otimes h_J h_J.$$

Dualising and using that for a function p in \mathbb{R}^n we have $\langle p \rangle_{I_1} - \langle p \rangle_K = \sum_{k=1}^{k_1} \langle \Delta_{I_1^{(k)}} p \rangle_{I_1}$, we reduce to estimating, for a fixed $k = 1, \dots, k_1$, as follows:

$$\begin{aligned} &\sum_{I_i^{(k_i)}=K}^K \sum_J |a_{K,(I_i)}| |I_1^{(k)}|^{-1/2} |\langle b, h_{I_1^{(k)}} \otimes h_J \rangle| |\langle f, h_{I_1} \otimes h_J \rangle| |\langle \langle g, h_{I_2} \rangle_1 \rangle_J \\ &= \sum_K \sum_{I^{(k_1-k)}=K} \sum_{\substack{I_1^{(k)}=I \\ I_2^{(k_2)}=K}} |a_{K,(I_i)}| |I|^{-1/2} |\langle b, h_I \otimes h_J \rangle| |\langle f, h_{I_1} \otimes h_J \rangle| |\langle \langle g, h_{I_2} \rangle_1 \rangle_J \\ &\leq \sum_K \sum_{I^{(k_1-k)}=K} |\langle b, h_I \otimes h_J \rangle| |I|^{1/2} |J|^{1/2} |\langle \Delta_{K \times J}^{k_1,0} f \rangle_{I \times J}| \langle |g| \rangle_{K \times J} \\ &\lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \iint_{\mathbb{R}^{n+m}} \left(\sum_{K,J} [M_{\mathcal{D}} \Delta_{K \times J}^{k_1,0} f]^2 \right)^{1/2} M_{\mathcal{D}} g \cdot \nu \\ &\lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}. \end{aligned}$$

Besides the various standard weighted estimates, we also used the weighted H^1 -BMO_{prod} duality estimate:

$$\sum_{I,J} |\langle b, h_I \otimes h_J \rangle| |c_{I,J}| \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu, \mathcal{D})} \iint_{\mathbb{R}^{n+m}} \left(\sum_{I,J} |c_{I,J}|^2 \frac{1_I \otimes 1_J}{|I||J|} \right)^{1/2} \nu.$$

For this see Proposition 4.1 in [16] (this is what is also used to prove Lemma 3.1). Notice also that there was even too much cancellation in this term (we were able to throw away the cancellative Haar function h_{I_2}). We have shown

$$\|S_n^1(A_5(b, f)) - A_5(b, S_n^1 f)\|_{L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} k_1 \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \|f\|_{L^p(\mu)}.$$

The other terms are similar, and so we are done with the i summation.

It remains to study $E := (S_n^1 S_m^2)^{b,1,2} - (S_n^1 S_m^2)^{b,1,1} - (S_n^1 S_m^2)^{b,2,2} + (S_n^1 S_m^2)^{b,2,1}$. After noting that

$$\begin{aligned} \langle b \rangle_{I_1 \times J_2} - \langle b \rangle_{I_1 \times J_1} - \langle b \rangle_{I_2 \times J_2} + \langle b \rangle_{I_2 \times J_1} &= \sum_{k=1}^{k_1} \sum_{v=1}^{v_2} \langle \Delta_{I_1^{(k)} \times J_2^{(v)}} b \rangle_{I_1 \times J_2} \\ &\quad - \sum_{k=1}^{k_1} \sum_{v=1}^{v_1} \langle \Delta_{I_1^{(k)} \times J_1^{(v)}} b \rangle_{I_1 \times J_1} - \sum_{k=1}^{k_2} \sum_{v=1}^{v_2} \langle \Delta_{I_2^{(k)} \times J_2^{(v)}} b \rangle_{I_2 \times J_2} + \sum_{k=1}^{k_2} \sum_{v=1}^{v_1} \langle \Delta_{I_2^{(k)} \times J_1^{(v)}} b \rangle_{I_2 \times J_1} \end{aligned}$$

we can reduce the term E to terms, which are very similar to what has been estimated above. Finally, we get

$$\| [S_n^1, [b, S_m^2]] \|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} (1 + \max k_i)(1 + \max v_i) \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

The case $[\pi_n^1, [b, \pi_m^2]]$. Here we study the case that

$$\pi_n^1 f = \sum_K a_K h_K \otimes \langle f \rangle_{K,1}, \quad \pi_m^2 f = \sum_V a_V \langle f \rangle_{V,2} \otimes h_V.$$

Here $K \in \mathcal{D}^n$ and $V \in \mathcal{D}^m$, $\|(a_K)\|_{\text{BMO}(\mathcal{D}^n)} \leq 1$ and $a_K \neq 0$ for only finitely many K , and similarly for a_V . Paraproducts also have the dual form (unlike shifts which are completely symmetric) – this is taken into account and we comment on it later on.

For $I = \pi_n^1(b \pi_m^2 f)$ we perform a one-parameter expansion in \mathbb{R}^m so that

$$I = \sum_{i=1}^2 \pi_n^1(a_i^2(b, \pi_m^2 f)) + \pi_n^1(w^2(b, \pi_m^2 f)),$$

where by adding and subtracting an average we get

$$\pi_n^1(w^2(b, \pi_m^2 f)) = \sum_{K,V} a_K a_V \langle [\langle b \rangle_{V,2} - \langle b \rangle_{K \times V}] \langle f \rangle_{V,2} \rangle_K h_K \otimes h_V + (\pi_n^1 \pi_m^2)^b f.$$

Here we have denoted $(\pi_n^1 \pi_m^2)^b f = \sum_{K,V} \langle b \rangle_{K \times V} a_K a_V \langle f \rangle_{K \times V} h_K \otimes h_V$. For the term $II = \pi_n^1 \pi_m^2(b f)$ we only add and subtract an average to the end that

$$II = \sum_{K,V} a_K a_V \langle [b - \langle b \rangle_{K \times V}] f \rangle_{K \times V} h_K \otimes h_V + (\pi_n^1 \pi_m^2)^b f.$$

The term $III = b\pi_m^2\pi_n^1f$ warrants a bi-parameter decomposition so that we get

$$III = \sum_{i=1}^8 A_i(b, \pi_m^2\pi_n^1f) + (\pi_n^1\pi_m^2)^b f.$$

Finally, the term $IV = \pi_m^2(b\pi_n^1f)$ is treated symmetrically to I so that

$$IV = \sum_{i=1}^2 \pi_m^2(a_i^1(b, \pi_n^1f)) + \sum_{K,V} a_K a_V \langle [\langle b \rangle_{K,1} - \langle b \rangle_{K \times V}] \langle f \rangle_{K,1} \rangle_V h_K \otimes h_V + (\pi_n^1\pi_m^2)^b f.$$

Now, we get

(4.3)

$$\begin{aligned} & [\pi_n^1, [b, \pi_m^2]]f = I - II - III + IV \\ &= - \sum_{i=1}^4 A_i(b, \pi_n^1\pi_m^2f) + \left\{ \sum_{i=1}^2 \pi_n^1(a_i^2(b, \pi_m^2f)) + \sum_{i=1}^2 \pi_m^2(a_i^1(b, \pi_n^1f)) - \sum_{i=5}^8 A_i(b, \pi_m^2\pi_n^1f) \right\} \\ &+ \sum_{K,V} \left[a_K a_V \langle [\langle b \rangle_{V,2} - \langle b \rangle_{K \times V}] \langle f \rangle_{V,2} \rangle_K h_K \otimes h_V \right. \\ &\left. + a_K a_V \langle [\langle b \rangle_{K,1} - \langle b \rangle_{K \times V}] \langle f \rangle_{K,1} \rangle_V h_K \otimes h_V - a_K a_V \langle [b - \langle b \rangle_{K \times V}] f \rangle_{K \times V} h_K \otimes h_V \right]. \end{aligned}$$

The terms with A_i , $i = 1, \dots, 4$, are readily in control. The terms inside the bracket are paired in the natural way: a_1^2 goes with A_5 , a_2^2 goes with A_6 , a_1^1 goes with A_7 and a_2^1 goes with A_8 . We study

$$\pi_n^1(a_1^2(b, \pi_m^2f)) - A_5(b, \pi_m^2\pi_n^1f) = \pi_n^1(a_1^2(b, \pi_m^2f)) - A_5(b, \pi_n^1\pi_m^2f).$$

As usual it is enough to study $\pi_n^1(a_1^2(b, f)) - A_5(b, \pi_n^1f)$. We see that it equals

$$\sum_{K,J} a_K \langle [\langle b, h_J \rangle_2 - \langle b, h_J \rangle_2] \langle f, h_J \rangle_2 \rangle_K h_K \otimes h_J h_J.$$

Expanding the product inside the average we see that this further equals

$$\sum_K a_K \sum_{I \subset K} \langle h_I h_I \rangle_K \langle b, h_I \otimes h_J \rangle \langle f, h_I \otimes h_J \rangle h_K \otimes h_J h_J.$$

Dualising this against a function g we are left with estimating as follows:

$$\begin{aligned} & \sum_K \frac{|a_K|}{|K|} \sum_{I \subset K} \langle b, h_I \otimes h_J \rangle \langle f, h_I \otimes h_J \rangle \langle g, h_K \rangle_1 \langle \cdot \rangle_J \\ & \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \int_{\mathbb{R}^m} \sum_K |a_K| M_{\mathcal{D}^m} \langle g, h_K \rangle_1 \langle S_{\mathcal{D}} f \cdot \nu \rangle_{K,1} \\ & \lesssim \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \iint_{\mathbb{R}^n \times \mathbb{R}^m} M_{\mathcal{D}^n}^1(S_{\mathcal{D}} f \cdot \nu) S_{\mathcal{D}, M}^1 g \\ & \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}. \end{aligned}$$

The three other natural terms coming from this bracket are estimated in a similar way.

Regarding the last term in (4.3) we note that after expanding the products it simply equals

$$\sum_{K,V} a_K a_V \sum_{\substack{I \subset K \\ J \subset V}} \langle h_I h_I \rangle_K \langle h_J h_J \rangle_V \langle b, h_I \otimes h_J \rangle \langle f, h_I \otimes h_J \rangle h_K \otimes h_V.$$

We dualise and estimate as follows:

$$\begin{aligned} & \sum_{K,V} \frac{|a_K|}{|K|} \frac{|a_V|}{|V|} |\langle g, h_K \otimes h_V \rangle| \sum_{\substack{I \subset K \\ J \subset V}} |\langle b, h_I \otimes h_J \rangle| |\langle f, h_I \otimes h_J \rangle| \\ & \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \sum_{K,V} |a_K| |a_V| \langle S_{\mathcal{D}} f \cdot \nu \rangle_{K \times V} |\langle g, h_K \otimes h_V \rangle| \\ & \lesssim \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \iint_{\mathbb{R}^n \times \mathbb{R}^m} \left(\sum_{K,V} \langle S_{\mathcal{D}} f \cdot \nu \rangle_{K \times V}^2 |\langle g, h_K \otimes h_V \rangle|^2 \frac{1_K \otimes 1_V}{|K||V|} \right)^{1/2} \\ & \leq \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \iint_{\mathbb{R}^n \times \mathbb{R}^m} M_{\mathcal{D}}(S_{\mathcal{D}} f \cdot \nu) S_{\mathcal{D}} g \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}. \end{aligned}$$

We have shown

$$\|[\pi_n^1, [b, \pi_m^2]]\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

The case $[S_n^1, [b, \pi_m^2]]$. Here we consider the mixed case that we have a one-parameter shift and a paraproduct like above. Expanding $I = S_n^1(b\pi_m^2 f)$, $II = S_n^1\pi_m^2(bf)$, $III = b\pi_m^2 S_n^1 f$ and $IV = \pi_m^2(bS_n^1 f)$ similarly as above we get

$$\begin{aligned} [S_n^1, [b, \pi_m^2]] &= I - II - III + IV = \sum_{i=1}^4 S_n^1(A_i(b, \pi_m^2 f)) - \sum_{i=1}^4 A_i(b, \pi_m^2 S_n^1 f) \\ &+ \left\{ \sum_{i=5}^8 S_n^1(A_i(b, \pi_m^2 f)) + \sum_{i=1}^2 \pi_m^2(a_i^1(b, S_n^1 f)) - \sum_{i=5}^8 A_i(b, \pi_m^2 S_n^1 f) - \sum_{i=1}^2 S_n^1 \pi_m^2(a_i^1(b, f)) \right\} \\ &+ \sum_V a_V \sum_K \sum_{\substack{J \subset V \\ I_i^{(k_i)} = K}} \langle h_J h_J \rangle_V a_{K, (I_i)} [\langle \langle b, h_J \rangle_2 \rangle_{I_2} - \langle \langle b, h_J \rangle_2 \rangle_{I_1}] \langle f, h_{I_1} \otimes h_J \rangle h_{I_2} \otimes h_V. \end{aligned}$$

We handle the last term first. As usual, we reduce to estimating, for a fixed $k = 1, \dots, k_1$, as follows:

$$\begin{aligned} & \sum_V \frac{|a_V|}{|V|} \sum_K \sum_{\substack{I^{(k_1-k)} = K \\ J \subset V}} |\langle b, h_I \otimes h_J \rangle| |I|^{-1/2} \sum_{\substack{I_1^{(k)} = I \\ I_2^{(k_2)} = K}} |a_{K, (I_i)}| |\langle f, h_{I_1} \otimes h_J \rangle| |\langle g, h_{I_2} \otimes h_V \rangle| \\ & \leq \sum_V |a_V| |V|^{-1/2} \sum_K \sum_{\substack{I^{(k_1-k)} = K \\ J \subset V}} |\langle b, h_I \otimes h_J \rangle| |I|^{1/2} |J|^{1/2} \langle |\Delta_J^2 f| \rangle_{I \times J} \langle |\Delta_{K \times V}^{k_2, 0} g| \rangle_{K \times V}. \end{aligned}$$

This is further dominated in the $\lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}}$ sense by $\|b\|_{\text{BMO}_{\text{prod}}(\nu)}$ multiplied with

$$\int_{\mathbb{R}^n} \sum_V |a_V| |V|^{1/2} \left\langle \left(\sum_J [M_{\mathcal{D}} \Delta_J^2 f]^2 \right)^{1/2} \nu \right\rangle_{V, 2} \left(\sum_K \langle |\Delta_{K \times V}^{k_2, 0} g| \rangle_{K \times V}^2 1_K \right)^{1/2}$$

$$\begin{aligned} &\lesssim \iint_{\mathbb{R}^n \times \mathbb{R}^m} M_{\mathcal{D}^m}^2 \left(\left(\sum_J [M_{\mathcal{D}} \Delta_J^2 f]^2 \right)^{1/2} \nu \right) \left(\sum_{K,V} [M_{\mathcal{D}} \Delta_{K \times V}^{k_2,0} g]^2 \right)^{1/2} \\ &\lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}. \end{aligned}$$

The sums where $i = 1, \dots, 4$ are obviously bounded. So it remains to bound the terms inside the bracket. This requires appropriately grouping the terms into pairs of differences. We simply pair a_1^1 with A_7 , a_2^1 with A_8 , and then we pair the remaining A_5 terms together and the A_6 terms together. This easily (after taking operators as common factors as usually) reduces to terms that we have already bounded. Therefore, we are done:

$$\| [S_n^1, [b, \pi_m^2]] \|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} (1 + \max k_i) \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

Rest of the cases. By duality and symmetry the only case we have to still consider is $[\pi_n^1, [b, \pi_m^2]]$, where this time one of the paraproducts is in the dual form and one is not:

$$\pi_n^1 f = \sum_K a_K \frac{1_K}{|K|} \otimes \langle f, h_K \rangle_1, \quad \pi_m^2 f = \sum_V a_V \langle f \rangle_{V,2} \otimes h_V.$$

When we decompose $[\pi_n^1, [b, \pi_m^2]]$, the usual terms with the A_i , a_i^1 and a_i^2 produce no surprises. What is left can be written in the form $E_1 + E_2$, where

$$\begin{aligned} E_1 &= \pi_n^1(W(b, \pi_m^2 f)) - \pi_n^1 \pi_m^2(w^1(b, f)) \\ &= - \sum_{K,V} a_K a_V \langle [\langle b \rangle_{K,1} - \langle b \rangle_{K \times V}] \langle f, h_K \rangle_1 \rangle_V \frac{1_K}{|K|} \otimes h_V \\ &= - \sum_{K,V} \sum_{J \subset V} a_K a_V \langle h_J h_J \rangle_V \langle f, h_K \otimes h_J \rangle \langle \langle b, h_J \rangle_2 \rangle_K \frac{1_K}{|K|} \otimes h_V \end{aligned}$$

and

$$\begin{aligned} E_2 &= -w^2(b, \pi_n^1 \pi_m^2 f) + \pi_m^2(b \pi_n^1 f) \\ &= \sum_{K,V} a_K a_V \langle (b - \langle b \rangle_{V,2}) \langle f, h_K \rangle_1 \rangle_{V,2} \frac{1_K}{|K|} \otimes h_V \\ &= \sum_{K,V} \sum_{J \subset V} a_K a_V \langle h_J h_J \rangle_V \langle f, h_K \otimes h_J \rangle \langle b, h_J \rangle_2 \frac{1_K}{|K|} \otimes h_V. \end{aligned}$$

Therefore, we have

$$\begin{aligned} E_1 + E_2 &= \sum_{K,V} \sum_{J \subset V} a_K a_V \langle h_J h_J \rangle_V \langle f, h_K \otimes h_J \rangle [\langle b, h_J \rangle_2 - \langle \langle b, h_J \rangle_2 \rangle_K] \frac{1_K}{|K|} \otimes h_V \\ &= \sum_{K,V} \sum_{\substack{I \subset K \\ J \subset V}} a_K a_V \langle h_J h_J \rangle_V |K|^{-1} \langle b, h_I \otimes h_J \rangle \langle f, h_K \otimes h_J \rangle h_I \otimes h_V. \end{aligned}$$

We now estimate this by duality, and first arrive at the obvious upper bound

$$(4.4) \quad \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \sum_{K,V} \frac{|a_K|}{|K|} \frac{|a_V|}{|V|} \iint_{K \times V} [S_{\mathcal{D}^n} \langle g, h_V \rangle_2 \otimes S_{\mathcal{D}^m} \langle f, h_K \rangle_1] \nu.$$

Define the auxiliary functions

$$\varphi_S^1(f) = \sum_K h_K \otimes S_{\mathcal{D}^m} \langle f, h_K \rangle_1 \text{ and } \varphi_S^2(g) = \sum_V S_{\mathcal{D}^n} \langle g, h_V \rangle_2 \otimes h_V.$$

Define also $\tilde{\pi}_n^1 f$ to be the same function as $\pi_n^1 f$ except that a_K is replaced by $|a_K|$. Define $\tilde{\pi}_m^2 g$ via the formula

$$\tilde{\pi}_m^2 g = \sum_V |a_V| \langle g, h_V \rangle_2 \otimes \frac{1_V}{|V|}.$$

That is, this is not π_m^2 where a_V is replaced by $|a_V|$, but rather the dual of π_m^2 where a_V is replaced by $|a_V|$ (using the definitions of paraproducts that are valid in this subsection). Notice that (4.4) equals

$$\begin{aligned} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} & \iint_{\mathbb{R}^n \times \mathbb{R}^m} \tilde{\pi}_n^1(\varphi_S^1(f)) \tilde{\pi}_m^2(\varphi_S^2(g)) \nu \\ & \leq \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \|\tilde{\pi}_n^1(\varphi_S^1(f))\|_{L^p(\mu)} \|\tilde{\pi}_m^2(\varphi_S^2(g))\|_{L^{p'}(\lambda^{1-p'})} \\ & \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}. \end{aligned}$$

Here we used the weighted boundedness of the paraproducts and the operators φ_S^i . For the latter, notice e.g. that $S_{\mathcal{D}^n}^1 \varphi_S^1 f = S_{\mathcal{D}} f$. We are done with the proof. \square

5. LOWER BOUNDS FOR COMMUTATORS $[b, \dots [b, [b, T]] \dots]$

Let K be a standard bi-parameter full kernel as in [34]. We also assume that K is uniformly non-degenerate. In our setup this means that for all $y \in \mathbb{R}^{n+m}$ and $r_1, r_2 > 0$ there exists $x \in \mathbb{R}^{n+m}$ such that $|x_1 - y_1| > r_1, |x_2 - y_2| > r_2$ and

$$(5.1) \quad |K(x, y)| \gtrsim \frac{1}{r_1^n r_2^m}.$$

For example, we can have

$$K(x, y) = K_{i,j}(x, y) = \frac{x_{1,i} - y_{1,i}}{|x_1 - y_1|^{n+1}} \frac{x_{2,j} - y_{2,j}}{|x_2 - y_2|^{m+1}}.$$

That is, $K = K_{i,j}$ is the full kernel of the bi-parameter Riesz transform $R_i^n \otimes R_j^m$, $i = 1, \dots, n, j = 1, \dots, m$. Regarding the assumed Hölder conditions of the kernel K , similarly as in [22], a weaker modulus of continuity should be enough, but we do not pursue this.

We record that (5.1) implies the following: given a rectangle $R = I \times J$ there exists a rectangle $\tilde{R} = \tilde{I} \times \tilde{J}$ such that $\ell(I) = \ell(\tilde{I}), \ell(J) = \ell(\tilde{J}), d(I, \tilde{I}) \sim \ell(I), d(J, \tilde{J}) \sim \ell(J)$ and such that for some $\sigma \in \mathbb{C}$ with $|\sigma| = 1$ we have for all $x \in \tilde{R}$ and $y \in R$ that

$$\text{Re } \sigma K(x, y) \gtrsim \frac{1}{|R|}.$$

This can be seen as follows. Let, for a big enough constant A , the centre $c_{\tilde{R}}$ of \tilde{R} to be the point x given by (5.1) applied to $y = c_R, r_1 = A\ell(I)$ and $r_2 = A\ell(J)$. Choose σ so that $\sigma K(c_{\tilde{R}}, c_R) = |K(c_{\tilde{R}}, c_R)|$. Finally, use mixed Hölder and size estimates repeatedly.

Let $k \geq 1$ and $b \in L_{\text{loc}}^k(\mathbb{R}^{n+m})$ be real-valued. Let $p > 1$ and $\mu, \lambda \in A_p$. Define $\Gamma = \Gamma(K, b, \mu, \lambda, k, p)$ via the formula

$$\Gamma = \sup \frac{1}{\mu(R)^{1/p}} \left\| x \mapsto 1_{\tilde{R}}(x) \int_A (b(x) - b(y))^k K(x, y) dy \right\|_{L^{p, \infty}(\lambda)},$$

where the supremum is taken over all rectangles R, \tilde{R} with $\ell(I) = \ell(\tilde{I})$, $\ell(J) = \ell(\tilde{J})$, $d(I, \tilde{I}) \sim \ell(I)$ and $d(J, \tilde{J}) \sim \ell(J)$, and all subsets $A \subset R$.

The moral is simply the following. Notice that

$$1_{\tilde{R}}(x) \int_A (b(x) - b(y))^k K(x, y) dy = 1_{\tilde{R}}(x) [b, \dots [b, [b, T]] \dots](1_A)(x),$$

if T is a bi-parameter singular integral with a full kernel K . If this iterated commutator maps $L^p(\mu) \rightarrow L^{p, \infty}(\lambda)$, then Γ is dominated by this norm. However, the constant Γ is significantly weaker and depends only on the kernel K and some off-diagonal assumptions.

The following proposition supplies the lower bound related to [33] in the case $b_1 = \dots = b_k = b$. See Equation (1.1) in the Introduction.

5.2. Proposition. *Suppose K is a uniformly non-degenerate bi-parameter full kernel, $k \geq 1$ and $b \in L_{\text{loc}}^k(\mathbb{R}^{n+m})$ is real-valued. Let $p > 1$, $\mu, \lambda \in A_p$ and $\nu = \mu^{1/p} \lambda^{-1/p}$. Then for $\Gamma = \Gamma(K, b, \mu, \lambda, k, p)$ we have*

$$\|b\|_{\text{bmo}(\nu^{1/k})} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \Gamma^{1/k}.$$

Proof. Let $R = I \times J$ be a fixed rectangle. As we saw above we can find $\tilde{R} = \tilde{I} \times \tilde{J}$ such that $\ell(I) = \ell(\tilde{I})$, $\ell(J) = \ell(\tilde{J})$, $d(I, \tilde{I}) \sim \ell(I)$, $d(J, \tilde{J}) \sim \ell(J)$ and such that for some $\sigma \in \mathbb{C}$ with $|\sigma| = 1$ we have for all $x \in \tilde{R}$ and $y \in R$ that

$$\text{Re } \sigma K(x, y) \gtrsim \frac{1}{|R|}.$$

For $t \in \mathbb{R}$ let $t_+ = \max(t, 0)$. For an arbitrary $\alpha \in \mathbb{R}$ and $x \in \tilde{R} \cap \{b \geq \alpha\}$ we have

$$\begin{aligned} \left(\frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k &\leq \frac{1}{|R|} \int_{R \cap \{b \leq \alpha\}} (b(x) - b(y))^k dy \\ &\lesssim \text{Re } \sigma \int_{R \cap \{b \leq \alpha\}} (b(x) - b(y))^k K(x, y) dy. \end{aligned}$$

Now, let α be a median of b on \tilde{R} so that

$$\min\{|\tilde{R} \cap \{b \leq \alpha\}|, |\tilde{R} \cap \{b \geq \alpha\}|\} \geq \frac{|\tilde{R}|}{2} = \frac{|R|}{2}.$$

In particular, it follows that

$$\frac{\lambda(\tilde{R} \cap \{b \geq \alpha\})}{\lambda(\tilde{R})} \geq [\lambda]_{A_p}^{-1} \left(\frac{|\tilde{R} \cap \{b \geq \alpha\}|}{|\tilde{R}|} \right)^p \gtrsim_{[\lambda]_{A_p}} 1.$$

As $\lambda(MR) \lesssim_{M, [\lambda]_{A_p}} \lambda(R)$ we have $\lambda(R) \sim_{[\lambda]_{A_p}} \lambda(\tilde{R})$. We then get

$$\lambda(R)^{1/p} \left(\frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k \lesssim_{[\lambda]_{A_p}} \lambda(\tilde{R} \cap \{b \geq \alpha\})^{1/p} \left(\frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k$$

$$\begin{aligned} &\lesssim \left\| x \mapsto 1_{\widehat{R}}(x) \int_{R \cap \{b \leq \alpha\}} (b(x) - b(y))^k K(x, y) dy \right\|_{L^{p, \infty}(\lambda)} \\ &\leq \Gamma \mu(R)^{1/p}. \end{aligned}$$

We have

$$\begin{aligned} 1 &= \left\langle \nu^{\frac{1}{k(k+1)} \cdot k} \nu^{-\frac{1}{k+1}} \right\rangle_R^{k+1} \\ &\leq \left\langle \nu^{1/k} \right\rangle_R^k \left\langle \nu^{-1} \right\rangle_R \leq \left\langle \nu^{1/k} \right\rangle_R^k \langle \lambda \rangle_R^{1/p} \langle \mu^{1-p'} \rangle_R^{1/p'} \leq \left\langle \nu^{1/k} \right\rangle_R^k [\mu]_{A_p}^{1/p} \langle \mu \rangle_R^{-1/p} \langle \lambda \rangle_R^{1/p}, \end{aligned}$$

and so

$$\mu(R)^{1/p} \lambda(R)^{-1/p} \lesssim_{[\mu]_{A_p}} \left\langle \nu^{1/k} \right\rangle_R^k.$$

Combining everything we get

$$\left(\frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \left\langle \nu^{1/k} \right\rangle_R^k \Gamma.$$

So we have proved $\int_R (\alpha - b)_+ \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \nu^{1/k}(R) \Gamma^{1/k}$, and the bound $\int_R (b - \alpha)_+ \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \nu^{1/k}(R) \Gamma^{1/k}$ is proved analogously. The claim $\|b\|_{\text{bmo}(\nu^{1/k})} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \Gamma^{1/k}$ follows. \square

REFERENCES

- [1] A. Barron, J. Pipher, Sparse domination for bi-parameter operators using square functions, preprint, arXiv:1709.05009, 2017.
- [2] S. Bloom, A commutator theorem and weighted BMO, *Trans. Amer. Math. Soc.* 292 (1985) 103–122.
- [3] R. Coifman, R. Rochberg, G. Weiss, Factorization theorems for Hardy spaces in several variables, *Ann. of Math. (2)* 103 (1976) 611–635.
- [4] D. Cruz-Uribe, C.J. Neugebauer, The structure of the reverse Hölder classes, *Trans. Amer. Math. Soc.* 347 (1995) 2941–2960.
- [5] D. Cruz-Uribe, J.M. Martell, C. Pérez, Sharp weighted estimates for classical operators, *Adv. Math.* 229(2012) 408–441.
- [6] S.-Y. A. Chang, R. Fefferman, A continuous version of duality of H^1 with BMO on the Bidisc, *Ann. of Math.* 112 (1980) 179–201.
- [7] S.-Y. A. Chang, R. Fefferman, Some recent developments in Fourier analysis and H^p theory on product domains, *Bull. Amer. Math. Soc.* 12 (1985) 1–43.
- [8] S.-Y. A. Chang, J. M. Wilson, T. H. Wolff, Some weighted norm inequalities concerning the Schrödinger operators, *Comment. Math. Helv.* 60 (1985) 217–246.
- [9] L. Dalenc, Y. Ou, Upper bound for multi-parameter iterated commutators *Publ. Mat.* 60 (2016) 191–220.
- [10] X.T. Duong, J. Li, Y. Ou, J. Pipher, B. Wick, Commutators of multi-parameter flag singular integrals and applications, *Anal. PDE* 12 (2019) 1325–1355.
- [11] S. H. Ferguson, M. T. Lacey, A characterization of product BMO by commutators, *Acta Math.* 189 (2002) 143–160.
- [12] S. Ferguson, 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.
- [13] A. Grau de la Herrán, Comparison of $T1$ conditions for multi-parameter operators, *Proc. Amer. Math. Soc.* 144 (2016) 2437–2443.
- [14] I. Holmes, M. Lacey, B. Wick, Commutators in the two-weight setting, *Math. Ann.* 367 (2017) 51–80.
- [15] I. Holmes, M. Lacey, B. Wick, Bloom’s inequality: commutators in a two-weight setting, *Arch. Math. (Basel)* 106 (2016) 53–63.
- [16] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators, *Anal. PDE* 11 (2018) 1693–1740.
- [17] I. Holmes, B. Wick, Two weight inequalities for iterated commutators with Calderón-Zygmund operators, *J. Operator Theory* 79 (2018) 33–54.

- [18] T. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, *Ann. of Math.* 175 (2012) 1473–1506.
- [19] T. Hytönen, Representation of singular integrals by dyadic operators, and the A_2 theorem, *Expo. Math.* 35 (2017) 166–205.
- [20] T. Hytönen, The two-weight inequality for the Hilbert transform with general measures, *Proc. Lond. Math. Soc.* 117 (2018) 483–526.
- [21] T. Hytönen, The Holmes-Wick theorem on two-weight bounds for higher order commutators revisited, *Arch. Math. (Basel)* 107 (2016) 389–395.
- [22] T. Hytönen, The L^p -to- L^q boundedness of commutators with applications to the Jacobian operator, preprint, arXiv:1804.11167, 2018.
- [23] J.-L. Journé, Calderón-Zygmund operators on product spaces, *Rev. Mat. Iberoam.* 1 (1985) 55–91.
- [24] I. Kunwar, Y. Ou, Two-weight inequalities for multilinear commutators, *New York J. Math.* 24 (2018) 980–1003.
- [25] M. Lacey, Two weight inequality for the Hilbert transform: a real variable characterization, II, *Duke Math. J.* 163 (2014) 2821–2840.
- [26] M. Lacey, E. Sawyer, I. Uriarte-Tuero, C.-Y. Shen, Two weight inequality for the Hilbert transform: a real variable characterization, I, *Duke Math. J.* 163 (2014) 2795–2820.
- [27] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Multiparameter Riesz commutators, *Amer. J. Math.* 131 (2009) 731–769.
- [28] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Iterated Riesz Commutators: A Simple Proof of Boundedness, *Contemp. Math.*, 505, Amer. Math. Soc., Providence, RI, 2010, 171–178 .
- [29] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Multi-parameter Div-Curl Lemmas, *Bull. London Math. Soc.* 10 (2012) 1123–1131.
- [30] A. Lerner, S. Ombrosi, I. Rivera-Ríos, On pointwise and weighted estimates for commutators of Calderón-Zygmund operators, *Adv. Math.* 319 (2017) 153–181.
- [31] A. Lerner, S. Ombrosi, I. Rivera-Ríos, Commutators of singular integrals revisited, *Bull. London Math. Soc.* 51 (2019) 107–119.
- [32] K. Li, H. Martikainen, E. Vuorinen, Bilinear Calderón-Zygmund theory on product spaces, arXiv:1712.08135, 2018.
- [33] K. Li, H. Martikainen, E. Vuorinen, Bloom type inequality for bi-parameter singular integrals: efficient proof and iterated commutators, *Int. Math. Res. Not. IMRN*, to appear, arXiv:1806.02742, 2018.
- [34] H. Martikainen, Representation of bi-parameter singular integrals by dyadic operators, *Adv. Math.* 229 (2012) 1734–1761.
- [35] Y. Ou, S. Petermichl, E. Strouse, Higher order Journé commutators and characterizations of multi-parameter BMO, *Adv. Math.* 291 (2016) 24–58.

(K.L.) BCAM (BASQUE CENTER FOR APPLIED MATHEMATICS), ALAMEDA DE MAZARREDO 14, 48009 BILBAO, SPAIN

E-mail address: kangwei.nku@gmail.com

(H.M.) DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI, P.O.B. 68, FI-00014 UNIVERSITY OF HELSINKI, FINLAND

E-mail address: henri.martikainen@helsinki.fi

(E.V.) CENTRE FOR MATHEMATICAL SCIENCES, UNIVERSITY OF LUND, P.O.B. 118, 22100 LUND, SWEDEN

E-mail address: j.e.vuorin@gmail.com