TWO-WEIGHT INEQUALITIES FOR MULTILINEAR COMMUTATORS IN PRODUCT SPACES

EMIL AIRTA, KANGWEI LI, AND HENRI MARTIKAINEN

ABSTRACT. This note is devoted to establishing two-weight estimates for commutators of singular integrals. We combine multilinearity with product spaces. A new type of two-weight extrapolation result is used to yield the quasi-Banach range of estimates.

1. INTRODUCTION

Commutators have the general form $[b,T]\colon f \mapsto bTf - T(bf)$. Here T is a singular integral operator

$$Tf(x) = \int_{\mathbb{R}^d} K(x,y)f(y)\,\mathrm{d}y.$$

Well-known examples include the Hilbert transform H in dimension $d=1$, which has the kernel $K(x,y) = \frac{1}{x-y}$, and the Riesz transforms R_j in dimensions $d \geq 2$, which have the kernel $K_j(x,y) = \frac{x_j - y_j}{|x-y|^{d+1}}$, $j = 1, \ldots, d$.

Our work revolves around the Coifman–Rochberg–Weiss [4] result, where the two-sided estimate

$$\|b\|_{\mathrm{BMO}} \lesssim \|[b,T]\|_{L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)} \lesssim \|b\|_{\mathrm{BMO}}, \qquad p \in (1,\infty),$$

was proved for a class of non-degenerate singular integrals T on $\mathbb{R}^d$. Here BMO stands for functions of bounded mean oscillation:

$$\|b\|_{\mathrm{BMO}} := \sup_I \fint_I |b - \langle b \rangle_I|,$$

where the supremum is over all cubes $I \subset \mathbb{R}^d$ and $\langle b \rangle_I = \fint_I b := \frac{1}{|I|}\int_I b$. The corresponding two-weight problem concerns estimates from $L^p(\mu)$ to $L^p(\lambda)$ for two different weights μ, λ and has recently attracted interest after the work by Holmes–Lacey–Wick [10]. See also e.g. [11, 14, 15].

In this note we establish that two-weight estimates for commutators can be proved under the joint difficulty of multilinearity and product spaces. Both have been considered separately before: see e.g. [1, 2, 10, 20, 22] for the multi-parameter work, and [12] and

---

2010 *Mathematics Subject Classification.* 42B20.

Key words and phrases. singular integrals, multilinear analysis, multi-parameter analysis, two-weight estimates, commutators.

E. A. was supported by the Emil Aaltonen Foundation and by the Academy of Finland through the grant 327271.

K. L. was supported by the National Natural Science Foundation of China through project number 12001400.

While at the University of Helsinki H. M. was supported by the Academy of Finland through project numbers 294840 and 327271, and by the three-year research grant 75160010 of the University of Helsinki.

[17] for the multilinear work. The recent satisfactory multilinear result of [17] is based on sparse domination and the approach cannot be used in our setting – this is due to the product space nature of the problem. For given exponents $1 < p_1, \dots, p_n \le \infty$ and $1/p = \sum_i 1/p_i > 0$, a natural form of a weighted estimate in the n-variable context has the form

$$\Big\| T(f_1, \dots, f_n) \prod_{i=1}^n w_i \Big\|_{L^p} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}.$$

The key thing is to only impose a *joint* condition on the tuple of weights $\vec{w} = (w_1, \dots, w_n) \in A_{\vec{p}}$ rather than to assume individual conditions $w_i^{p_i} \in A_{p_i}$. See Lerner, Ombrosi, Pérez, Torres and Trujillo-González [13] and for multi-parameter versions [21]. Naturally, this interplay is trickier still in our two-weight setting.

Our result is the following.

1.1. **Theorem.** *Let T be an n-linear bi-parameter Calderón-Zygmund operator. Let $\vec{p} = (p_1, \dots, p_n)$ with $1 < p_i \le \infty$ and*

$$\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i} > 0.$$

With a fixed $j \in \{1, \dots, n\}$ let $(w_1, \dots, w_n)$ and $(w_1, \dots, \lambda_j, \dots, w_n)$ be two tuples of weights in the genuinely multilinear bi-parameter weight class $A_{\vec{p}}$ and define the associated Bloom weight $\nu = w_j \lambda_j^{-1}$. If we have $b \in \operatorname{bmo}(\nu)$ and $\nu \in A_\infty$, then

$$\|[b, T]_j(f_1, \dots, f_n)\nu^{-1} w\|_{L^p} \lesssim \|b\|_{\operatorname{bmo}(\nu)} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}, \quad w = \prod_{i=1}^n w_i.$$

The corresponding lower bound holds if T is suitably non-degenerate.

For the exact definitions see the main text.

Extrapolation methods are important in our current work – they are used to yield the quasi-Banach range $p < 1$. The extrapolation theorem of Rubio de Francia says that if $\|g\|_{L^{p_0}(w)} \lesssim \|f\|_{L^{p_0}(w)}$ for some $p_0 \in (1, \infty)$ and all $w \in A_{p_0}$, then $\|g\|_{L^p(w)} \lesssim \|f\|_{L^p(w)}$ for all $p \in (1, \infty)$ and all $w \in A_p$. In [9] (see also [6]) a multivariable analogue was developed in the setting $w_i^{p_i} \in A_{p_i}$, $i = 1, \dots, n$. Very recently, in [18, 19, 24] it was shown that also the genuinely multilinear weighted estimates can be extrapolated. We prove a suitable two-weight adaptation that can be used in our current work.

1.2. **Theorem.** *Let $(f, f_1, \dots, f_n)$ be a tuple of non-negative functions. Let $1 \le p_i \le \infty$, $1 \le i \le n$, $\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i}$, and $j \in \{1, \dots, n\}$. Assume that for all $(w_1, \cdots, w_n)$, $(w_1, \dots, \lambda_j, \dots, w_n) \in A_{\vec{p}}$ with $w_j \lambda_j^{-1} \in A_\infty$, there holds that*

$$\Big\| f \lambda_j \prod_{\substack{i=1 \\ i \ne j}}^n w_i \Big\|_{L^p} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}.$$

Then for all $(w_1, \cdots, w_n), (w_1, \ldots, \lambda_j, \cdots, w_n) \in A_{\vec{q}}$ *with* $w_j \lambda_j^{-1} \in A_\infty$ *and* $1 < q_i \le \infty, i \ne j$, $1/q = 1/p_j + \sum_{\substack{i=1 \\ i \ne j}}^n 1/q_i > 0$, *there holds that*

$$\Big\| f\lambda_j \prod_{\substack{i=1 \\ i\ne j}}^n w_i \Big\|_{L^q} \lesssim \|f_j w_j\|_{L^{p_j}} \prod_{\substack{i=1 \\ i\ne j}}^n \|f_i w_i\|_{L^{q_i}}.$$

2. Preliminaries

Throughout this paper, $A \lesssim B$ means that $A \le CB$ with some constant C that we deem unimportant to track at that point. We write $A \sim B$ if $A \lesssim B \lesssim A$. Sometimes we e.g. write $A \lesssim_\epsilon B$ if we want to make the point that $A \le C(\epsilon)B$.

2.A. **Dyadic notation.** Given a dyadic grid $\mathcal{D}$ in $\mathbb{R}^d$, $I \in \mathcal{D}$ and $k \in \mathbb{Z}$, $k \ge 0$, we use the following notation:

(1) $\ell(I)$ is the side length of I.
(2) $I^{(k)} \in \mathcal{D}$ is the kth parent of I, i.e., $I \subset I^{(k)}$ and $\ell(I^{(k)}) = 2^k \ell(I)$.
(3) $\mathrm{ch}(I)$ is the collection of the children of I, i.e., $\mathrm{ch}(I) = \{J \in \mathcal{D} \colon J^{(1)} = I\}$.
(4) $E_I f = \langle f \rangle_I 1_I$ is the averaging operator, where $\langle f \rangle_I = \fint_I f = \frac{1}{|I|}\int_I f$.
(5) $\Delta_I f$ is the martingale difference $\Delta_I f = \sum_{J \in \mathrm{ch}(I)} E_J f - E_I f$.
(6) $\Delta_{I,k} f$ is the martingale difference block

$$\Delta_{I,k} f = \sum_{\substack{J \in \mathcal{D} \\ J^{(k)} = I}} \Delta_J f.$$

For an interval $J \subset \mathbb{R}$ we denote by J_l and J_r the left and right halves of J, respectively. We define $h_J^0 = |J|^{-1/2} 1_J$ and $h_J^1 = |J|^{-1/2}(1_{J_l} - 1_{J_r})$. Let now $I = I_1 \times \cdots \times I_d \subset \mathbb{R}^d$ be a cube, and define the Haar function h_I^η, $\eta = (\eta_1, \ldots, \eta_d) \in \{0,1\}^d$, by setting

$$h_I^\eta = h_{I_1}^{\eta_1} \otimes \cdots \otimes h_{I_d}^{\eta_d}.$$

If $\eta \ne 0$ the Haar function is cancellative: $\int h_I^\eta = 0$. We exploit notation by suppressing the presence of η, and write h_I for some h_I^η, $\eta \ne 0$. Notice that for $I \in \mathcal{D}$ we have $\Delta_I f = \langle f, h_I \rangle h_I$ (where the finite η summation is suppressed), $\langle f, h_I \rangle := \int f h_I$.

2.B. **Multi-parameter notation.** We will be working on the bi-parameter product space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We denote a general dyadic grid in $\mathbb{R}^{d_i}$ by $\mathcal{D}^i$. We denote cubes in $\mathcal{D}^i$ by I^i, J^i, K^i, etc. Thus, our dyadic rectangles take the forms $I^1 \times I^2$, $J^1 \times J^2$, $K^1 \times K^2$ etc. We usually denote the collection of dyadic rectangles by $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$.

If A is an operator acting on $\mathbb{R}^{d_1}$, we can always let it act on the product space $\mathbb{R}^d$ by setting $A^1 f(x) = A(f(\cdot, x_2))(x_1)$. Similarly, we use the notation $A^i f$ if A is originally an operator acting on $\mathbb{R}^{d_i}$. Our basic multi-parameter dyadic operators – martingale differences and averaging operators – are obtained by simply chaining together relevant one-parameter operators. For instance, a bi-parameter martingale difference is

$$\Delta_R f = \Delta_{I^1}^1 \Delta_{I^2}^2 f, \qquad R = I^1 \times I^2.$$

When we integrate with respect to only one of the parameters we may e.g. write

$$\langle f, h_{I^1} \rangle_1 (x_2) := \int_{\mathbb{R}^{d_1}} f(x_1, x_2) h_{I^1}(x_1) \, \mathrm{d}x_1$$

or

$$\langle f \rangle_{I^1,1}(x_2) := \fint_{I^1} f(x_1, x_2)\,\mathrm{d}x_1.$$

2.C. **Adjoints.** Consider an n-linear operator T on $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Let $f_i = f_i^1 \otimes f_i^2$, $i = 1, \dots, n+1$. We set up notation for the adjoints of T in the bi-parameter situation. We let T^{j*}, $j \in \{0, \dots, n\}$, denote the full adjoints, i.e., $T^{0*} = T$ and otherwise

$$\langle T(f_1, \dots, f_n), f_{n+1} \rangle = \langle T^{j*}(f_1, \dots, f_{j-1}, f_{n+1}, f_{j+1}, \dots, f_n), f_j \rangle.$$

A subscript 1 or 2 denotes a partial adjoint in the given parameter – for example, we define

$$\langle T(f_1, \dots, f_n), f_{n+1} \rangle = \langle T_1^{j*}(f_1, \dots, f_{j-1}, f_{n+1}^1 \otimes f_j^2, f_{j+1}, \dots, f_n), f_j^1 \otimes f_{n+1}^2 \rangle.$$

Finally, we can take partial adjoints with respect to different parameters in different slots also – in that case we denote the adjoint by $T_{1,2}^{j_1*, j_2*}$. It simply interchanges the functions $f_{j_1}^1$ and f_{n+1}^1 and the functions $f_{j_2}^2$ and f_{n+1}^2. Of course, we e.g. have $T_{1,2}^{j^*, j^*} = T^{j*}$ and $T_{1,2}^{0*, j^*} = T_2^{j*}$, so everything can be obtained, if desired, with the most general notation $T_{1,2}^{j_1*, j_2*}$. In any case, there are $(n+1)^2$ adjoints (including T itself). Similarly, the bi-parameter dyadic model operators that we later define always have $(n+1)^2$ different forms.

2.D. **Multilinear bi-parameter weights.** A weight $w(x_1, x_2)$ (i.e. a locally integrable a.e. positive function) belongs to the bi-parameter weight class $A_p = A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$, $1 < p < \infty$, if

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

where the supremum is taken over rectangles R – that is, over $R = I^1 \times I^2$ where $I^i \subset \mathbb{R}^{d_i}$ is a cube. In contrast to the one-parameter definition, we take supremum over rectangles instead of cubes.

We have

$$[w]_{A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})} < \infty \text{ iff } \max\Big(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{d_1}} [w(x_1, \cdot)]_{A_p(\mathbb{R}^{d_2})}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{d_2}} [w(\cdot, x_2)]_{A_p(\mathbb{R}^{d_1})} \Big) < \infty, \tag{2.1}$$

and that

$$\max\Big(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{d_1}} [w(x_1, \cdot)]_{A_p(\mathbb{R}^{d_2})}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{d_2}} [w(\cdot, x_2)]_{A_p(\mathbb{R}^{d_1})} \Big) \le [w]_{A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})},$$

while the constant $[w]_{A_p}$ is dominated by the maximum to some power. It is also useful that $\langle w \rangle_{I^2,2} \in A_p(\mathbb{R}^{d_1})$ uniformly on the cube $I^2 \subset \mathbb{R}^{d_2}$. For basic bi-parameter weighted theory see e.g. [10]. We say $w \in A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ if

$$[w]_{A_\infty} := \sup_R \langle w \rangle_R \exp\big(\langle \log w^{-1} \rangle_R \big) < \infty.$$

It is well-known that

$$A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}) = \bigcup_{1<p<\infty} A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}).$$

We also define

$$[w]_{A_1} = \sup_R \langle w \rangle_R \operatorname*{ess\,sup}_R w^{-1}.$$

The following multilinear reverse Hölder property is well-known – for the history and a very short proof see e.g. [17, Lemma 2.5]. The proof in our bi-parameter setting is the same.

2.2. **Lemma.** *Let $u_i \in (0,\infty)$ and $w_i \in A_\infty$, $i = 1,\dots,N$, be bi-parameter weights. Then for every rectangle R we have*

$$\prod_{i=1}^N \langle w_i\rangle_R^{u_i} \lesssim \Big\langle \prod_{i=1}^N w_i^{u_i}\Big\rangle_R.$$

Next we define multilinear bi-parameter Muckenhoupt weights. Original one-parameter versions appeared in [13]. Our definition in the bi-parameter case is the same as in [21].

2.3. **Definition.** Given $\vec p = (p_1,\dots,p_n)$ with $1 \le p_1,\dots,p_n \le \infty$ we say that $\vec w = (w_1,\dots,w_n) \in A_{\vec p} = A_{\vec p}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})$, if

$$0 < w_i < \infty, \qquad i = 1,\dots,n,$$

almost everywhere and

$$[\vec w]_{A_{\vec p}} := \sup_R \langle w^p\rangle_R^{\frac1p} \prod_{i=1}^n \langle w_i^{-p_i'}\rangle_R^{\frac{1}{p_i'}} < \infty,$$

where the supremum is over rectangles R,

$$w := \prod_{i=1}^n w_i \qquad \text{and} \qquad \frac1p = \sum_{i=1}^n \frac{1}{p_i}.$$

If $p_i = 1$ we interpret $\langle w_i^{-p_i'}\rangle_R^{\frac{1}{p_i'}}$ as $\operatorname{ess\,sup}_R w_i^{-1}$, and if $p = \infty$ we interpret $\langle w^p\rangle_R^{\frac1p}$ as $\operatorname{ess\,sup}_R w$.

Conveniently, we can characterize the class $A_{\vec p}$ using the standard A_p class. The lemma is proven in [13] and the bi-parameter analog of the same proof is recorded in [21].

2.4. **Lemma.** *Let $\vec p = (p_1,\dots,p_n)$ with $1 \le p_1,\dots,p_n \le \infty$, $1/p = \sum_{i=1}^n 1/p_i \ge 0$, $\vec w = (w_1,\dots,w_n)$ and $w = \prod_{i=1}^n w_i$. We have*

$$[w_i^{-p_i'}]_{A_{np_i'}} \le [\vec w]_{A_{\vec p}}^{p_i'}, \qquad i = 1,\dots,n,$$

and

$$[w^p]_{A_{np}} \le [\vec w]_{A_{\vec p}}^p.$$

In the case $p_i = 1$ the estimate is interpreted as $[w_i^{\frac1n}]_{A_1} \le [\vec w]_{A_{\vec p}}^{1/n}$, and in the case $p = \infty$ we have $[w^{-\frac1n}]_{A_1} \le [\vec w]_{A_{\vec p}}^{1/n}$.

Conversely, we have

$$[\vec w]_{A_{\vec p}} \le [w^p]_{A_{np}}^{\frac1p} \prod_{i=1}^n [w_i^{-p_i'}]_{A_{np_i'}}^{\frac{1}{p_i'}}.$$

Most of the proofs are duality based and this makes the following lemma relevant.

2.5. **Lemma** ([23, Lemma 3.1]). *Let* $\vec{p} = (p_1, \dots, p_n)$ *with* $1 < p_1, \dots, p_n < \infty$ *and* $\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i} \in (0,1)$. *Let* $\vec{w} = (w_1, \dots, w_n) \in A_{\vec{p}}$ *with* $w = \prod_{i=1}^n w_i$ *and define*

$$\begin{aligned}\vec{w}^{\,i} &= (w_1, \dots, w_{i-1}, w^{-1}, w_{i+1}, \dots, w_n),\\ \vec{p}^{\,i} &= (p_1, \dots, p_{i-1}, p', p_{i+1}, \dots, p_n).\end{aligned}$$

Then we have

$$[\vec{w}^{\,i}]_{A_{\vec{p}^{\,i}}} = [\vec{w}]_{A_{\vec{p}}}.$$

In the main theorems of this paper we will be using the multilinear bi-parameter weights

$$(w_1, \dots, w_n), (\lambda_1, w_2, \dots, w_n) \in A_{(p_1, \dots, p_n)}, \quad \text{and} \quad \nu := \lambda_1^{-1} w_1 \in A_\infty,$$

where $1 \le p_1, \dots, p_n \le \infty$, $1/p = \sum_{i=1}^n 1/p_i > 0$. Throughout this paper, we will be using notation $\sigma_i = w_i^{-p_i'}$, $\sigma_{n+1} = (\nu^{-1} w)^p$, and $\eta_1 = \lambda_1^{-p_1'}$ as they will appear regularly.

The assumption that $\nu \in A_\infty$ is necessary as it is not implied by the other two assumptions, see a counter-example in [17].

However, instead of the two separate conditions

$$(w_1, \dots, w_n) \in A_{(p_1, \dots, p_n)} \quad \text{and} \quad (\lambda_1, w_2, \dots, w_n) \in A_{(p_1, \dots, p_n)},$$

if we assume only that $(w_1, \dots, w_n, \nu w^{-1}) \in A_{(p_1, \dots, p_n, p')}$, where $\nu = \lambda_1^{-1} w_1$ and $w = \prod_{i=1}^n w_i$, that is

$$\sup_R \prod_{i=1}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \langle \nu^{-p} w^p \rangle_R^{\frac{1}{p}} \langle \nu \rangle_R < \infty,$$

we would automatically get that

$$\prod_{i=1}^n w_i \cdot \nu w^{-1} = \nu \in A_{n+1} \subset A_\infty$$

by Lemma 2.4.

Yet, it is unlikely that this assumption is enough for the boundedness of the commutator as conjectured for the linear case in [16]. Although, we will show below that this assumption is enough for the boundedness of Bloom type paraproducts in the Banach range and also sufficient to conclude the lower bound of the commutator.

On the other hand, the joint assumption for the weights is very natural for the two-weight commutator estimates since the assumption $(w_1, \dots, w_n, \nu w^{-1}) \in A_{(p_1, \dots, p_n, p')}$ is implied by the two separate multilinear weight conditions and $\nu \in A_\infty$.

This is easy to verify. Let $\sum_{i=1}^n \frac{1}{p_i} =: \frac{1}{p} > 1$ and assume that $(w_1, \dots, w_n), (\lambda_1, w_2, \dots, w_n) \in A_{(p_1, \dots, p_n)}$, and $\nu := \lambda_1^{-1} w_1 \in A_\infty$.

$$\begin{aligned}\prod_{i=1}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \langle (\nu w^{-1})^{-p} \rangle_R^{\frac{1}{p}} \langle \nu \rangle_R &= \prod_{i=1}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \langle \lambda_1^p \prod_{i=2}^n w_i^p \rangle_R^{\frac{1}{p}} \langle (\lambda_1^{-p_1'})^{\frac{1}{p_1'}} \prod_{i=2}^n (w_i^{-p_i'})^{\frac{1}{p_i'}} (w^p)^{\frac{1}{p}} \rangle_R \\ &\overset{(*)}{\lesssim} \prod_{i=1}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \langle \lambda_1^p \prod_{i=2}^n w_i^p \rangle_R^{\frac{1}{p}} \langle \lambda_1^{-p_1'} \rangle_R^{\frac{1}{p_1'}} \prod_{i=2}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \langle w^p \rangle_R^{\frac{1}{p}} \\ &\le [\vec{w}]_{A_{\vec{p}}} [(\lambda_1, w_2, \dots, w_n)]_{A_{\vec{p}}},\end{aligned}$$

where in the step $(*)$ we apply [17, Lemma 2.9] for $\nu \in A_\infty$.

Motivated by the above discussion we give the following definition, where p' does not appear hence $p > 1$ is not needed.

2.6. **Definition.** Given $\vec{p} = (p_1, \dots, p_n)$ with $1 \le p_1, \dots, p_n \le \infty$, we say that $\vec{w} = (w_1, \dots, w_n, w_{n+1}) \in A^*_{\vec{p}} = A^*_{\vec{p}}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$, if

$$0 < w_i < \infty, \qquad i = 1, \dots, n+1,$$

almost everywhere and

$$[\vec{w}]_{A^*_{\vec{p}}} := \sup_R \langle w \rangle_R \langle w_{n+1}^{-p} \rangle_R^{\frac{1}{p}} \prod_{i=1}^{n} \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} < \infty,$$

where the supremum is over rectangles R,

$$w := \prod_{i=1}^{n+1} w_i \qquad \text{and} \qquad \frac{1}{p} = \sum_{i=1}^{n} \frac{1}{p_i}.$$

If $p_i = 1$ we interpret $\langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}}$ as $\operatorname{ess\,sup}_R w_i^{-1}$, and if $p = \infty$ we interpret $\langle w^p \rangle_R^{\frac{1}{p}}$ as $\operatorname{ess\,sup}_R w$.

Morally the difference is that with $A^*_{\vec{p}}$ we do not necessarily have

$$\prod_{i=1}^{n} w_i^p \in A_\infty$$

or $\lambda_j^{-p_j} \in A_\infty$ compared to assuming the two separate $A_{\vec{p}}$ and $\nu \in A_\infty$ but we are equipped with $\nu \in A_{n+1}$.

Furthermore, using this definition, we can write the following joint condition

$$(w_1, \dots, w_n, \nu w^{-1}) \in A_{(p_1, \dots, p_n, p')}$$

as $(w_1, \dots, w_n, \nu w^{-1}) \in A^*_{(p_1, \dots, p_n)} = A^*_{\vec{p}}$.

A_∞ **extrapolation.** Besides of the extrapolation theorem proven in this paper, we also need to use the following A_∞-extrapolation result of [5].

2.7. **Lemma.** *Let (f, g) be a pair of non-negative functions. Suppose that there exists some $0 < p_0 < \infty$ such that for every $w \in A_\infty$ we have*

$$\int f^{p_0} w \lesssim \int g^{p_0} w. \tag{2.8}$$

Then for all $0 < p < \infty$ and $w \in A_\infty$ we have

$$\int f^p w \lesssim \int g^p w.$$

In addition, let $\{(f_i, g_i)\}_i$ be a sequence of pairs of non-negative functions defined on $\mathbb{R}^d$. Suppose that for some $0 < p_0 < \infty$, (f_i, g_i) satisfies inequality (2.8) for every i. Then, for all $0 < p, q < \infty$ and $w \in A_\infty(\mathbb{R}^d)$ we have

$$\Big\| \Big(\sum_i (f_i)^q \Big)^{\frac{1}{q}} \Big\|_{L^p(w)} \lesssim_{[w]_{A_\infty}} \Big\| \Big(\sum_i (g_i)^q \Big)^{\frac{1}{q}} \Big\|_{L^p(w)},$$

where $\{(f_j, g_j)\}_j$ is a sequence of pairs of non-negative functions defined on $\mathbb{R}^d$.

2.E. **Maximal functions.** Let $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$ be a fixed lattice of dyadic rectangles and define

$$M_{\mathcal{D}}(f_1, \dots, f_n) = \sup_{R \in \mathcal{D}} \prod_{i=1}^n \langle |f_i| \rangle_R 1_R.$$

2.9. **Proposition.** *If $1 < p_1, \dots, p_n \le \infty$ and $1/p = \sum_{i=1}^n 1/p_i$ we have*

$$\|M_{\mathcal{D}}(f_1, \dots, f_n) w\|_{L^p} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}$$

for all multilinear bi-parameter weights $\vec{w} \in A_{\vec{p}}$.

An efficient proof can be found in [21] (originally proved in [8]).

Also we often need the result of R. Fefferman [7]. Proof also recorded in [22, Appendix B]. Denote $\langle f \rangle_R^\mu := \frac{1}{\mu(R)} \int_R f \,\mathrm{d}\mu$ and define

$$M_{\mathcal{D}}^\mu f = \sup_R 1_R \langle |f| \rangle_R^\mu.$$

2.10. **Proposition.** *Let $\lambda \in A_p$, $p \in (1,\infty)$, be a bi-parameter weight. Then for all $s \in (1,\infty)$ we have*

$$\|M_{\mathcal{D}}^\lambda f\|_{L^s(\lambda)} \lesssim [\lambda]_{A_p}^{1+1/s} \|f\|_{L^s(\lambda)}.$$

2.F. **Square functions.** We begin with the classical (dyadic) square function in the bi-parameter framework. Let $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$ be a fixed lattice of dyadic rectangles. We define the square functions

$$S_{\mathcal{D}} f = \Big(\sum_{R \in \mathcal{D}} |\Delta_R f|^2 \Big)^{1/2}, \; S_{\mathcal{D}^1}^1 f = \Big(\sum_{I^1 \in \mathcal{D}^1} |\Delta_{I^1}^1 f|^2 \Big)^{1/2}$$

and define $S_{\mathcal{D}^2}^2 f$ analogously.

The lower bound estimate of the square function for A_∞ weights is essential for many estimates later on. The fact that the key weights w^p and $w_i^{-p_i'}$ are at least in A_∞ for the multilinear weights of Definition 2.3 allows us to use this lower bound estimate.

2.11. **Lemma.** *It holds*

$$\|f\|_{L^p(w)} \lesssim \|S_{\mathcal{D}^j}^j f\|_{L^p(w)} \lesssim \|S_{\mathcal{D}} f\|_{L^p(w)}$$

for all $p \in (0,\infty)$ and bi-parameter weights $w \in A_\infty$.

The first inequality is the classical result found e.g. in [25, Theorem 2.5] and the latter inequality can be deduced using the A_∞ extrapolation, Lemma 2.7.

Notice that by disjointness of supports we have, for example, for all $k = (k_1, k_2) \in \{0, 1, \dots\}^2$ that

$$S_{\mathcal{D}} f = \Big(\sum_{K = K^1 \times K^2 \in \mathcal{D}} |\Delta_{K,k} f|^2 \Big)^{1/2}, \qquad \Delta_{K,k} = \Delta_{K^1,k_1}^1 \Delta_{K^2,k_2}^2.$$

Next, we take the definition of the n-linear square functions from [21]. For $k = (k_1, k_2)$ we set

$$A_1(f_1, \dots, f_n) = A_{1,k}(f_1, \dots, f_n) = \Big(\sum_{K \in \mathcal{D}} \langle |\Delta_{K,k} f_1| \rangle_K^2 \prod_{i=2}^n \langle |f_i| \rangle_K^2 1_K \Big)^{\frac{1}{2}}.$$

In addition, we understand this so that $A_{1,k}$ can also take any one of the symmetric forms, where each $\Delta^j_{K^j,k_j}$ appearing in $\Delta_{K,k} = \Delta^1_{K^1,k_1}\Delta^2_{K^2,k_2}$ can alternatively be associated with any of the other functions $f_2, \dots, f_n$. That is, $A_{1,k}$ can, for example, also take the form

$$A_{1,k}(f_1,\dots,f_n) = \Big(\sum_{K\in\mathcal{D}} \langle|\Delta^2_{K^2,k_2} f_1|\rangle_K^2 \langle|\Delta^1_{K^1,k_1} f_2|\rangle_K^2 \prod_{i=3}^n \langle|f_i|\rangle_K^2 1_K\Big)^{\frac12}.$$

For $k=(k_1,k_2,k_3)$ we define

$$\begin{aligned}&A_{2,k}(f_1,\dots,f_n)\\ &= \Big(\sum_{K^2\in\mathcal{D}^2}\Big(\sum_{K^1\in\mathcal{D}^1} \langle|\Delta^2_{K^2,k_1} f_1|\rangle_K \langle|\Delta^1_{K^1,k_2} f_2|\rangle_K \langle|\Delta^1_{K^1,k_3} f_3|\rangle_K \prod_{i=4}^n \langle|f_i|\rangle_K 1_K\Big)^2\Big)^{\frac12},\end{aligned} \tag{2.12}$$

where we again understand this as a family of square functions. First, the appearing three martingale blocks can be associated with different functions, too. Second, we can have the K^1 summation out and the K^2 summation in (we can interchange them), but then we have two martingale blocks with K^2 and one martingale block with K^1.

Finally, for $k=(k_1,k_2,k_3,k_4)$ we define

$$A_{3,k}(f_1,\dots,f_n) = \sum_{K\in\mathcal{D}} \langle|\Delta_{K,(k_1,k_2)} f_1|\rangle_K \langle|\Delta_{K,(k_3,k_4)} f_2|\rangle_K \prod_{i=3}^n \langle|f_i|\rangle_K 1_K,$$

where this is a family with two martingale blocks in each parameter, which can be moved around.

2.13. **Theorem** ([21, Theorem 5.5.]). *If $1 < p_1,\dots,p_n \le \infty$ and $\frac1p = \sum_{i=1}^n \frac{1}{p_i} > 0$ we have*

$$\|A_{j,k}(f_1,\dots,f_n)w\|_{L^p} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}, \quad j=1,2,3,$$

for all multilinear bi-parameter weights $\vec{w} \in A_{\vec{p}}$.

Moreover, we need a certain linear estimate which appears regularly when dealing with the commutator estimates.

2.14. **Proposition** ([21, Proposition 5.8.]). *For $u \in A_\infty$ and $p,s \in (1,\infty)$ we have*

$$\Big\|\Big[\sum_m \Big(\sum_{K\in\mathcal{D}} \langle|\Delta_{K,k} f_m|\rangle_K^2 \frac{1_K}{\langle u\rangle_K^2}\Big)^{\frac s2}\Big]^{\frac1s} u^{\frac1p}\Big\|_{L^p} \lesssim \Big\|\Big(\sum_m |f_m|^s\Big)^{\frac1s} u^{-\frac{1}{p'}}\Big\|_{L^p}.$$

3. BMO SPACES

Let $\mathcal{D} = \mathcal{D}^1\times\mathcal{D}^2$ be a collection of dyadic rectangles on $\mathbb{R}^d = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$. For a function $b\in L^1_{\text{loc}}$ and a bi-parameter weight $\nu\in A_\infty$ we define the usual dyadic weighted little BMO norm of b as follows:

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

In fact, the direct definition is not used that often and we will mostly invoke it through the following H^1-BMO type inequalities. For $i=1$ and $i=2$ we have

$$|\langle b,f\rangle| \lesssim \|b\|_{\text{bmo}(\nu)}\|S^i_{\mathcal{D}^i} f\|_{L^1(\nu)} \lesssim \|b\|_{\text{bmo}(\nu)}\|S_{\mathcal{D}} f\|_{L^1(\nu)}.$$

The first estimate follows from the one-parameter result [26], see e.g. [10]. For the second inequality concerning square functions only see e.g. [1, Lemma 2.5].

Often when a supremum is taken over rectangles we also have a formulation of the norm uniformly each parameter separately. We have

$$\|b\|_{\mathrm{bmo}(\nu)} \sim \max\Big(\operatorname*{ess\,sup}_{x_1\in\mathbb{R}^{d_1}} \|b(x_1,\cdot)\|_{\mathrm{BMO}(\nu(x_1,\cdot))}, \operatorname*{ess\,sup}_{x_2\in\mathbb{R}^{d_2}} \|b(\cdot,x_2)\|_{\mathrm{BMO}(\nu(\cdot,x_2))}\Big) \tag{3.1}$$

where $\|\cdot\|\,\mathrm{BMO}(\rho)$ is the standard one-parameter dyadic weighted BMO space. For proof see e.g. [10].

We formulate the Muckenhoupt–Wheeden type estimates now.

3.2. **Lemma.** *Let $a \in \mathrm{BMO}$ and $w \in A_\infty$. It holds*

$$\sum_{I\in\mathcal{D}} \langle a, h_I\rangle \langle w\rangle_I \varphi_I \lesssim \|a\|_{\mathrm{BMO}} \Big\| \Big(\sum_I \varphi_I^2 \frac{1_I}{|I|}\Big)^{\frac12}\Big\|_{L^1(w)}.$$

In particular the above one is a special case of the two-weight version. We state this as a little bmo version.

3.3. **Lemma.** *Let $\sigma,\nu \in A_\infty$. Assume that $b \in \mathrm{bmo}(\nu)$. Then we have*

$$\sum_{R=R^1\times R^2} \langle b, h_R\rangle \langle \sigma\rangle_R \varphi_R \lesssim \|b\|_{\mathrm{bmo}(\nu)} \Big\| \Big(\sum_R \varphi_R^2 \frac{1_R}{|R|}\Big)^{\frac12}\Big\|_{L^1(\sigma\nu)}.$$

Also, we have

$$\sum_{R=R^1\times R^2} \Big\langle b, h_{R^1}\otimes \frac{1_{R^2}}{|R^2|}\Big\rangle \langle \sigma\rangle_R \varphi_R \lesssim \|b\|_{\mathrm{bmo}(\nu)} \Big\| \sum_{R^2} \Big(\sum_{R^1} \varphi_R^2 \frac{1_{R^1}}{|R^1|}\Big)^{\frac12} \otimes \frac{1_{R^2}}{|R^2|}\Big\|_{L^1(\sigma\nu)}$$

with a similar estimate when the cancellation is on the second parameter.

Proof. Let us consider the first estimate above and use the duality

$$\sum_{R=R^1\times R^2} \langle b, h_R\rangle \langle \sigma\rangle_R \varphi_R \lesssim \|b\|_{\mathrm{bmo}(\nu)} \int \Big(\sum_R \varphi_R^2 \langle \sigma\rangle_R^2 \frac{1_R}{|R|}\Big)^{\frac12} \nu.$$

By the reverse Hölder property of A_∞ weights, Lemma 2.2, we have

$$\langle \sigma\rangle_R \langle \nu\rangle_R \lesssim \langle \sigma\nu\rangle_R.$$

Hence, for all $R\in\mathcal{D}$ we have

$$\int \varphi_R \langle\sigma\rangle_R \frac{1_R}{|R|}\nu \lesssim \int \varphi_R \frac{1_R}{|R|}\sigma\nu.$$

The second part of the extrapolation result, Lemma 2.7, yields that

$$\int \Big(\sum_R \varphi_R^2 \langle \sigma\rangle_R^2 \frac{1_R}{|R|}\Big)^{\frac12} \nu \lesssim \int \Big(\sum_R \varphi_R^2 \frac{1_R}{|R|}\Big)^{\frac12} \sigma\nu$$

as desired.

For the second claim observe that, for example, we have

$$\sum_{R=R^1\times R^2} \Big\langle b, h_{R^1}\otimes \frac{1_{R^2}}{|R^2|}\Big\rangle \langle \sigma\rangle_R \varphi_R = \int_{\mathbb{R}^{d_2}} \sum_{R^2}\sum_{R^1} \langle b, h_{R^1}\rangle \langle \sigma\rangle_R \varphi_R \frac{1_{R^2}}{|R^2|}$$

$$\lesssim \|b\|_{\mathrm{bmo}(\nu)} \int_{\mathbb{R}^d} \sum_{R^2} \Big(\sum_{R^1} \varphi_R^2 \langle \sigma \rangle_R^2 \frac{1_{R^1}}{|R^1|} \Big)^{\frac{1}{2}} \otimes \frac{1_{R^2}}{|R^2|} \nu,$$

where we use the one-parameter duality for fixed variable on the second parameter. The proof is concluded as above. □

Using characterizations (3.1) and (2.1), we have

3.4. **Lemma.** *Let $\sigma, \nu \in A_\infty$. Assume that $b \in \mathrm{bmo}(\nu)$. For a fixed variable $x_1 \in \mathbb{R}^{d_1}$, we have*

$$\sum_{R^2} \langle b_{x_1}, h_{R^2} \rangle \langle \sigma_{x_1} \rangle_{R^2} \varphi_{R^2} \lesssim \|b\|_{\mathrm{bmo}(\nu)} \Big\| \Big(\sum_{R^2} \varphi_{R^2}^2 \frac{1_{R^2}}{|R^2|} \Big)^{\frac{1}{2}} \Big\|_{L^1_{x_2}(\sigma_{x_1} \nu_{x_1})},$$

where g_{x_1} denotes the one parameter function $g(x_1, \cdot)$. We have a similar estimate for a fixed variable on $\mathbb{R}^{d_2}$.

We omit the proof as it is analogous to the previous one.

4. Multilinear bi-parameter singular integrals

We call a function ω as a modulus of continuity if it is an increasing and subadditive function with $\omega(0) = 0$. A relevant quantity is the modified Dini condition

$$\|\omega\|_{\mathrm{Dini}_\alpha} := \int_0^1 \omega(t) \Big(1 + \log \frac{1}{t} \Big)^\alpha \frac{dt}{t}, \qquad \alpha \ge 0$$

that appears in practise as follows

$$\sum_{k=1}^\infty \omega(2^{-k}) k^\alpha = \sum_{k=1}^\infty \frac{1}{\log 2} \int_{2^{-k}}^{2^{-k+1}} \omega(2^{-k}) k^\alpha \frac{\mathrm{d}t}{t} \lesssim \int_0^1 \omega(t) \Big(1 + \log \frac{1}{t} \Big)^\alpha \frac{dt}{t}.$$

4.A. **Bi-parameter SIOs.** We consider an n-linear operator T on $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Let ω_i be a modulus of continuity on $\mathbb{R}^{d_i}$. We define that T is an n-linear bi-parameter (ω_1, ω_2)-SIO if it satisfies the full and partial kernel representations as defined below.

Full kernel representation. Let $f_i = f_i^1 \otimes f_i^2$, $i = 1, \dots, n+1$. For both $m \in \{1, 2\}$ there exists $i_1, i_2 \in \{1, \dots, n+1\}$ so that $\operatorname{spt} f_{i_1}^m \cap \operatorname{spt} f_{i_2}^m = \emptyset$. We demand that in this case we have the representation

$$\langle T(f_1, \dots, f_n), f_{n+1} \rangle = \int_{\mathbb{R}^{(n+1)d}} K(x_{n+1}, x_1, \dots, x_n) \prod_{i=1}^{n+1} f_i(x_i) \,\mathrm{d}x,$$

where

$$K \colon \mathbb{R}^{(n+1)d} \setminus \{(x_{n+1}, x_1, \dots, x_n) \in \mathbb{R}^{(n+1)d} \colon x_1^1 = \dots = x_{n+1}^1 \text{ or } x_1^2 = \dots = x_{n+1}^2\} \to \mathbb{C}$$

is a kernel satisfying a set of estimates which we specify next. The kernel K is assumed to satisfy the size estimate

$$|K(x_{n+1}, x_1, \dots, x_n)| \lesssim \prod_{m=1}^2 \frac{1}{\Big(\sum_{i=1}^n |x_{n+1}^m - x_i^m| \Big)^{d_m n}}.$$

In addition, we require the continuity estimate, for example, we demand that

$$|K(x_{n+1}, x_1, \dots, x_n) - K(x_{n+1}, x_1, \dots, x_{n-1}, (c^1, x_n^2))$$

$$
\begin{aligned}
&- K((x_{n+1}^1, c^2), x_1, \dots, x_n) + K((x_{n+1}^1, c^2), x_1, \dots, x_{n-1}, (c^1, x_n^2))| \\
&\lesssim \omega_1\Big(\frac{|x_n^1 - c^1|}{\sum_{i=1}^n |x_{n+1}^1 - x_i^1|}\Big) \frac{1}{\Big(\sum_{i=1}^n |x_{n+1}^1 - x_i^1|\Big)^{d_1 n}} \\
&\times \omega_2\Big(\frac{|x_{n+1}^2 - c^2|}{\sum_{i=1}^n |x_{n+1}^2 - x_i^2|}\Big) \frac{1}{\Big(\sum_{i=1}^n |x_{n+1}^2 - x_i^2|\Big)^{d_2 n}}
\end{aligned}
$$

whenever $|x_n^1 - c^1| \le 2^{-1} \max_{1\le i\le n} |x_{n+1}^1 - x_i^1|$ and $|x_{n+1}^2 - c^2| \le 2^{-1} \max_{1\le i \le n} |x_{n+1}^2 - x_i^2|$. Of course, we also require all the other natural symmetric estimates, where c^1 can be in any of the given $n+1$ slots and similarly for c^2. There are, of course, $(n+1)^2$ different estimates.

Moreover, we expect to have the following mixed continuity and size estimates. For example, we demand that

$$
\begin{aligned}
&|K(x_{n+1}, x_1, \dots, x_n) - K(x_{n+1}, x_1, \dots, x_{n-1}, (c^1, x_n^2))| \\
&\lesssim \omega_1\Big(\frac{|x_n^1 - c^1|}{\sum_{i=1}^n |x_{n+1}^1 - x_i^1|}\Big) \frac{1}{\Big(\sum_{i=1}^n |x_{n+1}^1 - x_i^1|\Big)^{d_1 n}} \cdot \frac{1}{\Big(\sum_{i=1}^n |x_{n+1}^2 - x_i^2|\Big)^{d_2 n}}
\end{aligned}
$$

whenever $|x_n^1 - c^1| \le 2^{-1} \max_{1\le i\le n} |x_{n+1}^1 - x_i^1|$. Again, we also require all the other natural symmetric estimates.

Partial kernel representations. Suppose now only that there exists $i_1, i_2 \in \{1, \dots, n+1\}$ so that $\operatorname{spt} f_{i_1}^1 \cap \operatorname{spt} f_{i_2}^1 = \emptyset$. Then we assume that

$$
\langle T(f_1, \dots, f_n), f_{n+1}\rangle = \int_{\mathbb{R}^{(n+1)d_1}} K_{(f_i^2)}(x_{n+1}^1, x_1^1, \dots, x_n^1) \prod_{i=1}^{n+1} f_i^1(x_i^1)\,\mathrm{d}x^1,
$$

where $K_{(f_i^2)}$ is a one-parameter ω_1-Calderón–Zygmund kernel with a constant depending on the fixed functions $f_1^2, \dots, f_{n+1}^2$. For example, this means that the size estimate takes the form

$$
|K_{(f_i^2)}(x_{n+1}^1, x_1^1, \dots, x_n^1)| \le C(f_1^2, \dots, f_{n+1}^2) \frac{1}{\Big(\sum_{i=1}^n |x_{n+1}^1 - x_i^1|\Big)^{d_1 n}}.
$$

The continuity estimates are analogous.

We assume the following $T1$ type control on the constant $C(f_1^2, \dots, f_{n+1}^2)$. We have

$$
C(1_{I^2}, \dots, 1_{I^2}) \lesssim |I^2| \tag{4.1}
$$

and

$$
C(a_{I^2}, 1_{I^2}, \dots, 1_{I^2}) + C(1_{I^2}, a_{I^2}, 1_{I^2}, \dots, 1_{I^2}) + \cdots + C(1_{I^2}, \dots, 1_{I^2}, a_{I^2}) \lesssim |I^2|
$$

for all cubes $I^2 \subset \mathbb{R}^{d_2}$ and all functions a_{I^2} satisfying $a_{I^2} = 1_{I^2} a_{I^2}$, $|a_{I^2}| \le 1$ and $\int a_{I^2} = 0$.

Analogous partial kernel representation on the second parameter is assumed when $\operatorname{spt} f_{i_1}^2 \cap \operatorname{spt} f_{i_2}^2 = \emptyset$ for some i_1, i_2.

4.B. **Multilinear bi-parameter Calderón-Zygmund operators.** We say that T satisfies the weak boundedness property if

$$|\langle T(1_R, \dots, 1_R), 1_R \rangle| \lesssim |R| \tag{4.2}$$

for all rectangles $R = I^1 \times I^2 \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

An SIO T satisfies the diagonal BMO assumption if the following holds. For all rectangles $R = I^1 \times I^2 \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and functions a_{I^i} with $a_{I^i} = 1_{I^i} a_{I^i}$, $|a_{I^i}| \le 1$ and $\int a_{I^i} = 0$ we have

$$|\langle T(a_{I^1} \otimes 1_{I^2}, 1_R, \dots, 1_R), 1_R \rangle| + \dots + |\langle T(1_R, \dots, 1_R), a_{I^1} \otimes 1_{I^2} \rangle| \lesssim |R| \tag{4.3}$$

and

$$|\langle T(1_{I^1} \otimes a_{I^2}, 1_R, \dots, 1_R), 1_R \rangle| + \dots + |\langle T(1_R, \dots, 1_R), 1_{I^1} \otimes a_{I^2} \rangle| \lesssim |R|.$$

An SIO T satisfies the product BMO assumption if it holds

$$S(1, \cdots, 1) \in \mathrm{BMO}_{\text{prod}}$$

for all the $(n+1)^2$ adjoints $S = T^{j_1*, j_2*}_{1,2}$. This can be interpreted in the sense that

$$\|S(1, \cdots, 1)\|_{\mathrm{BMO}_{\text{prod}}} = \sup_{\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2} \sup_{\Omega} \Big(\frac{1}{|\Omega|} \sum_{\substack{R = I^1 \times I^2 \in \mathcal{D} \\ R \subset \Omega}} |\langle S(1, \cdots, 1), h_R \rangle|^2 \Big)^{1/2} < \infty,$$

where $h_R = h_{I^1} \otimes h_{I^2}$ and the supremum is over all dyadic grids $\mathcal{D}^i$ on $\mathbb{R}^{d_i}$ and open sets $\Omega \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $0 < |\Omega| < \infty$, and the pairings $\langle S(1, \cdots, 1), h_R \rangle$ can be defined, in a natural way, using the kernel representations.

4.4. **Definition.** An n-linear bi-parameter (ω_1, ω_2)-SIO T satisfying the weak boundedness property, the diagonal BMO assumption and the product BMO assumption is called an n-linear bi-parameter (ω_1, ω_2)-Calderón–Zygmund operator ((ω_1, ω_2)-CZO).

We simplify the study of above operators through the following representation theorem.

4.5. **Proposition.** *Suppose T is an n-linear bi-parameter (ω_1, ω_2)-CZO. Then we have*

$$\langle T(f_1, \dots, f_n), f_{n+1} \rangle = C_T \mathbb{E}_\sigma \sum_{u = (u_1, u_2) \in \mathbb{N}^2} \omega_1(2^{-u_1}) \omega_2(2^{-u_2}) \langle U_{u,\sigma}(f_1, \dots, f_n), f_{n+1} \rangle,$$

where C_T enjoys a linear bound with respect to the CZO quantities and $U_{u,\sigma}$ denotes some n-linear bi-parameter dyadic operator (defined in the grid $\mathcal{D}_\sigma$) with the following property. We have that $U_u = U_{u,\sigma}$ can be decomposed using the standard dyadic model operators as follows:

$$U_u = C \sum_{i_1=0}^{u_1-1} \sum_{i_2=0}^{u_2-1} V_{i_1, i_2}, \tag{4.6}$$

where each $V = V_{i_1,i_2}$ is a dyadic model operator (a shift, a partial paraproduct or a full paraproduct) of complexity $k^m_{j,V}$, $j \in \{1, \dots, n+1\}$, $m \in \{1, 2\}$, satisfying

$$k^m_{j,V} \le u_m.$$

In above $\mathbb{E}_\sigma$ denotes the expectation over a natural probability space $\Omega = \Omega_1 \times \Omega_2$, the details of which are not relevant for us here, so that to each $\sigma = (\sigma_1, \sigma_2) \in \Omega$ we can associate a random collection of dyadic rectangles $\mathcal{D}_\sigma = \mathcal{D}_{\sigma_1} \times \mathcal{D}_{\sigma_2}$. The proposition is a consequence of [3, Theorem 5.35. and Lemma 5.12.].

It was proven in [3] that the minimal regularity we require is that $\omega_i \in \operatorname{Dini}_{\frac{1}{2}}$. For the optimal dependence the dyadic representation is in terms of certain modified model operators. The modified versions of the standard operators are much more difficult to handle and we are forced to rely on the lemma that these can be written as a sum of the standard ones. However, as it is explained in [3], this will cause a loss in the kernel regularity. Yet another problem appears when dealing with the genuinely multilinear weights. Thus in some cases, we need to stick to the usual Hölder type kernel regularity $\omega_i(t) = t^{\alpha_i}$. In the paper [21], it was proven that the standard model operators are bounded with the weights on the genuinely multilinear weight class introduced earlier. We will move on to introducing the model operators and state the very recent results for these.

4.C. **Dyadic model operators.** All the operators in this section are defined in some fixed rectangles $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$. We do not emphasise this dependence in the notation.

4.D. **Shifts.** Let $k = (k_1, \ldots, k_{n+1})$, where $k_i = (k_i^1, k_i^2) \in \{0, 1, \ldots\}^2$. An n-linear bi-parameter shift S_k takes the form

$$\langle S_k(f_1, \ldots, f_n), f_{n+1}\rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_i^{(k_i)} = K}} a_{K,(R_i)} \prod_{i=1}^{n+1} \langle f_i, \widetilde{h}_{R_i}\rangle.$$

Here $K, R_1, \ldots, R_{n+1} \in \mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, $R_i = I_i^1 \times I_i^2$, $R_i^{(k_i)} := (I_i^1)^{(k_i^1)} \times (I_i^2)^{(k_i^2)}$ and $\widetilde{h}_{R_i} = \widetilde{h}_{I_i^1} \otimes \widetilde{h}_{I_i^2}$. Here we assume that for $m \in \{1, 2\}$ there exist two indices $i_0^m, i_1^m \in \{1, \ldots, n+1\}$, $i_0^m \ne i_1^m$, so that $\widetilde{h}_{I^m_{i_0^m}} = h_{I^m_{i_0^m}}$, $\widetilde{h}_{I^m_{i_1^m}} = h_{I^m_{i_1^m}}$ and for the remaining indices $i \notin \{i_0^m, i_1^m\}$ we have $\widetilde{h}_{I_i^m} \in \{h^0_{I_i^m}, h_{I_i^m}\}$. Moreover, $a_{K,(R_i)} = a_{K,R_1,\ldots,R_{n+1}}$ is a scalar satisfying the normalization

$$|a_{K,(R_i)}| \le \frac{\prod_{i=1}^{n+1} |R_i|^{1/2}}{|K|^n}. \tag{4.7}$$

4.8. **Theorem** ([21, Theorem 6.2.]). *Suppose S_k is an n-linear bi-parameter shift, $1 < p_1, \ldots, p_n, \le \infty$ and $\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i} > 0$. Then we have*

$$\|S_k(f_1, \ldots, f_n)w\|_{L^p} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}$$

for all multilinear bi-parameter weights $\vec{w} \in A_{\vec{p}}$. The implicit constant does not depend on k.

4.E. **Partial paraproducts.** Let $k = (k_1, \ldots, k_{n+1})$, where $k_i \in \{0, 1, \ldots\}$. An n-linear bi-parameter partial paraproduct $(S\pi)_k$ with the paraproduct component on $\mathbb{R}^{d_2}$ takes the form

$$\langle (S\pi)_k(f_1,\ldots,f_n), f_{n+1}\rangle = \sum_{K=K^1\times K^2} \sum_{\substack{I^1_1,\ldots,I^1_{n+1} \\ (I^1_i)^{(k_i)}=K^1}} a_{K,(I^1_i)} \prod_{i=1}^{n+1} \langle f_i, \widetilde{h}_{I^1_i} \otimes u_{i,K^2}\rangle, \tag{4.9}$$

where the functions $\widetilde{h}_{I^1_i}$ and u_{i,K^2} satisfy the following. There are $i_0, i_1 \in \{1,\ldots,n+1\}$, $i_0 \ne i_1$, so that $\widetilde{h}_{I^1_{i_0}} = h_{I^1_{i_0}}$, $\widetilde{h}_{I^1_{i_1}} = h_{I^1_{i_1}}$ and for the remaining indices $i \notin \{i_0, i_1\}$ we have $\widetilde{h}_{I^1_i} \in \{h^0_{I^1_i}, h_{I^1_i}\}$. There is $i_2 \in \{1,\ldots,n+1\}$ so that $u_{i_2,K^2} = h_{K^2}$ and for the remaining indices $i \ne i_2$ we have $u_{i,K^2} = \frac{1_{K^2}}{|K^2|}$. Moreover, the coefficients are assumed to satisfy

$$\|(a_{K,(I^1_i)})_{K^2}\|_{\mathrm{BMO}} = \sup_{K^2_0\in\mathcal{D}^2} \Big(\frac{1}{|K^2_0|}\sum_{K^2\subset K^2_0} |a_{K,(I^1_i)}|^2\Big)^{1/2} \le \frac{\prod_{i=1}^{n+1}|I^1_i|^{\frac12}}{|K^1|^n}. \tag{4.10}$$

Of course, $(\pi S)_k$ is defined symmetrically.

4.11. **Theorem** ([21, Theorem 6.7.]). *Suppose $(S\pi)_k$ is an n-linear partial paraproduct, $1 < p_1,\ldots,p_n \le \infty$ and $\frac1p = \sum_{i=1}^n \frac{1}{p_i} > 0$. Then, for every $0<\beta\le 1$ we have*

$$\|(S\pi)_k(f_1,\ldots,f_n)w\|_{L^p} \lesssim_\beta 2^{\max_j k_j\beta}\prod_{i=1}^n \|f_iw_i\|_{L^{p_i}}$$

for all multilinear bi-parameter weights $\vec{w} \in A_{\vec{p}}$.

4.F. **Full paraproducts.** An n-linear bi-parameter full paraproduct Π takes the form

$$\langle \Pi(f_1,\ldots,f_n), f_{n+1}\rangle = \sum_{K=K^1\times K^2} a_K \prod_{i=1}^{n+1}\langle f_i, u_{i,K^1}\otimes u_{i,K^2}\rangle, \tag{4.12}$$

where the functions u_{i,K^1} and u_{i,K^2} are like in (4.9). The coefficients are assumed to satisfy

$$\|(a_K)\|_{\mathrm{BMO}_{\mathrm{prod}}} = \sup_{\Omega}\Big(\frac{1}{|\Omega|}\sum_{K\subset\Omega}|a_K|^2\Big)^{1/2} \le 1,$$

where the supremum is over open sets $\Omega \subset \mathbb{R}^d = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$ with $0<|\Omega|<\infty$.

4.13. **Theorem** ([21, Theorem 6.21.]). *Suppose Π is an n-linear bi-parameter full paraproduct, $1 < p_1,\ldots,p_n \le \infty$ and $1/p = \sum_{i=1}^n 1/p_i > 0$. Then we have*

$$\|\Pi(f_1,\ldots,f_n)w\|_{L^p} \lesssim \prod_{i=1}^n \|f_iw_i\|_{L^{p_i}}$$

for all multilinear bi-parameter weights $\vec{w} \in A_{\vec{p}}$.

In fact, the above theorem is a special case of the Bloom type inequality. The following operator and result have obvious extensions in the product BMO setting. We consider an n-linear bi-parameter paraproduct

$$\langle \Pi_b(f_1,\ldots,f_n), f_{n+1}\rangle = \sum_{K=K^1\times K^2}\langle b, v_{0,K^1}\otimes v_{0,K^2}\rangle \prod_{i=1}^{n+1}\langle f_i, v_{i,K^1}\otimes v_{i,K^2}\rangle. \tag{4.14}$$

Here we assume that for $m \in \{1,2\}$ there exist two indices $i_0^m, i_1^m \in \{0,\dots,n+1\}$, $i_0^m \neq i_1^m$, so that $v_{i_0^m,K^m} = h_{K^m}$, $v_{i_1^m,K^m} = h_{K^m}$ and for the remaining indices $i \notin \{i_0^m, i_1^m\}$ we have $v_{i,K^m} = \frac{1_{K^m}}{|K^m|}$. Moreover, here we will assume that we at least have $0 \in \{i_0^1, i_1^1\}$ or $0 \in \{i_0^2, i_1^2\}$.

Later on, paraproducts will also appear as a result of standard expansions of products

$$bf = \sum_{I^i \in \mathcal{D}^i} \langle b, h_{I^i}\rangle_i \langle f, h_{I^i}\rangle_i \otimes h_{I^i} h_{I^i} + \sum_{I^i \in \mathcal{D}^i} \langle b, h_{I^i}\rangle_i \langle f\rangle_{I^i,i} \otimes h_{I^i} + \sum_{I^i \in \mathcal{D}^i} \langle b\rangle_{I^i,i} \langle f, h_{I^i}\rangle_i \otimes h_{I^i}.$$

In the first term, the worst case is if $h_{I^i} h_{I^i}$ is non-cancellative hence equals to $1_{I^i}/|I^i|$. Often it is enough to consider the worst-case scenario.

We denote these expansions as $\Pi_{j_1,j_2}(b,f), (j_1,j_2) \in \{1,2,3\}^2$, where the indices dictates the from of the paraproduct. More specifically, in the above language of the multilinear paraproduct: if $j_m = 1$ then $i_0^m = 0$ and $i_1^m = 1$, if $j_m = 2$ then $i_0^m = 0$ and $i_1^m = 2$, and if $j_m = 3$ then $i_0^m = 1$ and $i_1^m = 2$. In all of the cases the unmentioned slot do not have the cancellation. Hence, notice that when $j_1 = 3 = j_2$ we have no cancellation for the function b meaning that it is not a paraproduct as such.

4.15. **Proposition.** *Let Π_b be a paraproduct as described above. Fix $\vec{p} = (p_1,\dots,p_n)$ so that $1 < p_i \le \infty$, define $\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i}$ and assume $1 < p < \infty$. Let $(w_1,\dots,w_n,\nu)$ be a tuple of weights. Assume that*

$$b \in \operatorname{bmo}(\nu) \qquad \text{and} \qquad (w_1,\dots,w_n,\nu w^{-1}) \in A^*_{\vec{p}}.$$

Then we have

$$\|\Pi_b(f_1,\dots,f_n)\nu^{-1}w\|_{L^p} \lesssim \|b\|_{\operatorname{bmo}(\nu)} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}. \tag{4.16}$$

Moreover, if $\lambda_j = w_j \nu^{-1}$ for some $j = 1,2,\dots n$ such that

$$(w_1,\dots,w_{j-1},\lambda_j,w_{j+1},\dots,w_n), (w_1,\dots,w_n) \in A_{\vec{p}},$$

then (4.16) holds for all $1 < p_i \le \infty$ such that $p \in (n^{-1},\infty)$.

Proof. It suffices to show that

$$|\langle \Pi_b(f_1,\dots,f_n), f_{n+1}\rangle| \lesssim \|b\|_{\operatorname{bmo}(\nu)} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}} \cdot \|f_{n+1}\nu w^{-1}\|_{L^{p'}}.$$

Case I. We have $0 \in \{i_0^1, i_1^1\}$ and $0 \in \{i_0^2, i_1^2\}$. We consider the concrete case

$$\begin{aligned}
&\langle \Pi_b(f_1,\dots,f_n), f_{n+1}\rangle \\
&= \sum_{K = K^1 \times K^2} \langle b, h_{K^1} \otimes h_{K^2}\rangle \Big\langle f_1, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|}\Big\rangle \Big\langle f_2, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2}\Big\rangle \prod_{i=3}^{n+1} \langle f_i\rangle_K.
\end{aligned}$$

We have

$$\begin{aligned}
\langle \Pi_b(f_1,\dots,f_n), f_{n+1}\rangle &\lesssim \|b\|_{\operatorname{bmo}(\nu)} \Big\| \Big(\sum_{K \in \mathcal{D}} \langle |\Delta^1_{K^1} f_1|\rangle_K^2 \langle |\Delta^2_{K^2} f_2|\rangle_K^2 \prod_{i=3}^{n+1} \langle |f_i|\rangle_K^2 1_K \Big)^{\frac{1}{2}} \Big\|_{L^1(\nu)} \\
&\lesssim \|b\|_{\operatorname{bmo}(\nu)} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}} \cdot \|f_{n+1}\nu w^{-1}\|_{L^{p'}}.
\end{aligned}$$

Here the first step used that $\nu \in A_\infty$ – which follows as $(w_1, \ldots, w_n, \nu w^{-1}) \in A^*_{\vec{p}}$ – and the estimate

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

The second step used Theorem 2.13 together with the assumption $(w_1, \ldots, w_n, \nu w^{-1}) \in A^*_{\vec{p}}$.

Case 2. We have $0 \notin \{i^1_0, i^1_1\}$ but $0 \in \{i^2_0, i^2_1\}$ (or the other way around). We consider the concrete case

$$\begin{aligned}
&\langle \Pi_b(f_1, \ldots, f_n), f_{n+1}\rangle \\
&= \sum_K \Big\langle b, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2}\Big\rangle \Big\langle f_1, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2}\Big\rangle \Big\langle f_2, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|}\Big\rangle \Big\langle f_3, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|}\Big\rangle \prod_{i=4}^{n+1} \langle f_i\rangle_K.
\end{aligned}$$

We have

$$\begin{aligned}
&\langle A(f_1, \ldots, f_n), f_{n+1}\rangle \\
&\lesssim \|b\|_{\mathrm{bmo}(\nu)} \Big\| \Big(\sum_{K^2} \Big(\sum_{K^1} \langle |\Delta^2_{K^2} f_1|\rangle_K \langle |\Delta^1_{K^1} f_2|\rangle_K \langle |\Delta^1_{K^1} f_3|\rangle_K \prod_{i=4}^{n} (\langle |f_i|\rangle_K 1_K \Big)^2 \Big)^{\frac12} \Big\|_{L^1(\nu)} \\
&\lesssim \|b\|_{\mathrm{bmo}(\nu)} \prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}} \cdot \|f_{n+1} \nu w^{-1}\|_{L^{p'}},
\end{aligned}$$

where we used the estimate

$$|\langle b, f\rangle| \lesssim \|b\|_{\mathrm{bmo}(\nu)} \|S^2_{\mathcal{D}^2} f\|_{L^1(\nu)}$$

and Theorem 2.13.

The second claim is obtained by using extrapolation, Theorem 1.2. □

Let λ and w be bi-parameter weights such that for some $1 < p < \infty$ we have $\lambda^{-p'}$, $w^{-p'} \in A_\infty$. Assume also that $\nu := \lambda^{-1} w \in A_\infty$ and $b \in \mathrm{bmo}(\nu)$. Then we have a weighted variant of the paraproduct operator

$$\langle \Pi_{b,\eta} f_1, f_2\rangle = \sum_{K = K^1 \times K^2} \langle b, h_K\rangle \langle f_1, h_K\rangle \langle f_2\rangle^\eta_K, \tag{4.17}$$

where $\eta = \lambda^{-p'}$.

4.18. **Proposition.** *let $p \in (1, \infty)$. Let λ and w be bi-parameter weights such that $\lambda^{-p'}$, $w^{-p'} \in A_\infty$. Let $\Pi_{b,\eta}$ be a weighted paraproduct operator defined via (4.17), we have*

$$\|\Pi_{b,\eta}(f)\lambda\|_{L^p} \lesssim \|b\|_{\mathrm{bmo}(\nu)} \|fw\|_{L^p}.$$

Proof. The result follows from a variant of techniques seen in the proof of Proposition 4.15. For example, by duality we have terms like (4.17). Introducing a weight averages $\langle \sigma\rangle_K \langle \sigma\rangle^{-1}_K = 1$, where $\sigma = w^{-p'}$, we can apply Lemma 3.3. Hence, we get

$$\begin{aligned}
|\langle \Pi_{b,\eta} f_1, f_2\rangle| &\lesssim \|b\|_{\mathrm{bmo}(\nu)} \int \Big(\sum_K \frac{\langle f_1, h_K\rangle^2}{\langle \sigma\rangle^2_K} (\langle f_2\rangle^\eta_K)^2 \frac{1_K}{|K|} \Big)^{\frac12} \sigma\nu \\
&\le \|b\|_{\mathrm{bmo}(\nu)} \int M^\eta_{\mathcal{D}} f_2 \Big(\sum_K \frac{\langle f_1, h_K\rangle^2}{\langle \sigma\rangle^2_K} \frac{1_K}{|K|} \Big)^2 \sigma\nu
\end{aligned}$$

$$\le \|b\|_{\mathrm{bmo}(\nu)} \|M_{\mathcal{D}}^{\eta} f_2\|_{L^{p'}(\eta)} \Big\| \Big(\sum_K \frac{\langle f_1, h_K\rangle^2}{\langle \sigma \rangle_K^2} \frac{1_K}{|K|} \Big)^2 \sigma^{\frac{1}{p}} \Big\|_{L^p}$$
$$\lesssim \|b\|_{\mathrm{bmo}(\nu)} \|f_2 \lambda^{-1}\|_{L^{p'}} \|f_1 w\|_{L^p}.$$

□

In the same setting as above we can have, for example, the following mixed type weighted paraproduct

$$\langle \Pi_{b,\eta} f_1, f_2\rangle = \sum_{K = K^1 \times K^2} \Big\langle b, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \langle f_1, h_K\rangle \langle \langle f_2, h_{K^2}\rangle_2 \rangle_{K^1}^{\langle \eta \rangle_{K^2,2}}.$$

Symmetrical definition when we have $\Big\langle b, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2} \Big\rangle$. We also consider the case

$$\langle \Pi_{b,\eta} f_1, f_2\rangle = \sum_{K = K^1 \times K^2} \Big\langle b, h_{K^1} \otimes h_{K^2} \Big\rangle \Big\langle f_1, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \langle \langle f_2, h_{K^2}\rangle_2 \rangle_{K^1}^{\langle \eta \rangle_{K^2,2}}.$$

4.19. **Proposition.** *For a weighted paraproduct operator $\Pi_{b,\eta}$ as described above, we have*

$$|\langle \Pi_{b,\eta} f_1, f_2\rangle| \lesssim \|b\|_{\mathrm{bmo}(\nu)} \|f w\|_{L^p} \|S_{\mathcal{D}}^i f_2 \lambda^{-1}\|_{L^{p'}},$$

where i is either 1 or 2 depending on which parameter the cancellation is.

Proof. Let us, for example, consider the paraproduct written above, where we have $\Big\langle b, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle$. Similar to the previous proof, we use Lemma 3.3 but this time the second claim. Then the main difference to the previous proof is that we face e.g.

$$\Big\| \Big(\sum_{K^2} M_{\mathcal{D}^1}^{\langle \eta \rangle_{K^2,2}} (\langle f, h_{K^2}\rangle_2)^2 \otimes \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \Big\|_{L^{p'}(\eta)}.$$

Nevertheless, the claim follows quite easily via an extrapolation trick (see [21, Lemma 9.2]), since for fixed $p' = 2$ we have

$$\int_{\mathbb{R}^{d_1}} \Big[M_{\mathcal{D}^1}^{\langle \eta \rangle_{K^2,2}} (\langle f, h_{K^2}\rangle_2) \Big]^2 \langle \eta \rangle_{K^2,2} \lesssim [\eta]_{A_\infty} \int_{\mathbb{R}^{d_1}} \langle f, h_{K^2}\rangle_2^2 \langle \eta \rangle_{K^2,2}.$$

□

For the references below, we state a lemma regarding the square functions of partial paraproducts. For the lemma, it is relevant in which slots the cancellation appears. The square function can be taken corresponding to the cancellation on the $(n+1)$-th slot. For example, if $(S\pi)_k$ is a form of partial paraproduct such that there is a cancellation on the $(n+1)$-th slot on the second parameter, then we have the boundedness of the second parameter square function of this operator, namely $S_{\mathcal{D}^2}(S\pi)_k$. Similarly, $S_{\mathcal{D}^1}(S\pi)_k$ and $S_{\mathcal{D}}(S\pi)_k$ must have the corresponding cancellation to be bounded.

4.20. **Lemma.** *Let U be a square function of partial paraproduct stated in above. Let $1 < p_i \le \infty$ and $\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i} > 0$. It holds*

$$\Big\| U(f_1, \ldots, f_n) w \Big\|_{L^p} \lesssim_\beta 2^{\max_j k_j \beta} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}},$$

where $w = \prod_{i=1}^n w_i$, $(w_1, \ldots, w_n) \in A_{(p_1, \ldots, p_n)}$.

Proof. The result follows almost identically to the proof of [21, Theorem 6.7.]. We take the partial paraproduct of the form

$$\Big(\sum_{K\in\mathcal{D}}\Big(\sum_{(I_i^1)^{(k_i)}=K^1} a_{K,(I_i^1)}\prod_{i=1}^{n}\Big\langle f_i,\widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle h^0_{I^1_{n+1}}\otimes h^0_{K^2}\Big)^2\Big)^{\frac12}.$$

Using the dualisation trick in [21] for $p>1$, we choose a sequence of functions $(f_{n+1,K})_K\in L^{p'}(\ell^2)$ with norm $\|(f_{n+1,K})_K\|_{L^{p_1}(\ell^2)}\le 1$, and we look at

$$\begin{aligned}&\Big|\sum_{K\in\mathcal{D}}\sum_{(I_i^1)^{(k_i)}=K^1} a_{K,(I_i^1)}\prod_{i=1}^{n}\Big\langle f_i,\widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\langle f_{n+1,K}w,h^0_{I^1_{n+1}}\otimes h^0_{K^2}\rangle\Big|\\&\le\sum_{K^1}\sum_{(I_i^1)^{(k_i)}=K^1}\frac{\prod_{i=1}^{n+1}|I_i^1|^{\frac12}}{|K^1|^2}\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}\Big|A_{K^2}(\langle f_1,\widetilde{h}_{I_1^1}\rangle,\dots,\langle f_{n+1,K}w,h^0_{I^1_{n+1}}\rangle)\Big|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac12},\end{aligned}$$

where

$$A_{K^2}(g_1,\dots,g_{n+1})=\Big(\prod_{i=1}^{n}\langle g_i\rangle_{K^2}\Big)\langle g_{n+1},h^0_{K^2}\rangle.$$

We write

$$|I_i^1|^{-\frac12}\Big\langle f_i,h^0_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle=\langle f_i\rangle_K+\sum_{\ell_3=0}^{k_i-1}\sum_{(L_i^1)^{(\ell_i)}=K^1}\Big\langle f_i,h_{L_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\langle h_{L_i^1}\rangle_{I_i^1}$$

for $i\in\{1,2,\dots,n\}$ whenever we have the non-cancellative Haar function, expect when complexity is zero.

We are reduced to bounding

$$\begin{aligned}&\sum_{K^1}\sum_{(L_i^1)^{(\ell_i)}=K^1}\frac{\prod_{i=1}^{n}|L_i^1|^{\frac12}}{|K^1|^n}\\&\times\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}\Big|A_{K^2}(\langle f_1,\widetilde{h}_{L_1^1}\rangle,\dots,\langle f_n,\widetilde{h}_{L_n^1}\rangle,\langle f_{n+1,K}w,h^0_{L^1_{n+1}}\rangle)\Big|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac12},\end{aligned}\tag{4.21}$$

where $\widetilde{h}_{L_i^1}=h_{L_i^1}$ for at least one index i, and $\ell_{n+1}=k_{n+1}$. Moreover, if $\widetilde{h}_{L_i^1}=h^0_{L_i^1}$, then we have complexity $\ell_i=0$.

We consider an example to see how we can use the idea in [21] in this setting. The goal is to prove

$$\|g\|_{L^1}\le\Big(\prod_{i=1}^{n}\|f_iw_i\|_{L^{p_i}}\Big)\|\widetilde{f}_{n+1}w^{-1}\|_{L^{p'}},$$

where

$$\widetilde{f}_1:=\Big(\sum_K|f_{1,K}w|^2\Big)^{\frac12}$$

and g equals to

$$\sum_{K^1}\sum_{(L_i^1)^{(\ell_i)}=K^1}\frac{\prod_{i=1}^{n}|L_i^1|^{\frac12}}{|K^1|^n}\frac{1_{K^1}}{|K^1|}$$

$$\times \Big(\sum_{K^2} \Big| A_{K^2}(\langle f_1, \widetilde{h}_{L_2^1}\rangle, \dots, \langle f_n, \widetilde{h}_{L_n^1}\rangle, \langle f_{n+1,K} w, h^0_{L^1_{n+1}}\rangle)\Big|^2 \frac{1_{K^2}}{|K^2|}\Big)^{\frac12}.$$

By extrapolation [18], we just need to prove that

$$\|gv\|_{L^{\frac{2}{n+1}}} \le \prod_{i=1}^n \|f_i v_i\|_{L^2} \|\widetilde{f}_{n+1} v_{n+1}\|_{L^2}, \qquad (v_1, \dots, v_{n+1}) \in A_{(2,\dots,2)}.$$

Following the proof in [21], everything will be the same except that for $\tilde{f}_{n+1}$, we need to control

$$\Big\| \Big(\sum_{K^1} |F_{n+1,K^1}|^2\Big)^{\frac12} v_{n+1}^{-1}\Big\|_{L^2} = \Big\| \Big(\sum_{K^1} |F_{n+1,K^1}|^2\Big)^{\frac12} \gamma_{n+1}^{\frac12}\Big\|_{L^2},$$

where

$$F_{n+1,K^1} = 1_{K^1} \sum_{(I^1_{n+1})^{(k_{n+1})} = K^1} \frac{|I^1_{n+1}|^{\frac12}}{|K^1|} \Big(\sum_{K^2} \frac{\langle |f_{n+1,K}| w, h^0_{I^1_{n+1}} \otimes h^0_{K^2}\rangle^2}{\langle \gamma_{n+1}\rangle_K^2} \frac{1_{K^2}}{|K^2|}\Big)^{\frac12},$$

and $\gamma_{n+1} = v_{n+1}^{-2}$. For brevity, in below we just write $\sum_{I^1_{n+1}}$ instead of $\sum_{(I^1_{n+1})^{(k_{n+1})} = K^1}$.

So it remains to prove some variant of Proposition 2.14, which is straightforward. In fact, for the above model case, since $\gamma_{n+1} \in A_{2(n+1)}$, we have

$$(\gamma_{n+1}^{-\frac12}, \gamma_{n+1}^{\frac{1}{2n+1}}, \cdots, \gamma_{n+1}^{\frac{1}{2n+1}}) \in A_{(2,\infty,\cdots,\infty)}.$$

Thus,

$$F_{n+1,K^1} \le 1_{K^1} \sum_{I^1_{n+1}} \frac{|I^1_{n+1}|^{\frac12}}{|K^1|} \Big(\sum_{K^2} \langle |f_{n+1,K}| w, h^0_{I^1_{n+1}} \otimes h^0_{K^2}\rangle^2 \langle \gamma_{n+1}^{-\frac{1}{2n+1}}\rangle_K^{2\cdot(2n+1)} \frac{1_{K^2}}{|K^2|}\Big)^{\frac12}.$$

If $k_{n+1} = 0$, we simply have

$$F_{n+1,K^1} \le \Big(\sum_{K^2} \big[M_{\mathcal{D}}(|f_{n+1,K}| w, \gamma_{n+1}^{-\frac{1}{2n+1}}, \cdots, \gamma_{n+1}^{-\frac{1}{2n+1}})\big]^2\Big)^{\frac12}.$$

Then it is just a matter of vector-valued estimates for the multilinear maximal function and we are done. If $k_1 > 0$, then let $s > 1$ be such that d_1/s' is sufficiently small, we have

$$\begin{aligned}
&F_{n+1,K^1} \\
&\le 2^{\frac{k_{n+1} d_1}{s'}} 1_{K^1} \Big(\sum_{I^1_{n+1}} \frac{|I^1_{n+1}|^{\frac{s}{2}}}{|K^1|^s} \Big(\sum_{K^2} \frac{\langle |f_{n+1,K}| w, h^0_{I^1_{n+1}} \otimes h^0_{K^2}\rangle^2}{\langle \gamma_{n+1}\rangle_K^2} \frac{1_{K^2}}{|K^2|}\Big)^{\frac{s}{2}}\Big)^{\frac1s} \\
&\le 2^{\frac{k_{n+1} d_1}{s'}} 1_{K^1} \\
&\quad \otimes \Big(\sum_{I^1_{n+1}} \frac{|I^1_{n+1}|^{\frac{s}{2}}}{|K^1|^s} \Big(\sum_{K^2} \langle |f_{n+1,K}| w, h^0_{I^1_{n+1}} \otimes h^0_{K^2}\rangle^2 \big\langle \langle \gamma_{n+1}\rangle_{K^1,1}^{-\frac{1}{2n+1}}\big\rangle_{K^2}^{2\cdot(2n+1)} \frac{1_{K^2}}{|K^2|}\Big)^{\frac{s}{2}}\Big)^{\frac1s} \\
&\le 2^{\frac{k_{n+1} d_1}{s'}} 1_{K^1}
\end{aligned}$$

$$\otimes\Big(\sum_{I^1_{n+1}}\frac{|I^1_{n+1}|^{\frac{s}{2}}}{|K^1|^s}\Big(\sum_{K^2}\big[M_{\mathcal{D}^2}(\langle|f_{n+1,K}|w,h^0_{I^1_{n+1}}\rangle,\langle\gamma_{n+1}\rangle_{K^1,1}^{-\frac{1}{2n+1}},\cdots,\langle\gamma_{n+1}\rangle_{K^1,1}^{-\frac{1}{2n+1}})\big]^2\Big)^{\frac{s}{2}}\Big)^{\frac{1}{s}}.$$

Then the fact that

$$(\langle\gamma_{n+1}\rangle_{K^1,1}^{-\frac{1}{2}},\langle\gamma_{n+1}\rangle_{K^1,1}^{\frac{1}{2n+1}},\cdots,\langle\gamma_{n+1}\rangle_{K^1,1}^{\frac{1}{2n+1}})\in A_{(2,\infty,\cdots,\infty)}(\mathbb{R}^{d_2})$$

with characteristic independent of K^1 gives us that

$$\begin{aligned}
&\Big\|\Big(\sum_{K^1}|F_{n+1,K^1}|^2\Big)^{\frac{1}{2}}\gamma_{n+1}^{\frac{1}{2}}\Big\|_{L^2}^2\\
&\le 2^{\frac{2k_{n+1}d_1}{s'}}\sum_{K^1}\int_{\mathbb{R}^{d_2}}\langle\gamma_{n+1}\rangle_{K^1,1}\\
&\times\Big(\sum_{I^1_{n+1}}\frac{|I^1_{n+1}|^{\frac{s}{2}}}{|K^1|^{\frac{s}{2}}}\Big(\sum_{K^2}\big[M_{\mathcal{D}^2}(\langle|f_{n+1,K}|w,h^0_{I^1_{n+1}}\rangle,\langle\gamma_{n+1}\rangle_{K^1,1}^{-\frac{1}{2n+1}},\cdots,\langle\gamma_{n+1}\rangle_{K^1,1}^{-\frac{1}{2n+1}})\big]^2\Big)^{\frac{s}{2}}\Big)^{\frac{2}{s}}\\
&\lesssim 2^{\frac{2k_{n+1}d_1}{s'}}\sum_{K^1}\int_{\mathbb{R}^{d_2}}\Big(\sum_{I^1_{n+1}}\frac{|I^1_{n+1}|^{\frac{s}{2}}}{|K^1|^{\frac{s}{2}}}\Big(\sum_{K^2}\langle|f_{n+1,K}|w,h^0_{I^1_{n+1}}\rangle^2\Big)^{\frac{s}{2}}\Big)^{\frac{2}{s}}\langle\gamma_{n+1}\rangle_{K^1,1}^{-1}\\
&\le 2^{\frac{2k_{n+1}d_1}{s'}}\sum_{K^1}\int_{\mathbb{R}^{d_2}}\Big(\sum_{I^1_{n+1}}\frac{|I^1_{n+1}|^{\frac{1}{2}}}{|K^1|^{\frac{1}{2}}}\Big(\sum_{K^2}\langle|f_{n+1,K}|w,h^0_{I^1_{n+1}}\rangle^2\Big)^{\frac{1}{2}}\Big)^2\langle\gamma_{n+1}\rangle_{K^1,1}^{-1}.
\end{aligned}$$

By Minkowski's inequality,

$$\Big(\sum_{K^2}\langle|f_{n+1,K}|w,h^0_{I^1_{n+1}}\rangle^2\Big)^{\frac{1}{2}}\le\Big\langle\Big(\sum_{K^2}|f_{n+1,K}w|^2\Big)^{\frac{1}{2}},h^0_{I^1_{n+1}}\Big\rangle.$$

We are left with estimating

$$\begin{aligned}
&\sum_{K^1}\int_{\mathbb{R}^{d_2}}\Big(\sum_{(I^1_{n+1})^{(k_{n+1})}=K^1}\frac{|I^1_{n+1}|^{\frac{1}{2}}}{|K^1|^{\frac{1}{2}}}\Big\langle\Big(\sum_{K^2}|f_{n+1,K}w|^2\Big)^{\frac{1}{2}},h^0_{I^1_{n+1}}\Big\rangle\Big)^2\langle\gamma_{n+1}\rangle_{K^1,1}^{-1}\\
&\qquad=\int_{\mathbb{R}^d}\sum_{K^1}\frac{\Big\langle\Big(\sum_{K^2}|f_{n+1,K}w|^2\Big)^{\frac{1}{2}}\Big\rangle_{K^1}^2}{\langle\gamma_{n+1}\rangle_{K^1,1}^2}1_{K^1}\gamma_{n+1}\\
&\qquad\le\int_{\mathbb{R}^d}\sum_{K^1}\Big\langle\Big(\sum_{K^2}|f_{n+1,K}w|^2\Big)^{\frac{1}{2}}\Big\rangle_{K^1}^2\langle\gamma_{n+1}^{-\frac{1}{2n+1}}\rangle_{K^1,1}^{2\cdot(2n+1)}1_{K^1}\gamma_{n+1}\\
&\qquad\lesssim\int_{\mathbb{R}^d}\sum_{K^1}\Big(\sum_{K^2}|f_{n+1,K}w|^2\Big)^{\frac{1}{2}\cdot 2}\gamma_{n+1}^{-1}\\
&\qquad=\|\widetilde{f}_{n+1}v_{n+1}\|_{L^2}^2.
\end{aligned}$$

This completes the proof. The case $p\le 1$ follows from extrapolation [19]. □

5. The upper bound

In this section, we prove the following theorem.

5.1. **Theorem.** *Let $\vec{p} = (p_1, \dots, p_n)$ so that $1 < p_i \le \infty$, define $1/p = \sum_{i=1}^n 1/p_i > 0$. Let $(w_1, \dots, w_n), (\lambda_1, w_2, \dots, w_n) \in A_{\vec{p}}$ and let the associated Bloom weight $\nu = w_1 \lambda_1^{-1} \in A_\infty$. Assume that $b \in \operatorname{bmo}(\nu)$.*

For a multilinear bi-parameter dyadic model operator U, defined in the section 4.C, we have

$$\|[b, U]_1(f_1, \dots, f_n)\nu^{-1} w\|_{L^p} \lesssim_k \|b\|_{\operatorname{bmo}(\nu)} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}.$$

Here the constant depends on the complexity $k = (k_1, \dots, k_n) = ((k_1^1, k_1^2), \dots, (k_{n+1}^1, k_{n+1}^2))$ whenever U is a shift or a partial paraproduct. Dependence of the complexity is

$$\begin{cases} C_\beta 2^{\max_i k_i \beta} & \text{for every } \beta \in (0,1], \quad \text{if } U \text{ is a partial paraproduct} \\ (1 + \max\{k_1^1, k_1^2, k_{n+1}^1, k_{n+1}^2\})^{\frac{1}{2}}, & \text{if } U \text{ is a shift.} \end{cases} \tag{5.2}$$

We divide the analysis of each model operator into different subsections.

The boundedness of these model operator commutators yields the boundedness of the commutators of Calderón-Zygmund operators via Proposition 4.5. Use of Proposition 4.5 and complexity dependences (5.2) restricts the kernel regularity of (ω_1, ω_2)-CZOs in Theorem 1.1. For the paraproduct free CZOs, we can use milder kernel regularity, where we have that $\omega_i \in \operatorname{Dini}_{3/2}, i = 1, 2$. By paraproduct free, we mean that the paraproducts in the dyadic representation of T vanish, which could also be stated in terms of (both partial and full) "$T1 = 0$" type conditions. In the paraproduct free case, the reader can think of convolution form SIOs. Otherwise, we must use the standard Hölder type kernel regularity $\omega_i(t) = t^{\alpha_i}, \alpha_i \in (0, 1]$.

In the proof, we consider the boundedness $\prod_{i=1}^n L^{p_i}(w_i^{p_i}) \to L^p(\nu^{-p} w^p)$ for $p > 1$ since Theorem 1.2 will extend the result to the quasi-Banach range. Recall the notation of dual weights: $\sigma_i = w_i^{-p_i'}, \sigma_{n+1} = (\nu^{-1} w)^p$, and $\eta_1 = \lambda_1^{-p_1'}$. Here we chose to consider the commutators acting on the first function slot as the other ones are symmetrical.

The shift case. We consider the following commutator

$$[b, S_k]_1(f_1, \dots, f_n) = bS_k(f_1, \dots, f_n) - S_k(bf_1, \dots, f_n),$$

where $S_k := S^{1,2}_{(k_1, \dots, k_{n+1})}$ is a standard multilinear bi-parameter shift.

The idea is to expand the commutator so that a product bf paired with Haar functions is expanded in the bi-parameter fashion only if both of the Haar functions are cancellative. In a mixed situation, we expand only in $\mathbb{R}^{d_1}$ or $\mathbb{R}^{d_2}$, and in the remaining fully non-cancellative situation we do not expand at all. This strategy has been important in the recent multi-parameter results – see e.g. [1,3,20,22].

We focus on a commutator, where the cancellation appears in a mixed situation on first and last slots, that is, we have a commutator that is expanded as follows

$$\sum_{j_1=1}^3 \Pi^1_{j_1}(b, S_k(f_1, \dots, f_n)) - \sum_{j_2=1}^3 S_k(\Pi^2_{j_2}(b, f_1), \dots, f_n). \tag{5.3}$$

This case essentially gathers all the methods for estimating these commutators. More involved expansions are considered with partial paraproducts.

Both terms are handled separately whenever we have a bounded paraproduct, that is $\Pi^i_{j_i}, j_i \neq 3$ (or bi-parameter Π_{j_1,j_2}, $(j_1,j_2) \neq (3,3)$). Otherwise, we need to add and subtract certain averages of the function b to obtain enough cancellation. We analyse the second term in (5.3) as the first term is similar (swap the roles of functions f_1 and f_{n+1} together with weights η_1 and w^p).

We begin with the term

$$
\begin{aligned}
&S_k(\Pi^2_1(b,f_1),\dots,f_n)\\
&= \sum_{K^1\times K^2\in\mathcal{D}^1\times\mathcal{D}^2} \sum_{\substack{I^j_i\in\mathcal{D}^j\\ (I^j_i)^{(k^j_i)}=K^j\\ i=1,\dots,n+1,j=1,2}} a_{K(I^j_i)} \Big\langle \sum_{J^2\in\mathcal{D}^2} \langle b, h_{J^2}\rangle_2 \langle f_1, h_{J^2}\rangle_2 \otimes h_{J^2}h_{J^2}, h^0_{I^1_1}\otimes h_{I^2_1}\Big\rangle \\
&\qquad\qquad \times \prod_{i=2}^n \langle f_i, \widetilde{h}_{I_i}\rangle h_{I^1_{n+1}} \otimes h^0_{I^2_{n+1}}.
\end{aligned}
$$

By the zero average of Haar functions, we always have $J^2 \subset I^2_1$. Now the important observation is that when $J^2 \subsetneq I^2_1$ we must have $h_{J^2}h_{J^2} = \frac{1_{J^2}}{|J^2|}$ and we can replace $\langle \frac{1_{J^2}}{|J^2|}, h_{I^2_1}\rangle$ with $\langle \frac{\eta_1 1_{J^2}}{\eta_1(J^2)}, h_{I^2_1}\rangle$. Thus, we can change the order of the operators, and we can split the dual form of the term as follows

$$
\begin{aligned}
&\langle S_k(\Pi^2_1(b,f_1),\dots,f_n), f_{n+1}\rangle\\
&= \sum_{K\in\mathcal{D}} \sum_{\substack{I^j_i\in\mathcal{D}^j\\ (I^j_i)^{(k^j_i)}=K^j\\ i=1,\dots,n+1,j=1,2}} a_{K(I^j_i)} \Big\langle \sum_{J^2\subsetneq I^2_1\in\mathcal{D}^2} \langle b, h_{J^2}\rangle_2 \langle f_1, h_{J^2}\rangle_2 \frac{\eta_1 1_{J^2}}{\eta_1(J^2)}, h^0_{I^1_1}\otimes h_{I^2_1}\Big\rangle \\
&\quad \times \prod_{i=2}^n \langle f_i, \widetilde{h}_{I_i}\rangle \langle f_{n+1}, h_{I^1_{n+1}} \otimes h^0_{I^2_{n+1}}\rangle + \sum_{K\in\mathcal{D}} \sum_{\substack{I^j_i\in\mathcal{D}^j\\ (I^j_i)^{(k^j_i)}=K^j\\ i=1,\dots,n+1,j=1,2}} a_{K(I^j_i)} \Big\langle \langle b, h_{I^2_1}\rangle_2 \langle f_1, h_{I^2_1}\rangle_2, h^0_{I^1_1}\Big\rangle\\
&\quad \times \langle h_{I^2_1}h_{I^2_1}, h_{I^2_1}\rangle \prod_{i=2}^n \langle f_i, \widetilde{h}_{I_i}\rangle \langle f_{n+1}, h_{I^1_{n+1}} \otimes h^0_{I^2_{n+1}}\rangle \\
&= \int_{\mathbb{R}^{d_1}} \sum_{J^2\in\mathcal{D}^2} \langle b, h_{J^2}\rangle_2 \langle f_1, h_{J^2}\rangle_2 \langle S^{1*}_k(f_2,\dots,f_{n+1})\rangle^{\eta_1}_{J^2,2}\\
&\qquad\qquad - \int_{\mathbb{R}^{d_1}} \sum_{J^2\in\mathcal{D}^2} \langle b, h_{J^2}\rangle_2 \langle f_1, h_{J^2}\rangle_2 \langle S^{1*}_{k,J^2}(f_2,\dots,f_{n+1})\rangle^{\eta_1}_{J^2,2} + E,
\end{aligned}
$$

where S^{1*}_{k,J^2} differs from the usual adjoint so that we have $I^2_1 \subset J^2$. We do not explicitly handle the term E as it is similar to the case $j_2 = 2$ (note that we have more cancellation than we need). Since the truncated operator S^{1*}_{k,J^2} can be dominated by the A_∞ weighted square functions lower bound, we can drop the dependence on cube J^2. Then, the estimations of the first two terms are very similar, hence one might think of S^{1*}_k as such or as $S^2_{\mathcal{D}} S^{1*}_k$ below. The boundedness follows simply by using Proposition 4.18 and the boundedness of multilinear shifts. Namely,

$$\begin{aligned}\int_{\mathbb{R}^{d_1}} \sum_{J^2\in\mathcal{D}^2} \langle b, h_{J^2}\rangle_2 \langle f_1, h_{J^2}\rangle_2 \langle S^{1*}_k(f_2,\ldots,f_{n+1})\rangle^{\eta_1}_{J^2,2} \\ &\lesssim \|b\|_{\operatorname{bmo}(\nu)} \|f_1 w_1\|_{L^{p_1}} \|S^{1*}_k(f_2,\ldots,f_{n+1})\|_{L^{p_1'}(\eta_1)} \\ &\lesssim \|b\|_{\operatorname{bmo}(\nu)} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}} \|f_{n+1}\nu w^{-1}\|_{L^{p'}}\end{aligned}$$

since $\eta_1^{1/p_1'} = \lambda_1^{-1} = \nu w^{-1}\prod_{i=2}^n w_i$ and $(w_2,\ldots,w_n,\nu w^{-1}) \in A_{(p_2,\ldots,p_n,p')}$ by Lemma 2.5.

The term, where $j_2 = 2$, is significantly more straightforward to estimate. We consider the dual form and estimate

$$\begin{aligned}&|\langle S(\Pi^2_2(b,f_1),\ldots,f_n), f_{n+1}\rangle| \\ &= \Big| \sum_{K\in\mathcal{D}} \sum_{\substack{I^j_i\in\mathcal{D}^j \\ (I^j_i)^{(k^j_i)}=K^j \\ i=1,\ldots,n+1, j=1,2}} a_{K(I^j_i)} \Big\langle \langle b, h_{I^2_1}\rangle_2 \langle f_1\rangle_{I^2_1,2}, h^0_{I^1_1}\Big\rangle \prod_{i=2}^n \langle f_i, \widetilde{h}_{I_i}\rangle \langle f_{n+1}, h_{I^1_{n+1}} \otimes h^0_{I^2_{n+1}}\rangle\Big| \\ &\le \int_{\mathbb{R}^{d_1}} \sum_{K^2} \sum_{(I^2_1)^{(k^2_1)}=K^2} |I^2_1|^{1/2} |\langle b, h_{I^2_1}\rangle_2| \langle |f_1|\rangle_{I^2_1,2} A_{K^2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1}) \\ &= \int_{\mathbb{R}^{d_1}} \sum_{K^2} \sum_{(I^2_1)^{(k^2_1)}=K^2} |I^2_1|^{1/2} |\langle b, h_{I^2_1}\rangle_2 \langle \sigma_1\rangle_{I^2_1,2}| \frac{\langle |f_1|\rangle_{I^2_1,2}}{\langle \sigma_1\rangle_{I^2_1,2}} A_{K^2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1})\end{aligned}$$

where $A_{K^2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}$, $i^m_1 \in \{2,\ldots,n\}$ is from family of operators such that the square sum

$$\Big(\sum_{K^2} A^2_{K^2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1}) 1_{K^2}\Big)^{\frac12}$$

is an $A_{2,k}$ type square function. We use Lemma 3.4 with a fixed variable on the first parameter and get

$$\begin{aligned}&|\langle S(\Pi^2_2(b,f_1),\ldots,f_n), f_{n+1}\rangle| \\ &\lesssim \|b\|_{\operatorname{bmo}(\nu)} \int \Big(\sum_{K^2} \sum_{(I^2_1)^{(k^2_1)}=K^2} \frac{\langle |f_1|\rangle^2_{I^2_1,2}}{\langle\sigma_1\rangle^2_{I^2_1,2}} A^2_{K^2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1}) 1_{I^2_1}\Big)^{1/2} \sigma_1\nu \\ &\le \|b\|_{\operatorname{bmo}(\nu)} \int M^{\sigma_1}_{\mathcal{D}^2}(f_1\sigma_1^{-1}) A_{2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1})\sigma_1\nu \\ &\le \|b\|_{\operatorname{bmo}(\nu)} \|M^{\sigma_1}_{\mathcal{D}^2}(f_1\sigma_1^{-1})\|_{L^{p_1}(\sigma_1)} \|A_{2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1})\lambda_1^{-1}\|_{L^{p_1'}} \\ &\lesssim \|b\|_{\operatorname{bmo}(\nu)} \|f_1 w_1\|_{L^{p_1}} \Big\| A_{2,(k^1_{n+1},k^1_{i^1_1},k^2_{i^2_1})}(f_2,\ldots,f_{n+1})\nu w^{-1}\prod_{i=2}^n w_i\Big\|_{L^{p_1'}}\end{aligned}$$

$$\lesssim \|b\|_{\mathrm{bmo}(\nu)} \prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}} \|f_{n+1} \nu w^{-1}\|_{L^{p'}}.$$

In the above estimates, it is enough to note that the maximal function is bounded since, by Fubini's theorem, we can work with a fixed variable on the first parameter and use the classical one-parameter result.

Lastly, we are left with the paraproducts of the illegal form

$$\begin{aligned}
&\langle \Pi_3^1(b, S_k(f_1, \dots, f_n)), f_{n+1}\rangle - \langle S_k(\Pi_3^2(b, f_1), \dots, f_n), f_{n+1}\rangle \\
&= \sum_K \sum_{\substack{I_i^j \in \mathcal{D}^j \\ (I_i^j)^{(k_i^j)} = K^j \\ i=1,\dots,n+1, j=1,2}} a_{K(I_i^j)} \langle f_1, h_{I_1^1}^0 \otimes h_{I_1^2}\rangle \prod_{i=2}^{n} \langle f_i, \widetilde{h}_{I_i}\rangle \Big\langle \langle b\rangle_{I_{n+1}^1,1} f_{n+1}, h_{I_{n+1}^1} \otimes h_{I_{n+1}^2}^0 \Big\rangle \\
&\quad - \sum_K \sum_{\substack{I_i^j \in \mathcal{D}^j \\ (I_i^j)^{(k_i^j)} = K^j \\ i=1,\dots,n+1, j=1,2}} a_{K(I_i^j)} \Big\langle \langle b\rangle_{I_1^2,2} f_1, h_{I_1^1}^0 \otimes h_{I_1^2} \Big\rangle \prod_{i=2}^{n} \langle f_i, \widetilde{h}_{I_i}\rangle \langle f_{n+1}, h_{I_{n+1}^1} \otimes h_{I_{n+1}^2}^0\rangle.
\end{aligned}$$

Here we introduce the martingale blocks to the function b. We write

$$\begin{aligned}
\langle b\rangle_{I_{n+1}^1,1} = \langle b\rangle_{I_{n+1}^1,1} - \langle b\rangle_{I_{n+1}^1 \times I_{n+1}^2} + \langle b\rangle_{I_{n+1}^1 \times I_{n+1}^2} - \langle b\rangle_{I_{n+1}^1 \times K^2} \\
+ \langle b\rangle_{I_{n+1}^1 \times K^2} - \langle b\rangle_{K^1 \times K^2} + \langle b\rangle_{K^1 \times K^2}
\end{aligned}$$

and likewise for $\langle b\rangle_{I_1^2,2}$. The extra $\langle b\rangle_{K^1\times K^2}$ simply cancels with the one from $\langle b\rangle_{I_1^2,2}$. Hence, in the commutator we can expand as follows

$$(\langle b\rangle_{I_{n+1}^1,1} - \langle b\rangle_{I_{n+1}^1 \times I_{n+1}^2}) 1_{I_{n+1}^2} = \sum_{J^2 \subset I_{n+1}^2} \Big\langle b, \frac{1_{I_{n+1}^1}}{|I_{n+1}^1|} \otimes h_{J^2} \Big\rangle h_{J^2}, \tag{5.4}$$

$$\langle b\rangle_{I_{n+1}^1 \times K^2} - \langle b\rangle_{K^1 \times K^2} = \sum_{I_{n+1}^1 \subsetneq J^1 \subset K^1} \Big\langle b, h_{J^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \langle h_{J^1}\rangle_{I_{n+1}^1}, \tag{5.5}$$

$$\langle b\rangle_{I_{n+1}^1 \times I_{n+1}^2} - \langle b\rangle_{I_{n+1}^1 \times K^2} = \sum_{I_{n+1}^2 \subsetneq J^2 \subset K^2} \Big\langle b, \frac{1_{I_{n+1}^1}}{|I_{n+1}^1|} \otimes h_{J^2} \Big\rangle \langle h_{J^2}\rangle_{I_{n+1}^2}. \tag{5.6}$$

Observe that we have omitted the terms raised from $\langle b\rangle_{I_1^2,2}$ because they are similar. On the other hand, we shall only work with (5.4) and (5.5) because (5.6) is analogous.

We begin with the dual form of (5.4)

$$\begin{aligned}
\Big| \sum_K \sum_{\substack{I_i^j \in \mathcal{D}^j \\ (I_i^j)^{(k_i^j)} = K^j \\ i=1,\dots,n+1, j=1,2}} a_{K(I_i^j)} \sum_{J^2 \subset I_{n+1}^2} \Big\langle b, \frac{1_{I_{n+1}^1}}{|I_{n+1}^1|} \otimes h_{J^2} \Big\rangle |I_{n+1}^2|^{-\frac{1}{2}} \langle f_1, h_{I_1^1}^0 \otimes h_{I_1^2}\rangle \\
\times \prod_{i=2}^{n} \langle f_i, \widetilde{h}_{I_i}\rangle \langle f_{n+1}, h_{I_{n+1}^1} \otimes h_{J^2}\rangle \Big|.
\end{aligned}$$

By similar arguments as that in the proof of Lemma 3.3, we have

$$\sum_{I_{n+1}^{(k_{n+1})}=K} |I_{n+1}^1|^{\frac12}|I_{n+1}^2|^{\frac12} \sum_{J^2\subset I_{n+1}^2} \Big\langle b, \frac{1_{I_{n+1}^1}}{|I_{n+1}^1|}\otimes h_{J^2}\Big\rangle |I_{n+1}^2|^{-\frac12}\langle f_{n+1}, h_{I_{n+1}^1}\otimes h_{J^2}\rangle$$
$$\lesssim \|b\|_{\mathrm{bmo}(\nu)} \sum_{I_{n+1}^{(k_{n+1})}=K} \int_{\mathbb{R}^{d_1}} h^0_{I^1_{n+1}} \int_{\mathbb{R}^{d_2}} \Big(\sum_{J^2\subset I^2_{n+1}} \langle f_{n+1}, h_{I_{n+1}^1}\otimes h_{J^2}\rangle^2 \frac{1_{J^2}}{|J^2|}\Big)^{\frac12}\nu$$
$$\lesssim \|b\|_{\mathrm{bmo}(\nu)} \sum_{(I_{n+1}^1)^{(k_{n+1}^1)}=K^1} \int_{\mathbb{R}^d} h^0_{I^1_{n+1}}\otimes 1_{K^2}\Big(\sum_{J^2}\frac{\langle f_{n+1}, h_{I_{n+1}^1}\otimes h_{J^2}\rangle^2}{\langle \sigma_{n+1}\rangle^2_{I^1_{n+1}\times J^2}}\frac{1_{J^2}}{|J^2|}\Big)^{\frac12}\sigma_{n+1}\nu.$$

Then by standard calculus, we can reduce the problem to

$$\int_{\mathbb{R}^d}\sum_{K^1} A_{K^1,k_1^2,k^1_{i^1_1},k^2_{i^2_1}}(f_1,\cdots,f_n)1_{K^1}\sum_{(I_{n+1}^1)^{(k_{n+1}^1)}=K^1} h^0_{I^1_{n+1}}\Big(\sum_{J^2}\frac{\langle f_{n+1}, h_{I_{n+1}^1}\otimes h_{J^2}\rangle^2}{\langle \sigma_{n+1}\rangle^2_{I^1_{n+1}\times J^2}}\frac{1_{J^2}}{|J^2|}\Big)^{\frac12}\sigma_{n+1}\nu,$$

where $A_{K^1,k_1^2,k^1_{i^1_1},k^2_{i^2_1}}(f_1,\cdots,f_n)$ is defined such that

$$\Big(\sum_{K^1}\big[A_{K^1,k_1^2,k^1_{i^1_1},k^2_{i^2_1}}(f_1,\cdots,f_n)\big]^2 1_{K^1}\Big)^{\frac12}$$

is an $A_{2,k}$ type square function. Notice that $\sigma_{n+1}\nu = (\nu^{-1}w)^p\nu = (w^p)^{\frac1p}((\nu^{-1}w)^p)^{\frac{1}{p'}} \in A_\infty$. The rest follows from estimates such as Hölder's inequality, Theorem 2.13, and Proposition 2.14.

Finally, we consider the dual form of (5.5)

$$\Big|\sum_K \sum_{\substack{I_i^j\in\mathcal{D}^j\\ (I_i^j)^{(k_i^j)}=K^j\\ i=1,\dots,n+1, j=1,2}} a_{K(I_i^j)} \sum_{I^1_{n+1}\subsetneq J^1\subset K^1}\Big\langle b, h_{J^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\langle h_{J^1}\rangle_{I^1_{n+1}}\prod_{i=1}^{n+1}\langle f_i,\widetilde{h}_{I_i}\rangle\Big|$$
$$\le \int_{\mathbb{R}^{d_2}}\sum_{K^1}\sum_{\substack{J^1\subset K^1\\ \ell(J^1)>2^{-k^1_{n+1}}\ell(K^1)}} |J^1|^{-\frac12}|\langle b, h_{J^1}\rangle_1|\sum_{K^2}\sum_{\substack{I_i^j\in\mathcal{D}^j\\ (I_i^j)^{(k_i^j)}=K^j\\ I^1_{n+1}\subset J^1\\ i=1,\dots,n+1,j=1,2}} |a_{K(I_i^j)}|\prod_{i=1}^{n+1}|\langle f_i,\widetilde{h}_{I_i}\rangle|\frac{1_{K^2}}{|K^2|}$$
$$\le \int_{\mathbb{R}^{d_2}}\sum_{K^1}\sum_{j^1_{n+1}=0}^{k^1_{n+1}-1}\sum_{(J^1)^{(j^1_{n+1})}=K^1}|J^1|^{\frac12}|\langle b, h_{J^1}\rangle_1|$$
$$\times\sum_{K^2} A_{K,(k_1^2,k^1_{i^1_1},k^2_{i^2_1})}(f_1,\dots,f_n)\langle|\Delta^1_{J^1,k^1_{n+1}-j^1_{n+1}}f_{n+1}|\rangle_{J^1\times K^2}1_{K^2},$$

where $A_{K,(k_1^2,k^1_{i_1^1},k^2_{i_1^2})}$ is defined such that

$$\Big(\sum_{K^1}\Big(\sum_{K^2} A_{K,(k_1^2,k^1_{i_1^1},k^2_{i_1^2})}(f_1,\dots,f_n)1_K\Big)^2\Big)^{\frac12}$$

is an $A_{2,k}$ type square function.

This resembles the term that we faced earlier with paraproduct Π_2. The only meaningful difference is the extra summation. The estimations are similar when we divide and multiply with $\langle \sigma_{n+1}\rangle_{J^1\times K^2}$. To be more precise, that is, we write

$$\begin{aligned}
&\int_{\mathbb{R}^{d_2}} \sum_{j^1_{n+1}=0}^{k^1_{n+1}-1} \sum_{(J^1)^{(j^1_{n+1})}=K^1} |J^1|^{\frac12}|\langle b,h_{J^1}\rangle_1||\langle|\Delta^1_{J^1,k^1_{n+1}-j^1_{n+1}} f_{n+1}|\rangle_{J^1\times K^2}1_{K^2}\\
&\quad= \int_{\mathbb{R}^{d_2}} \sum_{\substack{(J^1)^{(j^1_{n+1})}=K^1\\ 0\le j^1_{n+1}\le k^1_{n+1}-1}} |J^1|^{\frac12}|\langle b,h_{J^1}\rangle_1||\langle \sigma_{n+1}\rangle_{J^1\times K^2}\frac{\langle|\Delta^1_{J^1,k^1_{n+1}-j^1_{n+1}} f_{n+1}|\rangle_{J^1\times K^2}}{\langle \sigma_{n+1}\rangle_{J^1\times K^2}}1_{K^2}\\
&\quad\lesssim \|b\|_{\mathrm{bmo}(\nu)}\int_{\mathbb{R}^d}\Big(\sum_{\substack{(J^1)^{(j^1_{n+1})}=K^1\\ 0\le j^1_{n+1}\le k^1_{n+1}-1}} \frac{\langle|\Delta^1_{J^1,k^1_{n+1}-j^1_{n+1}} f_{n+1}|\rangle^2_{J^1\times K^2}}{\langle \sigma_{n+1}\rangle^2_{J^1\times K^2}}1_{J^1}\Big)^{\frac12}\nu\sigma_{n+1}1_{K^2},
\end{aligned}$$

where we have used Lemma 3.3. The rest of the argument is rather standard and thus the object is bounded by

$$(1+k^1_{n+1})^{\frac12}\|b\|_{\mathrm{bmo}(\nu)}\prod_{i=1}^n \|f_iw_i\|_{L^{p_i}}\|f_{n+1}\nu w^{-1}\|_{L^{p'}},$$

where dependence $(1+k^1_{n+1})^{\frac12}$ emerges from the summation of $0\le j^1_{n+1}\le k^1_{n+1}-1$. This completes the analysis of the commutator of this form.

Although other forms of shifts lead to different expansions, the methods shown above are sufficient to handle those as well. Since we are dealing with multilinear shifts, we now encounter terms in the shift case that are non-cancellative. In comparison, this does not happen in the linear case in [22], where we always expand in the bi-parameter fashion. For example, if we look at the term $bS(f_1,\dots,f_n)-S(\Pi^{1,2}_{3,3}(b,f_1),\dots,f_n)$, we have

$$(b-\langle b\rangle_{I_1^1\times I_1^2})1_{I^1_{n+1}\times I^2_{n+1}}.$$

We write

$$\begin{aligned}
(b-\langle b\rangle_{I_1^1\times I_1^2})1_{I^1_{n+1}\times I^2_{n+1}} &= ((b-\langle b\rangle_{I^1_{n+1}\times I^2_{n+1}})+(\langle b\rangle_{I^1_{n+1}\times I^2_{n+1}}-\langle b\rangle_{I_1^1\times I_1^2}))1_{I^1_{n+1}\times I^2_{n+1}}\\
&= (b-\langle b\rangle_{I^1_{n+1},1}-\langle b\rangle_{I^2_{n+1},2}+\langle b\rangle_{I^1_{n+1}\times I^2_{n+1}})1_{I^1_{n+1}\times I^2_{n+1}}\\
&\quad+(\langle b\rangle_{I^1_{n+1},1}-\langle b\rangle_{I^1_{n+1}\times I^2_{n+1}})1_{I^1_{n+1}\times I^2_{n+1}}\\
&\quad+(\langle b\rangle_{I^2_{n+1},2}-\langle b\rangle_{I^1_{n+1}\times I^2_{n+1}})1_{I^1_{n+1}\times I^2_{n+1}}\\
&\quad+(\langle b\rangle_{I^1_{n+1}\times I^2_{n+1}}-\langle b\rangle_{I_1^1\times I_1^2})1_{I^1_{n+1}\times I^2_{n+1}}.
\end{aligned}$$

The above terms are expanded to the martingale blocks and differences in a standard way like terms (5.4) and (5.5). Note that the first term on the right-hand side produces a bi-parameter martingale difference inside of the rectangle $I^1_{n+1} \times I^2_{n+2}$. We will analyse similar terms in the following subsection.

Partial paraproducts. As explained earlier, we will now focus on more involved expansions of the commutator. We show the most representative case out of those. Although we demonstrated the main ideas of the estimates already in the shift case, we need to use more complex estimates due to the more complicated structure of the partial paraproducts.

We do not repeat the expansion strategy and instead straight away consider separately

$$\begin{aligned}&\langle (S\pi)(\Pi^{1,2}_{j_1,j_2}(b,f_1),\dots,f_n),f_{n+1}\rangle \\ &= \sum_{K^1,K^2}\sum_{(I^1_i)^{(k_i)}=K^1} a_{K(I^1_i)}\langle \Pi^{1,2}_{j_1,j_2}(b,f_1),h_{I^1_1}\otimes h_{K^2}\rangle \prod_{i=2}^{n+1}\Big\langle f_i,\widetilde h_{I^1_i}\otimes \frac{1_{K^2}}{|K^2|}\Big\rangle\end{aligned}$$

for $(j_1,j_2)\neq(3,3)$. We collect most of the mixed index ($j_1\neq j_2$) cases, as the methods can be attained from these.

Let us begin with the term, where $j_1=1, j_2=2$, that equals

$$\begin{aligned}&\sum_{K^1,K^2}\sum_{(I^1_i)^{(k_i)}=K^1} a_{K(I^1_i)}\langle \Pi^{1,2}_{1,2}(b,f_1),h_{I^1_1}\otimes h_{K^2}\rangle \prod_{i=2}^{n+1}\Big\langle f_i,\widetilde h_{I^1_i}\otimes \frac{1_{K^2}}{|K^2|}\Big\rangle \\ &= \sum_{J^1,K^2}\langle b,h_{J^1}\otimes h_{K^2}\rangle\Big\langle f_1,h_{J^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle \\ &\qquad\times\Big\langle \sum_{K^1}\sum_{(I^1_i)^{(k_i)}=K^1} a_{K(I^1_i)}\prod_{i=2}^{n+1}\Big\langle f_i,\widetilde h_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle h_{I^1_1}\Big\rangle^{\langle\eta_1\rangle_{K^2,2}}_{J^1} \\ &- \sum_{J^1,K^2}\langle b,h_{J^1}\otimes h_{K^2}\rangle\Big\langle f_1,h_{J^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle \\ &\qquad\times\Big\langle \sum_{K^1}\sum_{\substack{(I^1_i)^{(k_i)}=K^1\\ I^1_1\subset J^1}} a_{K(I^1_i)}\prod_{i=2}^{n+1}\Big\langle f_i,\widetilde h_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle h_{I^1_1}\Big\rangle^{\langle\eta_1\rangle_{K^2,2}}_{J^1} \\ &+ \sum_{K^1,K^2}\sum_{(I^1_i)^{(k_i)}=K^1} a_{K(I^1_i)}\langle b,h_{I^1_1}\otimes h_{K^2}\rangle |I^1_1|^{-\frac12}\prod_{i=1}^{n+1}\Big\langle f_i,\widetilde h_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle.\end{aligned}$$

Similarly to the previously seen techniques, for the second term we use the square function lower bound to get rid of the restriction $I^1_1\subset J^1$. Thus, via Proposition 4.19 we can bound the first two terms by

$$\|b\|_{\mathrm{bmo}(\nu)}\|f_1w_1\|_{L^{p_1}}\|S_{\mathcal D}(S\pi)_k(f_2,\dots,f_{n+1})\lambda_1^{-1}\|_{L^{p_1'}}.$$

Clearly, Lemma 4.20 is enough to conclude the claim. The estimate for the remaining term is easier. We apply Lemma 3.3 and note that we have more cancellation than we need. Hence, we control

$$|I_1^1|^{-\frac{1}{2}}\Big|\Big\langle f_1, h_{I_1^1} \otimes \frac{1_{K^2}}{|K^2|}\Big\rangle\Big|\langle \sigma_1\rangle_{I_1^1\times K^2}^{-1} 1_{I_1^1\times K^2} \le M_{\mathcal{D}}^{\sigma_1}(f_1\sigma_1^{-1})1_{I_1^1\times K^2}.$$

Thus, we are left to estimate

$$\|b\|_{\mathrm{bmo}(\nu)}\|M_{\mathcal{D}}^{\sigma_1}(f_1\sigma_1^{-1})S_{\mathcal{D}}(S\pi)_k(f_2,\dots,f_{n+1})\sigma_1\nu\|_{L^1}.$$

The desired estimate follows by Hölder's inequality and Lemma 4.20. We remark that the remaining term essentially contains the idea to handle $\Pi_{1,1}$.

The term with $\Pi_{2,1}$ is analogous to the previous one. We remark that in this case, the weighted paraproduct operator has the weight $\langle \eta_1\rangle_{I_1^1}$ as the localization of the operator is at that level on the first parameter. The cases $\Pi_{3,1}$ and $\Pi_{1,3}$ can be handled similarly. For the sake of the completeness, we give a sketch of the case $\Pi_{3,1}$. As before, we write

$$\begin{aligned}
&\langle (S\pi)(\Pi_{3,1}(b,f_1),\dots,f_n), f_{n+1}\rangle \\
&\quad= \sum_{K^1,J^2}\sum_{(I_1^1)^{(k_1)}=K^1}\Big\langle b, \frac{1_{I_1^1}}{|I_1^1|}\otimes h_{J^2}\Big\rangle \langle f_1, h_{I_1^1}\otimes h_{J^2}\rangle \\
&\qquad\qquad \times \Big\langle \sum_{K^2\supsetneq J^2}\sum_{\substack{(I_i^1)^{(k_i)}=K^1\\ i\ne 1}} a_{K(I_i^1)}\prod_{i=2}^{n+1}\Big\langle f_i, \widetilde{h}_{I_i^1}\otimes \frac{1_{K^2}}{|K^2|}\Big\rangle h_{K^2}\Big\rangle_{J^2} \\
&\quad+ \sum_{K^1,K^2}\sum_{(I_i^1)^{(k_i)}=K^1} a_{K(I_i^1)}\langle b, \frac{1_{I_1^1}}{|I_1^1|}\otimes h_{K^2}\rangle |K^2|^{-\frac{1}{2}}\langle f_1, h_{I_1^1}\otimes h_{K^2}\rangle \prod_{i=2}^{n+1}\Big\langle f_i, \widetilde{h}_{I_i^1}\otimes \frac{1_{K^2}}{|K^2|}\Big\rangle.
\end{aligned}$$

For the second term, we again use Lemma 3.3 and treat

$$|K^2|^{-\frac{1}{2}}|\langle f_1, h_{I_1^1}\otimes h_{K^2}\rangle|\langle\sigma_1\rangle_{I_1^1\times K^2}^{-1}1_{I_1^1\times K^2} \le M_{\mathcal{D}^2}^{\langle\sigma_1\rangle_{I_1^1,1}}(\langle f_1, h_{I_1^1}\rangle_1\langle\sigma_1\rangle_{I_1^1,1}^{-1})1_{I_1^1\times K^2}.$$

Then after applying Hölder's inequality twice we reduce the problem to bounding

$$\begin{aligned}
\|b\|_{\mathrm{bmo}(\nu)}\Big\|\Big(\sum_{K^1}\sum_{(I_1^1)^{(k_1)}=K^1}\big[M_{\mathcal{D}^2}^{\langle\sigma_1\rangle_{I_1^1,1}}(\langle f_1, h_{I_1^1}\rangle_1\langle\sigma_1\rangle_{I_1^1,1}^{-1})\big]^2 1_{I_1^1}\Big)^{\frac{1}{2}}\Big\|_{L^{p_1}(\sigma_1)} \\
\times \|S_{\mathcal{D}}(S\pi)_k(f_2,\dots,f_{n+1})\lambda_1^{-1}\|_{L^{p_1'}}.
\end{aligned}$$

The estimate is done by Proposition 2.14 and Lemma 4.20. For the first term, we split as usual to

$$\begin{aligned}
&\sum_{K^1,J^2}\sum_{(I_1^1)^{(k_1)}=K^1}\Big\langle b, \frac{1_{I_1^1}}{|I_1^1|}\otimes h_{J^2}\Big\rangle\langle f_1, h_{I_1^1}\otimes h_{J^2}\rangle \\
&\qquad\qquad \times\Big\langle \sum_{K^2}\sum_{\substack{(I_i^1)^{(k_i)}=(I_1^1)^{(k_1)}\\ i\ne 1}} a_{K(I_i^1)}\prod_{i=2}^{n+1}\Big\langle f_i, \widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle h_{K^2}\Big\rangle_{J^2}^{\langle\eta_1\rangle_{I_1^1,1}} \\
&- \sum_{K^1,J^2}\sum_{(I_1^1)^{(k_1)}=K^1}\Big\langle b, \frac{1_{I_1^1}}{|I_1^1|}\otimes h_{J^2}\Big\rangle\langle f_1, h_{I_1^1}\otimes h_{J^2}\rangle
\end{aligned}$$

$$\times \Big\langle \sum_{K^2 \subset J^2} \sum_{\substack{(I_i^1)^{(k_i)}=(I_1^1)^{(k_1)} \\ i \ne 1}} a_{K(I_i^1)} \prod_{i=2}^{n+1} \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle h_{K^2} \Big\rangle_{J^2}^{\langle \eta_1 \rangle_{I_1^1,1}}.$$

We focus on the first term as the other one is very similar once square function lower bound is applied inside of the average over J^2. Rewrite the first term as

$$\sum_{J^1,J^2} \Big\langle b, \frac{1_{J^1}}{|J^1|} \otimes h_{J^2} \Big\rangle \langle f_1, h_{J^1} \otimes h_{J^2} \rangle \Big\langle \langle (S\pi)_k(f_2, \cdots, f_{n+1}), h_{J^1} \rangle_1 \Big\rangle_{J^2}^{\langle \eta_1 \rangle_{J^1,1}}.$$

Then the estimate is done by Proposition 4.19 and Lemma 4.20.

We continue with the term, where $j_1 = 2, j_2 = 3$, that is,

$$\sum_{K^1,K^2} \sum_{(I_i^1)^{(k_i)}=K^1} a_{K(I_i^1)} \langle \Pi^{1,2}_{2,3}(b, f_1), h_{I_1^1} \otimes h_{K^2} \rangle \prod_{i=2}^{n+1} \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle$$

$$= \sum_{K^1,K^2} \sum_{(I_1^1)^{(k_1)}=K^1} \Big\langle b, h_{I_1^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \langle \langle f_1, h_{K^2} \rangle_2 \rangle_{I_1^1} \sum_{\substack{(I_i^1)^{(k_i)}=K^1 \\ i \ne 1}} a_{K(I_i^1)} \prod_{i=2}^{n+1} \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle.$$

Note that we can rewrite the above as

$$\sum_{J^1,J^2} \Big\langle b, h_{J^1} \otimes \frac{1_{J^2}}{|J^2|} \Big\rangle \langle \sigma_1 \rangle_{J^1 \times J^2} \frac{\Big\langle f_1, \frac{1_{J^1}}{|J^1|} \otimes h_{J^2} \Big\rangle}{\langle \sigma_1 \rangle_{J^1 \times J^2}} \langle (S\pi)_k(f_2, \cdots, f_n), h_{J^1} \otimes h_{J^2} \rangle.$$

Then there is nothing new here; by Lemma 3.3 we have that the above is dominated by

$$\|b\|_{\mathrm{bmo}(\nu)} \Big\| \sum_{J^2} \Big(\sum_{J^1} [M_{\mathcal{D}^1}^{\langle \sigma_1 \rangle_{J^2,2}} (\langle f_1, h_{J^2} \rangle_2 \langle \sigma_1 \rangle_{J^2,2}^{-1})]^2$$
$$\times \langle (S\pi)_k(f_2, \cdots, f_n), h_{J^1} \otimes h_{J^2} \rangle^2 \frac{1_{J^1}}{|J^1|} \Big)^{\frac{1}{2}} \frac{1_{J^2}}{|J^2|} \Big\|_{L^1(\sigma_1 \nu)}.$$

The estimate is then completed by applying Hölder's inequality twice, Proposition 2.14 and Lemma 4.20.

Symmetrically, we can work with $\Pi_{3,2}$. Lastly, we focus on terms with $\Pi_{3,3}$ type illegal paraproducts. We choose here the type of term which we did not consider in the shift section:

$$\langle (S\pi)_k(\Pi_{3,3}(b, f_1), \ldots, f_n) - b(S\pi)_k(f_1 \ldots, f_n), f_{n+1} \rangle.$$

Notice that we have

(5.7)

$$\langle \Pi^{1,2}_{3,3}(b, f_1), h_{I_1^1} \otimes h_{K^2} \rangle \Big\langle f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle - \langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle \Big\langle b f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle$$
$$= \langle b \rangle_{I_1^1 \times K^2} \langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle \Big\langle f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle - \langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle \Big\langle b f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle$$
$$= \Big(\langle b \rangle_{I_1^1 \times K^2} - \langle b \rangle_{I^1_{n+1} \times K^2} \Big) \langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle \Big\langle f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle$$
$$- \langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle \Big\langle (b - \langle b \rangle_{I^1_{n+1} \times K^2}) f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle.$$

Now on the right-hand side of the above equation (5.7), we have two distinct cases where the first part is similar to the ones seen in the analysis of the shift commutator. We begin with this familiar case. However, now without using the sharper (5.5) expansion since, in this case, it does not matter if we have a square root dependence or a linear one. Observe that

$$|\langle b\rangle_{I_1^1\times K^2}-\langle b\rangle_{I_{n+1}^1\times K^2}|\lesssim \|b\|_{\mathrm{bmo}(\nu)}(\nu_{I_1^1,K^2}+\nu_{I_{n+1}^1,K^2}),$$

where

$$\nu_{Q_1,K^2}:=\sum_{\substack{J^1\in\mathcal{D}^1\\ Q^1\subsetneq J^1\subset K^1}}\langle \nu\rangle_{J^1\times K^2},\quad Q^1\in\{I_1^1,I_{n+1}^1\}.$$

Then our term is bounded by

$$\|b\|_{\mathrm{bmo}(\nu)}\sum_K\sum_{(I_i^1)^{(k_i)}=K^1}|a_{K,(I_i^1)}||\langle f_1,h_{I_1^1}\otimes h_{K^2}\rangle|\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big| \\ \times(\nu_{I_1^1,K^2}+\nu_{I_{n+1}^1,K^2}).$$

We first consider $\nu_{I_{n+1}^1,K^2}$. We fix $j_{n+1}\in\{1,\dots,k_{n+1}\}$ and it suffices to bound

$$\sum_{K^1}\sum_{(I_i^1)^{(k_i)}=K^1}\sum_{K^2}|a_{K,(I_i^1)}|\langle\nu\rangle_{(I_{n+1}^1)^{(j_{n+1})}\times K^2}|\langle f_1,h_{I_1^1}\otimes h_{K^2}\rangle|\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big| \\ \lesssim\sum_{K^1}\sum_{(I_i^1)^{(k_i)}=K^1}\frac{\prod_{i=1}^{n+1}|I_i^1|^{\frac12}}{|K^1|^n}\int_{\mathbb{R}^{d_1}}\frac{1_{(I_{n+1}^1)^{(j_{n+1})}}}{|(I_{n+1}^1)^{(j_{n+1})}|} \\ \times\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}|\langle f_1,h_{I_1^1}\otimes h_{K^2}\rangle|^2\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac12}\nu,$$

where we have applied Lemma 3.2. Recall the strategy in [21], when $\widetilde{h}_{I_i^1}=h_{I_i^1}$ we do not do anything and when $\widetilde{h}_{I_i^1}=h_{I_i^1}^0$ and $I_i^1\neq K^1$ we expand

$$|I_j^1|^{-\frac12}\langle f_j,h_{I_j^1}^0\rangle_1=\langle f\rangle_{I_j^1,1}=\langle f_j\rangle_{K^1,1}+\sum_{i_j=1}^{k_j}\langle\Delta^1_{(I_j^1)^{(i_j)}}f_j\rangle_{(I_j^1)^{(i_j-1)},1}.$$

We have

$$\sum_{K^1}\sum_{(I_i^1)^{(k_i)}=K^1}\frac{\prod_{i=1}^{n+1}|I_i^1|^{\frac12}}{|K^1|^n}\int_{\mathbb{R}^{d_1}}\frac{1_{(I_{n+1}^1)^{(j_{n+1})}}}{|(I_{n+1}^1)^{(j_{n+1})}|}|I_{n+1}^1|^{\frac12} \\ \times\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}|\langle f_1,h_{I_1^1}\otimes h_{K^2}\rangle|^2\prod_{i=2}^{n}\Big|\Big\langle f_i,\widetilde{h}_{I_i^1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big|^2|\langle f_{n+1}\rangle_{K^1\times K^2}|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac12}\nu \\ \le\sum_{K^1}\sum_{\substack{(I_i^1)^{(k_i)}=K^1\\ i\neq n+1}}\frac{\prod_{i=1}^{n}|I_i^1|^{\frac12}}{|K^1|^n}\int_{\mathbb{R}^d}1_{K^1}$$

$$\times \Big(\sum_{K^2} |\langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle|^2 \prod_{i=2}^n \Big| \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big|^2 \langle |f_{n+1}| \rangle^2_{K^1 \times K^2} \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \nu.$$

Since $(w_1, \cdots, w_n, \nu w^{-1}) \in A_{(p_1, \cdots, p_n, p')}$, the same proof as in [21, Section 6.B] yields the desired estimate. The proof of $\langle \Delta^1_{(I^1_{n+1})^{(i_{n+1})}} f_{n+1} \rangle_{(I^1_{n+1})^{(i_{n+1}-1)},1}$ with $i_{n+1} \ge j_{n+1}$ is similar. So we only focus on $i_{n+1} < j_{n+1}$. By simple calculus, we reduce to bounding

$$\sum_{K^1} \sum_{\substack{(I_i^1)^{(k_i)} = K^1 \\ i \ne n+1}} \frac{\prod_{i=1}^n |I_i^1|^{\frac{1}{2}}}{|K^1|^n} \int_{\mathbb{R}^{d_1}} \frac{1_{(L^1_{n+1})^{(j_{n+1}-i_{n+1})}}}{|(L^1_{n+1})^{(j_{n+1}-i_{n+1})}|} |L^1_{n+1}|^{\frac{1}{2}}$$
$$\times \int_{\mathbb{R}^{d_2}} \Big(\sum_{K^2} |\langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle|^2 \prod_{i=2}^n \Big| \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big|^2 \Big| \Big\langle f_{n+1}, h_{L^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big|^2 \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \nu.$$

Denote $(L^1_{n+1})^{(j_{n+1}-i_{n+1})} = Q^1_{n+1}$, and write

$$\Big\langle f_{n+1}, h_{L^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle = \frac{\Big\langle f_{n+1}, h_{L^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle}{\langle \sigma_{n+1} \rangle_{Q^1_{n+1} \times K^2}} \langle \sigma_{n+1} \rangle_{Q^1_{n+1} \times K^2}.$$

By the reverse Hölder and A_∞ extrapolation, we can get σ_{n+1} out of the square sum. Then using

$$\frac{\Big| \Big\langle f_{n+1}, h_{L^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big|}{\langle \sigma_{n+1} \rangle_{Q^1_{n+1} \times K^2}} \le M_{\mathcal{D}^2}^{\langle \sigma_{n+1} \rangle_{Q^1_{n+1},1}} (\langle f_{n+1}, h_{L^1_{n+1}} \rangle_1 \langle \sigma_{n+1} \rangle^{-1}_{Q^1_{n+1},1}) 1_{K^2}$$

we arrive at

$$\sum_{K^1} \sum_{\substack{(I_i^1)^{(k_i)} = K^1 \\ i \ne n+1}} \frac{\prod_{i=1}^n |I_i^1|^{\frac{1}{2}}}{|K^1|^n} \int_{\mathbb{R}^d} 1_{K^1} F_{n+1,K^1}$$
$$\times \Big(\sum_{K^2} |\langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle|^2 \prod_{i=2}^n \Big| \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big|^2 \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \nu \sigma_{n+1},$$

where

$$F_{n+1,K^1} := \sum_{(Q^1_{n+1})^{(k_{n+1}-j_{n+1})} = K^1} \frac{1_{Q^1_{n+1}}}{|Q^1_{n+1}|} \sum_{(L^1_{n+1})^{(j_{n+1}-i_{n+1})} = Q^1_{n+1}} |L^1_{n+1}|^{\frac{1}{2}}$$
$$\times M_{\mathcal{D}^2}^{\langle \sigma_{n+1} \rangle_{Q^1_{n+1},1}} (\langle f_{n+1}, h_{L^1_{n+1}} \rangle_1 \langle \sigma_{n+1} \rangle^{-1}_{Q^1_{n+1},1}).$$

By Hölder's inequality, it suffices to bound the $L^p(w^p)$ norm of

$$\Big(\sum_{K^1} 1_{K^1} \Big[\sum_{\substack{(I_i^1)^{(k_i)} = K^1 \\ i \ne n+1}} \frac{\prod_{i=1}^n |I_i^1|^{\frac{1}{2}}}{|K^1|^n} \Big(\sum_{K^2} |\langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle|^2 \prod_{i=2}^n \Big| \Big\langle f_i, \widetilde{h}_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big|^2 \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \Big]^2 \Big)^{\frac{1}{2}}$$

and

$$\Big\|\Big(\sum_{K^1} F^2_{n+1,K^1}\Big)^{\frac12}\Big\|_{L^{p'}(\sigma_{n+1})}.$$

Simply control the outer ℓ^2 norm by ℓ^1 norm – then we can again use the estimate in [21, Section 6.B] to conclude the first term. For the second term, note that

$$\begin{aligned}
&\Big\|\Big(\sum_{K^1} F^2_{n+1,K^1}\Big)^{\frac12}\Big\|_{L^{p'}(\sigma_{n+1})}\\
&=\Big\|\Big(\sum_{K^1}\Big[\sum_{(Q^1_{n+1})^{(k_{n+1}-j_{n+1})}=K^1}\frac{1_{Q^1_{n+1}}}{|Q^1_{n+1}|}\sum_{(L^1_{n+1})^{(j_{n+1}-i_{n+1})}=Q^1_{n+1}}|L^1_{n+1}|^{\frac12}\\
&\qquad\times M_{\mathcal{D}^2}^{\langle\sigma_{n+1}\rangle_{Q^1_{n+1},1}}(\langle f_{n+1},h_{L^1_{n+1}}\rangle_1\langle\sigma_{n+1}\rangle^{-1}_{Q^1_{n+1},1})\Big]^2\Big)^{\frac12}\Big\|_{L^{p'}(\sigma_{n+1})}\\
&=\Big\|\Big(\sum_{Q^1_{n+1}}1_{Q^1_{n+1}}\Big[\sum_{(L^1_{n+1})^{(j_{n+1}-i_{n+1})}=Q^1_{n+1}}\frac{|L^1_{n+1}|^{\frac12}}{|Q^1_{n+1}|}\\
&\qquad\times M_{\mathcal{D}^2}^{\langle\sigma_{n+1}\rangle_{Q^1_{n+1},1}}(\langle f_{n+1},h_{L^1_{n+1}}\rangle_1\langle\sigma_{n+1}\rangle^{-1}_{Q^1_{n+1},1})\Big]^2\Big)^{\frac12}\Big\|_{L^{p'}(\sigma_{n+1})},
\end{aligned}$$

which again can be handled exactly as in [21, p.23]. Now we turn to consider the case $Q^1=I^1_1$. Similarly,

$$\begin{aligned}
&\sum_{K^1}\sum_{(I^1_i)^{(k_i)}=K^1}\sum_{K^2}|a_{K,(I^1_i)}||\langle\nu\rangle_{(I^1_1)^{(j_1)}\times K^2}|\langle f_1,h_{I^1_1}\otimes h_{K^2}\rangle|\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big|\\
&\lesssim\sum_{K^1}\sum_{(I^1_i)^{(k_i)}=K^1}\frac{\prod_{i=1}^{n+1}|I^1_i|^{\frac12}}{|K^1|^n}\int_{\mathbb{R}^{d_1}}\frac{1_{(I^1_1)^{(j_1)}}}{|(I^1_1)^{(j_1)}|}\\
&\qquad\times\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}|\langle f_1,h_{I^1_1}\otimes h_{K^2}\rangle|^2\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac12}\nu.
\end{aligned}$$

Similar as [21], we may without loss of generality assume either $\widetilde{h}_{I^1_i}=h_{I^1_i}$ or otherwise $I^1_i=K^1$. As before, by reverse Hölder and A_∞ extrapolation the object is dominated by

$$\begin{aligned}
&\sum_{K^1}\sum_{(I^1_i)^{(k_i)}=K^1}\frac{\prod_{i=1}^{n+1}|I^1_i|^{\frac12}}{|K^1|^n}\int_{\mathbb{R}^{d_1}}\frac{1_{(I^1_1)^{(j_1)}}}{|(I^1_1)^{(j_1)}|}\\
&\qquad\times\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}\frac{|\langle f_1,h_{I^1_1}\otimes h_{K^2}\rangle|^2}{\langle\sigma_1\rangle^2_{(I^1_1)^{(j_1)}\times K^2}}\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac12}\nu\sigma_1.
\end{aligned}$$

Next, we write

$$\prod_{i=2}^{n+1}\Big|\Big\langle f_i,\widetilde{h}_{I^1_i}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle\Big|\le M_{\mathcal{D}^2}(\langle f_2,\widetilde{h}_{I^1_2}\rangle_1,\cdots,\langle f_{n+1},\widetilde{h}_{I^1_{n+1}}\rangle_1).$$

Then by Hölder's inequality the estimate is reduced to

$$A := \Big\| \Big(\sum_{K^1} \Big(\sum_{(I_1^1)^{(k_1)} = K^1} |I_1^1|^{\frac{1}{2}} \frac{1_{(I_1^1)^{(j_1)}}}{|(I_1^1)^{(j_1)}|} \Big(\sum_{K^2} \frac{|\langle f_1, h_{I_1^1} \otimes h_{K^2} \rangle|^2}{\langle \sigma_1 \rangle^2_{(I_1^1)^{(j_1)} \times K^2}} \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \Big)^2 \Big)^{\frac{1}{2}} \Big\|_{L^{p_1}(\sigma_1)}$$

and

$$B := \Big\| \Big(\sum_{K^1} 1_{K^1} \Big[\sum_{\substack{(I_i^1)^{(k_i)} = K^1 \\ i \neq 1}} \frac{\prod_{i=2}^{n+1} |I_i^1|^{\frac{1}{2}}}{|K^1|^n} M_{\mathcal{D}^2}(\langle f_2, \widetilde{h}_{I_2^1} \rangle_1, \cdots, \langle f_{n+1}, \widetilde{h}_{I_{n+1}^1} \rangle_1) \Big]^2 \Big)^{\frac{1}{2}} \Big\|_{L^{p_1'}(\eta_1)}.$$

Again, the estimate of A can be found in [21, Section 6.B] and we omit the details. For B, we shall prove

$$B \lesssim \|f_{n+1} \nu w^{-1}\|_{L^{p'}} \prod_{i=2}^n \|f_i w_i\|_{L^{p_i}}.$$

By the extrapolation theorem, it suffices to prove

$$\Big\| \Big(\sum_{K^1} 1_{K^1} \Big[\sum_{\substack{(I_i^1)^{(k_i)} = K^1 \\ i \neq 1}} \frac{\prod_{i=2}^{n+1} |I_i^1|^{\frac{1}{2}}}{|K^1|^n} M_{\mathcal{D}^2}(\langle f_2, \widetilde{h}_{I_2^1} \rangle_1, \cdots, \langle f_{n+1}, \widetilde{h}_{I_{n+1}^1} \rangle_1) \Big]^2 \Big)^{\frac{1}{2}} v \Big\|_{L^{\frac{2}{n}}} \le \prod_{i=2}^{n+1} \|f_i v_i\|_{L^2},$$

provided $(v_2, \cdots, v_{n+1}) \in A_{(2,\cdots,2)}$ and $v = \prod_{i=2}^{n+1} v_i$. Note that for a fixed K^2, if we denote $\zeta_i = v_i^{-2}$, $2 \le i \le n+1$, then

$$\begin{aligned} \prod_{i=2}^{n+1} \Big\langle |\langle f_i, \widetilde{h}_{I_i^1} \rangle_1| \Big\rangle_{K^2} &= \prod_{i=2}^{n+1} \frac{\Big\langle |\langle f_i, \widetilde{h}_{I_i^1} \rangle_1| \Big\rangle_{K^2}}{\langle \zeta_i \rangle_{K^1 \times K^2}} \langle \zeta_i \rangle_{K^1 \times K^2} \\ &\lesssim \frac{1}{\langle v^{\frac{2}{n}} \rangle^n_{K^1 \times K^2}} \prod_{i=2}^{n+1} \frac{\Big\langle |\langle f_i, \widetilde{h}_{I_i^1} \rangle_1| \Big\rangle_{K^2}}{\langle \zeta_i \rangle_{K^1 \times K^2}} \\ &\le \inf_{x \in K^1 \times K^2} \Big(M_{\mathcal{D}}^{v^{\frac{2}{n}}} \big(\big[\prod_{i=2}^{n+1} M_{\mathcal{D}^2}^{\langle \zeta_i \rangle_{K^1,1}} (|\langle f_i, \widetilde{h}_{I_i^1} \rangle_1| \langle \zeta_i \rangle^{-1}_{K^1,1}) 1_{K^1} \big]^{\frac{1}{n}} v^{-\frac{2}{n}} \big) \Big)^n. \end{aligned}$$

Whence

$$\begin{aligned} &1_{K^1} M_{\mathcal{D}^2}(\langle f_2, \widetilde{h}_{I_2^1} \rangle_1, \cdots, \langle f_{n+1}, \widetilde{h}_{I_{n+1}^1} \rangle_1) \\ &\qquad \le \Big(M_{\mathcal{D}}^{v^{\frac{2}{n}}} \big(\big[\prod_{i=2}^{n+1} M_{\mathcal{D}^2}^{\langle \zeta_i \rangle_{K^1,1}} (|\langle f_i, \widetilde{h}_{I_i^1} \rangle_1| \langle \zeta_i \rangle^{-1}_{K^1,1}) 1_{K^1} \big]^{\frac{1}{n}} v^{-\frac{2}{n}} \big) \Big)^n \end{aligned}$$

and by the vector-valued estimate for $M_{\mathcal{D}}^{v^{\frac{2}{n}}}$ and Hölder's inequality, we have

$$\Big\| \Big(\sum_{K^1} 1_{K^1} \Big[\sum_{\substack{(I_i^1)^{(k_i)} = K^1 \\ i \neq 1}} \frac{\prod_{i=2}^{n+1} |I_i^1|^{\frac{1}{2}}}{|K^1|^n} M_{\mathcal{D}^2}(\langle f_2, \widetilde{h}_{I_2^1} \rangle_1, \cdots, \langle f_{n+1}, \widetilde{h}_{I_{n+1}^1} \rangle_1) \Big]^2 \Big)^{\frac{1}{2}} v \Big\|_{L^{\frac{2}{n}}}$$

$$\lesssim \prod_{i=2}^{n+1} \Big\| \Big(\sum_{K^1} 1_{K^1} \Big[\sum_{(I_i^1)^{(k_i)}=K^1} \frac{|I_i|^{\frac12}}{|K^1|} M_{\mathcal{D}^2}^{\langle \zeta_i\rangle_{K^1,1}} (|\langle f_i, \widetilde{h}_{I_i^1}\rangle_1| \langle \zeta_i\rangle^{-1}_{K^1,1}) \Big]^2 \Big)^{\frac12} \Big\|_{L^2(\zeta_i)}.$$

Recall that when $\widetilde{h}_{I_i^1} = h^0_{I_i^1}$, then according to our convention $I_i^1 = K^1$ and

$$\sum_{(I_i^1)^{(k_i)}=K^1} \frac{|I_i|^{\frac12}}{|K^1|} M_{\mathcal{D}^2}^{\langle \zeta_i\rangle_{K^1,1}} (|\langle f_i, \widetilde{h}_{I_i^1}\rangle_1| \langle \zeta_i\rangle^{-1}_{K^1,1}) = M_{\mathcal{D}^2}^{\langle \zeta_i\rangle_{K^1,1}} (|\langle f_i\rangle_{K^1}| \langle \zeta_i\rangle^{-1}_{K^1,1}) \le M_{\mathcal{D}}^{\zeta_i}(f_i \zeta_i^{-1}).$$

Again the rest can be estimated as in [21, Section 6.B].

Next, we consider the latter part of (5.7). Notice that by Lemma 3.3 we have

$$\begin{aligned}
&|I^1_{n+1}|^{\frac12} |K^2| \Big\langle (b - \langle b\rangle_{I^1_{n+1}\times K^2}) f_{n+1}, h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \\
&= \sum_{I\times J \subset I^1_{n+1}\times K^2} \langle b, h_I\otimes h_J\rangle \langle f_{n+1}, h_I \otimes h_J\rangle + \sum_{I \subset I^1_{n+1}} \Big\langle b, h_I \otimes \frac{1_{K^2}}{|K^2|}\Big\rangle \langle f_{n+1}, h_I \otimes 1_{K^2}\rangle \\
&\quad + \sum_{J\subset K^2} \Big\langle b, \frac{1_{I^1_{n+1}}}{|I^1_{n+1}|} \otimes h_J \Big\rangle \langle f_{n+1}, 1_{I^1_{n+1}} \otimes h_J\rangle \\
&\lesssim \|b\|_{\mathrm{bmo}(\nu)} \int_{I^1_{n+1}\times K^2} \Big(\sum_{R\in\mathcal{D}} \frac{\langle f_{n+1}, h_R\rangle^2}{\langle \sigma_{n+1}\rangle_R^2} \frac{1_R}{|R|}\Big)^{\frac12} \nu\sigma_{n+1} \\
&\qquad + \|b\|_{\mathrm{bmo}(\nu)} \int_{I^1_{n+1}\times K^2} \Big(\sum_{I\in\mathcal{D}^1} \frac{\langle f_{n+1}, h_I \otimes \frac{1_{K^2}}{|K^2|}\rangle^2}{\langle \sigma_{n+1}\rangle^2_{I\times K^2}} \frac{1_I}{|I|}\Big)^{\frac12} \nu\sigma_{n+1} \\
&\qquad\quad + \|b\|_{\mathrm{bmo}(\nu)} \int_{I^1_{n+1}\times K^2} \Big(\sum_{J\in\mathcal{D}^2} \frac{\langle f_{n+1}, \frac{1_{I^1_{n+1}}}{|I^1_{n+1}|} \otimes h_J\rangle^2}{\langle \sigma_{n+1}\rangle^2_{I^1_{n+1}\times J^2}} \frac{1_J}{|J|}\Big)^{\frac12} \nu\sigma_{n+1}.
\end{aligned}$$

Then e.g. dominating

$$\Big(\sum_{I\in\mathcal{D}^1} \frac{\langle f_{n+1}, h_I \otimes \frac{1_{K^2}}{|K^2|}\rangle^2}{\langle \sigma_{n+1}\rangle^2_{I\times K^2}} \frac{1_I}{|I|}\Big)^{\frac12} \le \Big(\sum_{I\in\mathcal{D}^1} \big[M_{\mathcal{D}^2}^{\langle \sigma_{n+1}\rangle_{I,1}} (\langle f_{n+1}, h_I\rangle \langle \sigma_{n+1}\rangle^{-1}_{I,1})\big]^2 \frac{1_I}{|I|}\Big)^{\frac12}$$

allows us to view these square functions (which are bounded on $L^{p'}(\sigma_{n+1})$) as the new f_{n+1}. So that by Hölder's inequality, the related term in the commutator boils down to estimating the partial paraproduct

$$\Big\| \sum_{K^1,K^2} \sum_{(I_j^1)^{(k_j)}=K^1} a_{K(I_i^1)} \langle f_1, h_{I_1^1} \otimes h_{K^2}\rangle \prod_{i=2}^n \Big\langle f_i, h_{I_i^1} \otimes \frac{1_{K^2}}{|K^2|}\Big\rangle h^0_{I^1_{n+1}} \otimes \frac{1_{K^2}}{|K^2|} w \Big\|_{L^p},$$

which is exactly the standard one.

Following the expansion methods and estimations introduced earlier, we can handle the other forms of commutators similarly. Compared to the shift case, the more difficult challenges arise from the terms of forms, where we have

$$(\langle b\rangle_{I_1^1\times K^2} - \langle b\rangle_K) \langle f_1, h_{I_1^1} \otimes h_{K^2}\rangle,$$

$$(\langle b\rangle_{I^1_1\times K^2}-\langle b\rangle_K)\Big\langle f_1, h_{I^1_1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle,$$

$$(\langle b\rangle_{I^1_1\times K^2}-\langle b\rangle_K)\langle f_1, h^0_{I^1_1}\otimes h_{K^2}\rangle,$$

and

$$(\langle b\rangle_{I^1_1\times K^2}-\langle b\rangle_K)\Big\langle f_1, h^0_{I^1_1}\otimes\frac{1_{K^2}}{|K^2|}\Big\rangle.$$

We already handled the first and the symmetric case of the last one. By modifying the above methods, we can estimate the other two terms.

Full paraproducts. Although the full paraproducts have the more complicated product BMO coefficients, they do not require as much analysis as the partial paraproducts. Since no unseen methods are needed to conclude the boundedness of full paraproduct commutators, we omit the details.

6. The lower bound

Let K be a standard bi-parameter full kernel as described earlier. In this section, we additionally assume that K is a multilinear non-degenerate kernel. That is, for any given rectangle $R=I^1\times I^2$ there exists $\widetilde{R}=\widetilde{I}^1\times\widetilde{I}^2$ such that $\ell(I^i)=\ell(\widetilde{I}^i)$, $d(I^i,\widetilde{I}^i)\sim\ell(I^i)$, and there exists some $\zeta\in\mathbb{C}$ with $|\zeta|=1$ such that for all $x\in\widetilde{R}$ and $y_1,\dots,y_n\in R$ there holds

$$\operatorname{Re}\zeta K(x,y_1,\dots,y_n)\gtrsim\frac{1}{|R|^n}.$$

We are going to assume the weak type boundedness of the commutator. Suppose that

$$\sup_{A\subset R}\frac{1}{\prod_{i=1}^n\sigma_i(R)^{\frac{1}{p_i}}}\Big\|1_{\widetilde{R}}[b,T]_j(1_R\sigma_1,\dots,1_A\sigma_j,\dots,1_R\sigma_n)\nu^{-1}w\Big\|_{L^{p,\infty}}<\infty,$$

where recall that $\sigma_i=w_i^{-p_i'}$ and $\nu=\lambda_j^{-1}w_j$. Clearly, this is a weaker assumption than

$$\Big\|[b,T]_j\colon\prod_{i=1}^n L^{p_i}(w_i^{p_i})\to L^p(\nu^{-p}w^p)\Big\|<\infty.$$

We do not assume the two separate $A_{\vec{p}}$ conditions here. It is enough to assume that we have the two tuples $(w_1,\dots,w_n),(w_1,\dots,\lambda_j,\dots,w_n)$ of weights satisfying

$$(w_1,\dots,w_n,\nu w^{-1})\in A^*_{\vec{p}}.$$

Let us denote $\nu^{-p}w^p$ by σ_{n+1}.

We employ the idea of the median method to prove that

$$b\in\operatorname{bmo}_\nu(\sigma_j):=\{b\in L^1_{\text{loc}}\colon\sup_{R\in\mathcal{D}}\inf_{c\in\mathbb{R}}\frac{1}{\nu\sigma_j(R)}\int_R|b-c|\sigma_j<\infty\}$$

under the weaker assumption above. We additionally need to assume that $\nu\sigma_j\in A_\infty$ since when $\nu,\sigma_j,\nu\sigma_j\in A_\infty$ it follows that this is equivalent with the Bloom type little BMO definition, see the Appendix.

6.1. *Remark.* We get $\nu\sigma_j\in A_\infty$ for free whenever $\lambda_j^{-p_j'}\in A_\infty$ since

$$\nu\sigma_j=\lambda_j^{-1}w_j^{1-p_j'}=(\lambda_j^{-p_j'})^{\frac{1}{p_j'}}(\sigma_j)^{\frac{1}{p_j}}\in A_\infty.$$

Fix rectangle $R \in \mathcal{D}$. We take arbitrary $\alpha \in \mathbb{R}$ and $x \in \widetilde{R} \cap \{b \geq \alpha\}$, where $\widetilde{R}$ is a rectangle that satisfies the non-degeneracy property. Thus, we have

$$\frac{1}{|R|}\int_R (\alpha - b)_+ \sigma_j \prod_{\substack{i=1\\ i\neq j}}^n \frac{\sigma_i(R)}{|R|}$$
$$\lesssim \operatorname{Re}\zeta \int_{R\cap\{b\leq\alpha\}} \int_R \cdots \int_R (b(x)-b(y_j))K(x,y_1,\ldots,y_n)\prod_{i=1}^n \sigma_i(y_i)\,\mathrm{d}y_i.$$

We let α be the median of b on $\widetilde{R}$, i.e.

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

As $\sigma_{n+1} \in A_\infty$ we have that $\sigma_{n+1}(\widetilde{R}\cap\{b\geq\alpha\}) \sim \sigma_{n+1}(\widetilde{R}) \sim \sigma_{n+1}(R)$.

Thus, we get

$$\sigma_{n+1}(R)^{\frac{1}{p}}\frac{1}{|R|}\int_R(\alpha-b)_+\sigma_j\prod_{\substack{i=1\\ i\neq j}}^n\frac{\sigma_i(R)}{|R|} \lesssim \|C_b^K(\sigma_1,\ldots,\sigma_n)\|_{L^{p,\infty}(\sigma_{n+1})} \lesssim \prod_{i=1}^n \sigma_i(R)^{\frac{1}{p_i}}, \tag{6.2}$$

where

$$C_b^K(\sigma_1,\ldots,\sigma_n)(x)$$
$$:= 1_{\widetilde{R}\cap\{b\geq\alpha\}}(x)\operatorname{Re}\zeta\int_{R\cap\{b\leq\alpha\}}\int_R\cdots\int_R(b(x)-b(y_j))K(x,y_1,\ldots,y_n)\prod_{i=1}^n\sigma_i(y_i)\,\mathrm{d}y_i.$$

Recall that

$$1 \leq \prod_{i=1}^n \langle\sigma_i\rangle_R^{\frac{1}{p_i'}}\langle\sigma_{n+1}\rangle_R^{\frac{1}{p}}\langle\nu\rangle_R \leq [(w_1,\ldots,w_n,\nu w^{-1})]_{A^*_{\vec{p}}} < \infty.$$

Rearranging terms in (6.2) and using the observation, we get

$$\frac{1}{\langle\nu\rangle_R\sigma_j(R)}\int_R(\alpha-b)_+\sigma_j \lesssim 1.$$

By the reverse Hölder property, we have

$$\frac{1}{\langle\nu\rangle_R\sigma_j(R)} \gtrsim \frac{1}{\nu\sigma_j(R)}.$$

By symmetrical estimates, we also get

$$\frac{1}{\nu\sigma_j(R)}\int_R(b-\alpha)_+\sigma_j \lesssim 1.$$

This completes the proof.

7. Two-weight extrapolation

This section is devoted to proving Theorem 1.2.

The strategy of the proof will be similar as in [18] and [19]. We only prove the case

$$q_n \neq p_n, 1 < q_n \leq \infty, q_i = p_i \text{ for all } 2 \leq i \leq n-1.$$

Let us first recall the following lemma, whose proof can be found in [19, Lemma 2.14].

7.1. **Lemma.** *Let* $w_i^{\frac{1}{n-1+\frac{1}{p_i}}} \in A_{\frac{n}{n-1+\frac{1}{p_i}}}$, $1 \leq i \leq n-1$. *Let* $\widehat{w} = (\prod_{i=1}^{n-1} w_i)^\rho \in A_{n\rho}$, *where* $\rho = (1 + \sum_{i=1}^{n-1} \frac{1}{p_i})^{-1}$. *Then* $(w_1, \cdots, w_n) \in A_{\vec{p}}$ *if and only if*

$$W := w_n \widehat{w}^{\frac{1}{p_n'}} \in A_{p_n, p}(\widehat{w}).$$

Note that it is also recorded in [19, Lemma 2.14] that if $(w_1, \cdots, w_n) \in A_{\vec{p}}$, then we always have

$$\widehat{w} = (\prod_{i=1}^{n-1} w_i)^\rho \in A_{n\rho}, \qquad w_i^{\frac{1}{n-1+\frac{1}{p_i}}} \in A_{\frac{n}{n-1+\frac{1}{p_i}}}, \quad i = 1, \cdots, n-1.$$

With this at hand, since we have

$$(w_1, w_2, \cdots, w_n) \in A_{(p_1, \cdots, p_{n-1}, q_n)}, \quad (\lambda_1, w_2, \cdots, w_n) \in A_{(p_1, \cdots, p_{n-1}, q_n)},$$

recalling that

$$\frac{1}{q} = \frac{1}{p_1} + \cdots + \frac{1}{p_{n-1}} + \frac{1}{q_n},$$

we have

$$\widehat{w} = (\prod_{i=1}^{n-1} w_i)^\rho \in A_{n\rho}, \quad \widehat{\lambda} = (\lambda_1 \prod_{i=2}^{n-1} w_i)^\rho \in A_{n\rho}$$

and

$$W_w = w_n \widehat{w}^{\frac{1}{q_n'}} \in A_{q_n, q}(\widehat{w}), \quad W_\lambda = w_n \widehat{\lambda}^{\frac{1}{q_n'}} \in A_{q_n, q}(\widehat{\lambda}).$$

Then the goal is to prove

$$\|f \lambda_1 w_1^{-1} W_w\|_{L^q(\widehat{w})} \lesssim \|f_n \widehat{w}^{-1} W_w\|_{L^{q_n}(\widehat{w})} \prod_{i=1}^{n-1} \|f_i w_i\|_{L^{p_i}},$$

which can also be written as

$$\|f W_\lambda\|_{L^q(\widehat{\lambda})} \lesssim \|f_n \widehat{\lambda}^{-1} W_\lambda\|_{L^{q_n}(\widehat{\lambda})} \prod_{i=1}^{n-1} \|f_i w_i\|_{L^{p_i}}.$$

We split the proof to the following cases:

Case 1: $1/s := 1/q - 1/p = 1/q_n - 1/p_n > 0$. Without loss of generality we may assume

$$0 < \|f_n w_n\|_{L^{q_n}} = \|f_n \widehat{w}^{-1} W_w\|_{L^{q_n}(\widehat{w})} = \|f_n \widehat{\lambda}^{-1} W_\lambda\|_{L^{q_n}(\widehat{\lambda})} < \infty.$$

Let

$$h = \frac{f_n w_n^{q_n'}}{\|f_n w_n\|_{L^{q_n}}} = \frac{f_n \widehat{w}^{-1} W_w^{q_n'}}{\|f_n \widehat{w}^{-1} W_w\|_{L^{q_n}(\widehat{w})}} = \frac{f_n \widehat{\lambda}^{-1} W_\lambda^{q_n'}}{\|f_n \widehat{\lambda}^{-1} W_\lambda\|_{L^{q_n}(\widehat{\lambda})}},$$

so that we have $\|h\|_{L^{q_n}(w_n^{-q_n'})} = 1$. Define

$$\mathcal{R}'h = \sum_{k=0}^{\infty} \frac{(M'_{\widehat{\lambda}} M'_{\widehat{w}})^{(k)} h}{2^k \|M'_{\widehat{\lambda}} M'_{\widehat{w}}\|^k_{L^{(1+\frac{q_n'}{q})'}(w_n^{-q_n'})}} =: \sum_{k=0}^{\infty} \frac{(M'_{\widehat{\lambda}} M'_{\widehat{w}})^{(k)} h}{2^k \|M'_{\widehat{\lambda}} M'_{\widehat{w}}\|^k},$$

where

$$M'_{\widehat{w}} g = M_{\widehat{w}}(g W_w^{-q_n'}) W_w^{q_n'}, \qquad M'_{\widehat{\lambda}} g = M_{\widehat{\lambda}}(g W_\lambda^{-q_n'}) W_\lambda^{q_n'}.$$

Let us explain why $\mathcal{R}'$ is well-defined. Indeed, since $W_w \in A_{q_n,q}(\widehat{w})$, we have $W_w^{-q_n'} \in A_{1+\frac{q_n'}{q}}(\widehat{w})$ and $M'_{\widehat{w}}$ is bounded on $L^{(1+\frac{q_n'}{q})'}(W_w^{-q_n'}\widehat{w}) = L^{(1+\frac{q_n'}{q})'}(w_n^{-q_n'})$ (see [21, Lemma 8.2]). Likewise $M'_{\widehat{\lambda}}$ is bounded on $L^{(1+\frac{q_n'}{q})'}(w_n^{-q_n'})$. Now set

$$H = \mathcal{R}'(h^{\frac{q_n}{(1+\frac{q_n'}{q})'}})^{\frac{(1+\frac{q_n'}{q})'}{q_n}}.$$

Then the above discussion easily yields

$$h \le H, \quad \|H\|_{L^{q_n}(w_n^{-q_n'})} \lesssim \|h\|_{L^{q_n}(w_n^{-q_n'})} = 1$$

and

$$M'_{\widehat{w}}\Big(H^{\frac{q_n}{(1+\frac{q_n'}{q})'}}\Big) \le M'_{\widehat{\lambda}} M'_{\widehat{w}}\Big(H^{\frac{q_n}{(1+\frac{q_n'}{q})'}}\Big) \le 2\|M'_{\widehat{\lambda}} M'_{\widehat{w}}\| H^{\frac{q_n}{(1+\frac{q_n'}{q})'}}$$

$$M'_{\widehat{\lambda}}\Big(H^{\frac{q_n}{(1+\frac{q_n'}{q})'}}\Big) \le M'_{\widehat{\lambda}} M'_{\widehat{w}}\Big(H^{\frac{q_n}{(1+\frac{q_n'}{q})'}}\Big) \le 2\|M'_{\widehat{\lambda}} M'_{\widehat{w}}\| H^{\frac{q_n}{(1+\frac{q_n'}{q})'}},$$

which give that

$$[H^{\frac{q_n}{(1+\frac{q_n'}{q})'}} W_w^{-q_n'}]_{A_1(\widehat{w})} \le 2\|M'_{\widehat{\lambda}} M'_{\widehat{w}}\| \quad \text{and} \quad [H^{\frac{q_n}{(1+\frac{q_n'}{q})'}} W_\lambda^{-q_n'}]_{A_1(\widehat{\lambda})} \le 2\|M'_{\widehat{\lambda}} M'_{\widehat{w}}\|.$$

Finally, set $v_n = H^{-\frac{q_n}{s}} w_n^{1+\frac{q_n'}{s}}$. It remains to check

(7.2) $$(w_1, \cdots, w_{n-1}, v_n), (\lambda_1, \cdots, w_{n-1}, v_n) \in A_{\vec{p}}.$$

Equivalently, we check

$$v_n \widehat{w}^{\frac{1}{p_n'}} \in A_{p_n,p}(\widehat{w}) \quad \text{and} \quad v_n \widehat{\lambda}^{\frac{1}{p_n'}} \in A_{p_n,p}(\widehat{\lambda}),$$

which will be completely similar as that in [19, p. 106]. Indeed, once we have (7.2), then

$$\|f W_\lambda\|_{L^q(\widehat{\lambda})} = \|f v_n \widehat{\lambda}^{\frac{1}{q}+\frac{1}{q_n'}} H^{\frac{q_n}{s}} w_n^{-\frac{q_n'}{s}}\|_{L^q} \le \|f v_n \widehat{\lambda}^{\frac{1}{q}+\frac{1}{q_n'}}\|_{L^p} \|H^{\frac{q_n}{s}} w_n^{-\frac{q_n'}{s}}\|_{L^s}$$

$$\lesssim \|f v_n \widehat{\lambda}^{\frac{1}{q}+\frac{1}{q_n'}}\|_{L^p} = \|f v_n \widehat{\lambda}^{\frac{1}{p}+\frac{1}{p_n'}}\|_{L^p} \lesssim \|f_n v_n\|_{L^{p_n}} \prod_{i=1}^{n-1} \|f_i w_i\|_{L^{p_i}}.$$

The proof is completed by noticing that

$$\|f_n v_n\|_{L^{p_n}} = \big\| h w_n^{-q_n'} \|f_n w_n\|_{L^{q_n}} v_n \big\|_{L^{p_n}} \le \|f_n w_n\|_{L^{q_n}} \|H^{1-\frac{q_n}{s}} w_n^{-q_n'+1+\frac{q_n'}{s}}\|_{L^{p_n}}$$

$$= \|f_n w_n\|_{L^{q_n}} \|H^{\frac{q_n}{p_n}} w_n^{-\frac{q_n'}{p_n}}\|_{L^{p_n}} \lesssim \|f_n w_n\|_{L^{q_n}}.$$

Case 2: $1/s := 1/p - 1/q = 1/p_n - 1/q_n > 0$. Note that this case allows $q_n = \infty$. As observed in the above, $W_w^{-q_n'} \in A_{1+\frac{q_n'}{q}}(\widehat{w})$ and thus $M_{\widehat{w}}$ is bounded on $L^{1+\frac{q_n'}{q}}(W_w^{-q_n'}\widehat{w}) = L^{1+\frac{q_n'}{q}}(w_n^{-q_n'})$. Likewise, $M_{\widehat{\lambda}}$ is bounded on $L^{1+\frac{q_n'}{q}}(w_n^{-q_n'})$. Denote by $\|M_{\widehat{\lambda}}M_{\widehat{w}}\|$ the norm of $M_{\widehat{\lambda}}M_{\widehat{w}}$ on $L^{1+\frac{q_n'}{q}}(w_n^{-q_n'})$. We introduce the following Rubio de Francia algorithm:

$$\mathcal{R}g = \sum_{k=0}^{\infty} \frac{(M_{\widehat{\lambda}}M_{\widehat{w}})^{(k)}g}{2^k\|M_{\widehat{\lambda}}M_{\widehat{w}}\|^k}.$$

By duality, there exists some $0 \le h \in L^{\frac{s}{p}}(W_\lambda^q\widehat{\lambda})$ such that $\|h\|_{L^{\frac{s}{p}}(W_\lambda^q\widehat{\lambda})} = 1$ and

$$\|fW_\lambda\|_{L^q(\widehat{\lambda})} = \|f^p\|^{\frac{1}{p}}_{L^{\frac{q}{p}}(W_\lambda^q\widehat{\lambda})} = \Big(\int f^p h W_\lambda^q\widehat{\lambda}\Big)^{\frac{1}{p}}.$$

Set

$$H = \mathcal{R}\Big(h^{\frac{s}{p(1+\frac{q_n'}{q})}} w_n^{\frac{q_n'}{1+\frac{q_n'}{q}}} (W_\lambda^q\widehat{\lambda})^{\frac{1}{1+\frac{q_n'}{q}}}\Big)^{\frac{p(1+\frac{q_n'}{q})}{s}} w_n^{-\frac{q_n'p}{s}} (W_\lambda^q\widehat{\lambda})^{-\frac{p}{s}}.$$

Then it is easy to check that

$$h \le H, \quad \|H\|_{L^{\frac{s}{p}}(W_\lambda^q\widehat{\lambda})} \lesssim \|h\|_{L^{\frac{s}{p}}(W_\lambda^q\widehat{\lambda})} = 1$$

and

$$\Big[w_n^{\frac{q_n'}{1+\frac{q_n'}{q}}} (W_\lambda^q\widehat{\lambda})^{\frac{1}{1+\frac{q_n'}{q}}} H^{\frac{s}{p(1+\frac{q_n'}{q})}}\Big]_{A_1(\widehat{w})} \le 2\|M_{\widehat{\lambda}}M_{\widehat{w}}\|;$$

$$\Big[w_n^{\frac{q_n'}{1+\frac{q_n'}{q}}} (W_\lambda^q\widehat{\lambda})^{\frac{1}{1+\frac{q_n'}{q}}} H^{\frac{s}{p(1+\frac{q_n'}{q})}}\Big]_{A_1(\widehat{\lambda})} \le 2\|M_{\widehat{\lambda}}M_{\widehat{w}}\|.$$

Denote $v_n = H^{\frac{1}{p}} W_\lambda^{\frac{q}{p}} \widehat{\lambda}^{-\frac{1}{p_n'}}$, we claim

$$(w_1, \cdots, w_{n-1}, v_n), (\lambda_1, \cdots, w_{n-1}, v_n) \in A_{\vec{p}}. \tag{7.3}$$

Assume (7.3) for the moment, then

$$\|fW_\lambda\|_{L^q(\widehat{\lambda})} = \Big(\int f^p h W_\lambda^q\widehat{\lambda}\Big)^{\frac{1}{p}} \le \|fv_n\widehat{\lambda}^{\frac{1}{p_n'}}\|_{L^p(\widehat{\lambda})} \lesssim \|f_nv_n\|_{L^{p_n}} \prod_{i=1}^{n-1} \|f_iw_i\|_{L^{p_i}}.$$

We can conclude this case by noticing that

$$\begin{aligned}\|f_nv_n\|_{L^{p_n}} &\le \|f_nw_n\|_{L^{q_n}} \|v_nw_n^{-1}\|_{L^s} = \|f_nw_n\|_{L^{q_n}} \|H^{\frac{1}{p}} W_\lambda^{\frac{q}{p}} \widehat{\lambda}^{-\frac{1}{p_n'}} w_n^{-1}\|_{L^s} \\ &= \|f_nw_n\|_{L^{q_n}} \|H^{\frac{1}{p}} W_\lambda^{\frac{q}{s}} \widehat{\lambda}^{\frac{1}{s}}\|_{L^s} \lesssim \|f_nw_n\|_{L^{q_n}}.\end{aligned}$$

It remains to prove (7.3). Similar as before, it suffices to prove

$$v_n\widehat{w}^{\frac{1}{p_n'}} \in A_{p_n,p}(\widehat{w}) \quad \text{and} \quad v_n\widehat{\lambda}^{\frac{1}{p_n'}} \in A_{p_n,p}(\widehat{\lambda}).$$

Since

$$1 - \frac{p(1+\frac{q_n'}{q})}{s} = \frac{pq_n'}{qp_n'},$$

for arbitrary rectangle Q, direct calculus gives us

$$\begin{aligned}
&\Big(\frac{1}{\widehat{w}(Q)}\int_Q v_n^p \widehat{w}^{\frac{p}{p_n'}}\widehat{w}\Big)^{\frac1p} = \Big(\frac{1}{\widehat{w}(Q)}\int_Q HW_\lambda^q \widehat{\lambda}^{-\frac{p}{p_n'}}\widehat{w}^{\frac{p}{p_n'}+1}\Big)^{\frac1p}\\
&= \Big(\frac{1}{\widehat{w}(Q)}\int_Q H(W_\lambda^q\widehat{\lambda})^{\frac{p}{s}} w_n^{\frac{pq_n'}{s}} W_\lambda^{\frac{pq_n'}{p_n'}} \widehat{\lambda}^{-\frac{p}{p_n'}}\widehat{w}^{\frac{p}{p_n'}+1}\Big)^{\frac1p}\\
&\le \Big(\frac{1}{\widehat{w}(Q)}\int_Q w_n^{\frac{q_n'}{1+\frac{q_n'}{q}}}(W_\lambda^q\widehat{\lambda})^{\frac{1}{1+\frac{q_n'}{q}}} H^{\frac{s}{p(1+\frac{q_n'}{q})}}\widehat{w}\Big)^{\frac{1+\frac{q_n'}{q}}{s}} \Big(\frac{1}{\widehat{w}(Q)}\int_Q W_\lambda^q \widehat{\lambda}^{-\frac{q}{q_n'}}\widehat{w}^{\frac{q}{q_n'}+1}\Big)^{\frac{q_n'}{qp_n'}}\\
&\lesssim \inf_Q \big(w_n^{\frac{q_n'}{s}}(W_\lambda^q\widehat{\lambda})^{\frac1s}H^{\frac1p}\big)\Big(\frac{1}{\widehat{w}(Q)}\int_Q W_w^q\widehat{w}\Big)^{\frac{q_n'}{qp_n'}}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\Big(\frac{1}{\widehat{w}(Q)}\int_Q v_n^p \widehat{w}^{\frac{p}{p_n'}}\widehat{w}\Big)^{\frac{q_n'}{qp_n'}}\Big(\frac{1}{\widehat{w}(Q)}\int_Q v_n^{-p_n'}\Big)^{\frac{1}{p_n'}}\\
&\lesssim \Big(\frac{1}{\widehat{w}(Q)}\int_Q W_w^q\widehat{w}\Big)^{\frac{q_n'}{qp_n'}}\Big(\frac{1}{\widehat{w}(Q)}\int_Q w_n^{\frac{p_n'q_n'}{s}}(W_\lambda^q\widehat{\lambda})^{\frac{p_n'}{s}}H^{\frac{p_n'}{p}}v_n^{-p_n'}\Big)^{\frac{1}{p_n'}}\\
&= \Big(\frac{1}{\widehat{w}(Q)}\int_Q W_w^q\widehat{w}\Big)^{\frac{q_n'}{qp_n'}}\Big(\frac{1}{\widehat{w}(Q)}\int_Q w_n^{-q_n'}\Big)^{\frac{1}{p_n'}} \le [W_w]_{A_{q_n,q}(\widehat{w})}^{\frac{q_n'}{p_n'}}
\end{aligned}$$

This proves $v_n\widehat{w}^{\frac{1}{p_n'}} \in A_{p_n,p}(\widehat{w})$. The proof of $v_n\widehat{\lambda}^{\frac{1}{p_n'}} \in A_{p_n,p}(\widehat{\lambda})$ is similar.

Appendix A. Pointwise sparse bound of oscillation in rectangles

We provide a proof of the pointwise sparse bound

$$|b-\langle b\rangle_R^\sigma|1_R \lesssim \sum_{R'\in S_\gamma(R)} \langle |b-\langle b\rangle_{R'}^\sigma|\rangle_{R'}^\sigma 1_{R'},$$

where $\sigma \in A_\infty$ and $S_\gamma(R)$ is the dyadic sparse family inside the rectangle R attained by iteratively bisecting the size of R.

For brevity, set $\operatorname{osc}(b,R) = |b-\langle b\rangle_R^\sigma|1_R$. Fix a rectangle $R = R^1\times R^2 \in \mathcal{D}$ and we call $\mathcal{D}(R)$ the dyadic system inside of the rectangle R attained by iteratively bisecting the size of R. Let $\alpha = 2\langle \operatorname{osc}(b,R)\rangle_R^\sigma$ and then we form Calderón-Zygmund decomposition of $\operatorname{osc}(b,R)$ at level α with respect to the σ. We get a collection of maximal rectangles $\{R_i\}$ in $\mathcal{D}(R)$ with the property

$$\langle \operatorname{osc}(b,R)\rangle_{R_i}^\sigma > \alpha.$$

We take $E_0 = R\setminus\bigcup_i R_i$ and hence $|b-\langle b\rangle_R^\sigma|1_E \le \alpha$. By maximality, we have

$$\sum_i \sigma(R_i) < \alpha^{-1}\sum_i\int_{R_i}\operatorname{osc}(b,R)\sigma \le \frac12\sigma(R_i).$$

Thus, we get

$$\operatorname{osc}(b,R) = |b-\langle b\rangle_R^\sigma|1_E + \sum_i |b-\langle b\rangle_R^\sigma|1_{R_i}$$

$$\leq \alpha 1_E + \sum_i |\langle b\rangle^\sigma_{R_i} - \langle b\rangle^\sigma_R| 1_{R_i} + \sum_i |b - \langle b\rangle^\sigma_{R_i}| 1_{R_i}$$
$$\lesssim (D_\sigma + 1)\alpha 1_R + \sum_i |b - \langle b\rangle^\sigma_{R_i}| 1_{R_i},$$

where we use the maximality

$$\langle |b - \langle b\rangle^\sigma_R|\rangle_{(R_i)^{(1)}} \leq \alpha$$

and doubling property of $\sigma \in A_\infty$. Via recursive argument for $\sum_i |b - \langle b\rangle^\sigma_{R_i}| 1_{R_i}$ we complete the proof.

Now having this sparse domination in our hand, we can prove the following proposition

A.1. **Proposition.** *Let* $\nu, \sigma \in A_\infty$. *If* $\nu\sigma \in A_\infty$, *then it holds* $\mathrm{bmo}_\sigma(\nu) = \mathrm{bmo}(\nu)$, *where*

$$\mathrm{bmo}_\sigma(\nu) := \{b \in L^1_{loc} \colon \sup_R \frac{1}{\nu\sigma(R)} \int_R |b - \langle b\rangle^\sigma_R| \sigma < \infty\}.$$

The proof can be adapted from [17, Lemma 2.13], we omit the details.

References

[1] E. Airta, *Two-weight commutator estimates: general multi-parameter framework*, Publ. Mat. **64** (2020), no. 2, 681–729. ↑1, 10, 22

[2] E. Airta, K. Li, H. Martikainen, and E. Vuorinen, *Some new weighted estimates on product space*, Indiana Univ. Math. J. (2019), to appear, available at https://www.iumj.indiana.edu/IUMJ/Preprints/8807.pdf. ↑1

[3] E. Airta, H. Martikainen, and E. Vuorinen, *Modern singular integral theory with mild kernel regularity* (2019), preprint, available at https://arxiv.org/abs/2006.05807. ↑14, 22

[4] 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. ↑1

[5] D. Cruz-Uribe, J. M. Martell, and C. Pérez, *Extrapolation from A_∞ weights and applications*, J. Funct. Anal. **213** (2004), no. 2, 412–439. ↑7

[6] J. Duoandikoetxea, *Extrapolation of weights revisited: new proofs and sharp bounds*, J. Funct. Anal. **260** (2011), no. 6, 1886–1901. ↑2

[7] R. Fefferman, *Strong differentiation with respect to measures*, Amer. J. Math. **103** (1981), 33–40. ↑8

[8] L. Grafakos, L. Liu, R. H. Torres, and C. Pérez, *The multilinear strong maximal function*, J. Geom. Anal. **2** (2011), 118–149. ↑8

[9] L. Grafakos and J. M. Martell, *Extrapolation of weighted norm inequalities for multivariable operators and applications*, J. Geom. Anal. **14** (2004), no. 1, 19–46. ↑2

[10] 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. ↑1, 4, 10

[11] T. Hytönen, *The L^p-to-L^q boundedness of commutators with applications to the Jacobian operator* (2018), preprint, available at https://arxiv.org/abs/1804.11167. ↑1

[12] I. Kunwar and Y. Ou, *Two-weight inequalities for multilinear commutators*, New York J. Math. **24** (2018), 980–1003. ↑1

[13] A. K. Lerner, S. Ombrosi, C. Pérez, R. H. Torres, and R. Trujillo-González, *New maximal functions and multiple weights for the multilinear Calderón-Zygmund theory*, Adv. Math. **220** (2009), no. 4, 1222–1264. ↑2, 5

[14] A. K. Lerner, S. Ombrosi, and I. P. Rivera-Ríos, *On pointwise and weighted estimates for commutators of Calderón-Zygmund operators*, Adv. Math. **319** (2017), 153–181. ↑1

[15] A. K. Lerner, S. Ombrosi, and I. P. Rivera-Ríos, *Commutators of singular integrals revisited*, Bull. Lond. Math. Soc. **51** (2019), no. 1, 107–119. ↑1

[16] ———, *On two weight estimates for iterated commutators*, J. Funct. Anal. **281** (2021), no. 8, Paper No. 109153, 46. ↑6

[17] K. Li, *Multilinear commutators in the two-weight setting*, Bull. Lond. Math. Soc. (2021), to appear, available at https://arxiv.org/abs/2006.09071. ↑2, 5, 6, 42

[18] K. Li, J. M. Martell, and S. Ombrosi, *Extrapolation for multilinear Muckenhoupt classes and applications*, Adv. Math. **373** (2020), 107286. ↑2, 20, 38

[19] K. Li, J.M. Martell, H: Martikainen, S. Ombrosi, and E. Vuorinen, *End-point estimates, extrapolation for multilinear Muckenhoupt classes, and applications*, Trans. Amer. Math. Soc. **374** (2021), no. 1, 97–135. MR4188179 ↑2, 21, 38, 39

[20] K. Li, H. Martikainen, and E. Vuorinen, *Bloom type upper bounds in the product BMO setting*, J. Geom. Anal. **30** (2019), 3181–3203. ↑1, 22

[21] K. Li, H. Martikainen, and E. Vuorinen, *Genuinely multilinear weighted estimates for singular integrals in product spaces* (2020), preprint, available at https://arxiv.org/abs/2011.04459. ↑2, 5, 8, 9, 14, 15, 18, 19, 20, 31, 32, 33, 34, 35, 39

[22] ———, *Bloom-Type Inequality for Bi-Parameter Singular Integrals: Efficient Proof and Iterated Commutators*, Int. Math. Res. Not. IMRN **11** (2021), 8153–8187. ↑1, 8, 22, 27

[23] K. Li, K. Moen, and W. Sun, *The sharp weighted bound for multilinear maximal functions and Calderón–Zygmund operators*, J. Fourier Anal. Appl. **20** (2014), 751–756. ↑6

[24] B. Nieraeth, *Quantitative estimates and extrapolation for multilinear weight classes*, Math. Ann. **375** (2019), no. 1-2, 453–507. ↑2

[25] M. Wilson, *Weighted Littlewood-Paley theory and exponential-square integrability*, Lecture Notes in Mathematics, vol. 1924, Springer, Berlin, 2008. ↑8

[26] S. Wu, *A wavelet characterization for weighted Hardy spaces*, Rev. Mat. Iberoam. **8** (1992), no. 3, 329–349. ↑10

(E.A.) Department of Mathematics and Statistics, University of Helsinki, P.O.B. 68, FI-00014 University of Helsinki, Finland

Email address: emil.airta@helsinki.fi

(K.L.) Center for Applied Mathematics, Tianjin University, Weijin Road 92, 300072 Tianjin, China

Email address: kli@tju.edu.cn

(H.M.) Department of Mathematics and Statistics, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA

Email address: henri@wustl.edu