GENUINELY MULTILINEAR WEIGHTED ESTIMATES FOR SINGULAR INTEGRALS IN PRODUCT SPACES

KANGWEI LI, HENRI MARTIKAINEN, AND EMIL VUORINEN

ABSTRACT. We prove genuinely multilinear weighted estimates for singular integrals in product spaces. The estimates complete the qualitative weighted theory in this setting. Such estimates were previously known only in the one-parameter situation. Extrapolation gives powerful applications – for example, a free access to mixed-norm estimates in the full range of exponents.

1. INTRODUCTION

For given exponents $1 < p_1, \ldots, p_n < \infty$ and $1/p = \sum_i 1/p_i > 0$, a natural form of a weighted estimate in the n-variable context has the form

$$\Big\| g \prod_{i=1}^{n} w_i \Big\|_{L^p} \lesssim \prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}}$$

for some functions $f_1, \ldots, f_n$ and g. It is natural to initially assume that $w_i^{p_i} \in A_{p_i}$, where A_q stands for the classical Muckenhoupt weights. Even with this assumption the target weight only satisfies $\prod_{i=1}^{n} w_i^p \in A_{np} \supsetneq A_p$ making the case $n \geq 2$ have a different flavour than the classical case $n = 1$. Importantly, it turns out to be very advantageous – we get to the application later – to only impose a weaker *joint* condition on the tuple of weights $\vec{w} = (w_1, \ldots, w_n)$ rather than to assume individual conditions on the weights $w_i^{p_i}$. This gives the problem a genuinely multilinear nature. For many fundamental mappings $(f_1, \ldots, f_n) \mapsto g(f_1, \ldots, f_n)$, such as the n-linear maximal function, these joint conditions on the tuple $\vec{w}$ are necessary and sufficient for the weighted bounds.

Genuinely multilinear weighted estimates were first proved for n-linear *one-parameter* singular integral operators (SIOs) by Lerner, Ombrosi, Pérez, Torres and Trujillo-González in the extremely influential paper [35]. A basic model of an n-linear SIO T in $\mathbb{R}^d$ is obtained by setting

$$T(f_1, \ldots, f_n)(x) = U(f_1 \otimes \cdots \otimes f_n)(x, \ldots, x), \qquad x \in \mathbb{R}^d, \ f_i \colon \mathbb{R}^d \to \mathbb{C},$$

where U is a linear SIO in $\mathbb{R}^{nd}$. See e.g. Grafakos–Torres [19] for the basic theory. Estimates for SIOs play a fundamental role in pure and applied analysis – for example, L^p estimates for the homogeneous fractional derivative $D^\alpha f = \mathcal{F}^{-1}(|\xi|^\alpha \widehat{f}(\xi))$ of a product of two or more functions, the *fractional Leibniz rules*, are used in the area of dispersive equations, see e.g. Kato–Ponce [33] and Grafakos–Oh [18].

In the usual one-parameter context of [35] there is a general philosophy that the maximal function controls SIOs T – in fact, we have the concrete estimate

$$(1.1) \qquad \|T(f_1,\ldots,f_n)w\|_{L^p} \lesssim \|M(f_1,\ldots,f_n)w\|_{L^p}, \qquad p > 0, \, w^p \in A_\infty.$$

Thus, the heart of the matter of [35] reduces to the maximal function

$$M(f_1,\ldots,f_n) = \sup_I 1_I \prod_{i=1}^n \langle |f_i| \rangle_I,$$

where $\langle |f_i| \rangle_I = \fint_I |f_i| = \frac{1}{|I|} \int_I |f_i|$ and the supremum is over cubes $I \subset \mathbb{R}^d$.

In this paper we prove genuinely multilinear weighted estimates for multi-parameter SIOs in the product space $\mathbb{R}^d = \prod_{i=1}^m \mathbb{R}^{d_i}$. For the classical linear multi-parameter theory and some of its original applications see e.g. [7, 8, 22, 23, 24, 25, 26, 32]. Multilinear multi-parameter estimates arise naturally in applications whenever a multilinear phenomena, like the fractional Leibniz rules, are combined with product type estimates, such as those that arise when we want to take different partial fractional derivatives $D_{x_1}^\alpha D_{x_2}^\beta f$. We refer to our recent work [40] for a thorough general background on the subject.

It is already known [16] that the multi-parameter maximal function $(f_1,\ldots,f_n) \mapsto \sup_R 1_R \prod_{i=1}^n \langle |f_i| \rangle_R$, where the supremum is over rectangles $R = \prod_{i=1}^m I^i \subset \prod_{i=1}^m \mathbb{R}^{d_i}$ with sides parallel to the axes, satisfies the desired genuinely multilinear weighted estimates. However, in contrast to the one-parameter case, there is no known general principle which would automatically imply the corresponding weighted estimate for multi-parameter SIOs from the maximal function estimate. In particular, no estimate like (1.1) is known. In the paper [40] we developed the general theory of bilinear bi-parameter SIOs including weighted estimates under the more restrictive assumption $w_i^{p_i} \in A_{p_i}$. In fact, we only reached these weighted estimates without any additional cancellation assumptions of the type $T1 = 0$ in [1].

There are no genuinely multilinear weighted estimates for **any** multi-parameter SIOs in the literature – not even for the bi-parameter analogues (see e.g. [40, Appendix A]) of Coifman–Meyer [9] type multilinear multipliers. Almost ten years after the maximal function result [16] we establish these missing bounds – not only for some special SIOs – but for a very general class of n-linear m-parameter SIOs. With weighted bounds previously being known both in the linear multi-parameter setting [22, 23, 28] and in the multilinear one-parameter setting [35], we finally establish a holistic view completing the theory of qualitative weighted estimates in the joint presence of multilinearity and product space theory.

With the understanding that a Calderón–Zygmund operator (CZO) is an SIO satisfying natural $T1$ type assumptions, our main result reads as follows.

1.2. **Theorem.** *Suppose T is an n-linear m-parameter CZO in $\mathbb{R}^d = \prod_{i=1}^m \mathbb{R}^{d_i}$. If $1 < p_1,\ldots,p_n \le \infty$ and $1/p = \sum_{i=1}^n 1/p_i > 0$, we have*

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

for all n-linear m-parameter weights $\vec{w} = (w_1,\ldots,w_n) \in A_{\vec{p}}$, $\vec{p} = (p_1,\ldots,p_n)$. Here $\vec{w} \in A_{\vec{p}}$ if

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

where the supremum is over all rectangles $R \subset \mathbb{R}^d$.

For the exact definitions, see the main text.

Recent extrapolation methods are crucial both for the proof and for the applications. 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 [17] (see also [15]) a multivariable analogue was developed in the setting $w_i^{p_i} \in A_{p_i}$, $i = 1, \ldots, n$. Such extrapolation results are already of fundamental use in proving other estimates – often just to even deduce the full n-linear range of unweighted estimates $\prod_{j=1}^n L^{p_j} \to L^p$, $\sum_j 1/p_j = 1/p$, $1 < p_j < \infty$, $1/n < p < \infty$, from some particular single tuple $(p_1, \ldots, p_n, p)$. Indeed, reaching $p \leq 1$ can often be a crucial challenge, particularly so in multi-parameter settings where many other tools are completely missing.

Very recently, in [37] it was shown that also the genuinely multilinear weighted estimates can be extrapolated. In the subsequent paper [38] (see also [45]) a key advantage of extrapolating using the general weight classes was identified: it is possible to both start the extrapolation and, importantly, to reach – as a consequence of the extrapolation – weighted estimates with $p_i = \infty$. See Theorem 3.12 for a formulation of these general extrapolation principles. Moreover, extrapolation is flexible in the sense that one can extrapolate both 1-parameter and m-parameter, $m \geq 2$, weighted estimates.

These new extrapolation results are extremely useful e.g. in proving mixed-norm estimates – for example, in the bi-parameter case they yield that

$$\|T(f_1, \ldots, f_n)\|_{L^p(\mathbb{R}^{d_1}; L^q(\mathbb{R}^{d_2}))} \lesssim \prod_{i=1}^n \|f_i\|_{L^{p_i}(\mathbb{R}^{d_1}; L^{q_i}(\mathbb{R}^{d_2}))},$$

where $1 < p_i, q_i \leq \infty$, $\frac{1}{p} = \sum_i \frac{1}{p_i} > 0$ and $\frac{1}{q} = \sum_i \frac{1}{q_i} > 0$. The point is that even all of the various cases involving ∞ become immediate. See e.g. [14, 38, 40] for some of the previous mixed-norm estimates. Compared to [38] we can work with completely general n-linear m-parameter SIOs instead of bi-parameter tensor products of 1-parameter SIOs, and the proof is much simplified due to the optimal weighted estimates, Theorem 1.2.

We also use extrapolation to give a new short proof of the boundedness of the multi-parameter n-linear maximal function [16] – see Proposition 4.1.

On the technical level there is no existing approach to our result: the modern one-parameter tools (such as sparse domination in the multilinear setting, see e.g. [6]) are missing and many of the bi-parameter methods [40] used in conjunction with the assumption that each weight individually satisfies $w_i^{p_i} \in A_{p_i}$ are of little use. Aside from maximal function estimates, multi-parameter estimates require various square function estimates (and combinations of maximal function and square function estimates). Similarly as one cannot use $\prod_i M f_i$ instead of $M(f_1, \ldots, f_n)$ due to the nature of the multilinear weights, it is also not possible to use classical square function estimates separately for the functions f_i. Now, this interplay makes it impossible to decouple estimates to terms like $\|Mf_1 \cdot w_1\|_{L^{p_1}} \|Sf_2 \cdot w_2\|_{L^{p_2}}$, since neither of them would be bounded separately as $w_1^{p_1} \notin A_{p_1}$ and $w_2^{p_2} \notin A_{p_2}$. However, such decoupling of estimates has previously seemed almost indispensable.

Our proof starts with the reduction to dyadic model operators [2] (see also [13, 31, 41, 43, 47]), which is a standard idea. After this we introduce a family of n-linear multi-parameter square function type objects A_k. On the idea level, a big part of the proof

works by taking a dyadic model operator S and finding an appropriate square function A_k so that

$$\|S(f_1, \ldots, f_n)w\|_{L^p} \lesssim \|A_k(f_1, \ldots, f_n)w\|_{L^p}.$$

This requires different tools depending on the model operator in question and is a new way to estimate model operators that respects the n-linear structure fully. We then prove that all of our operators A_k satisfy the genuinely n-linear weighted estimates

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

This is done with an argument that is based on using duality and lower square function estimates iteratively until all of the cancellation present in these square functions has been exploited.

Aside from the full range of mixed-norm estimates, the weighted estimates immediately give other applications as well. We present here a result on commutators, which greatly generalises [40]. Commutator estimates appear all over analysis implying e.g. factorizations for Hardy functions [10], certain div-curl lemmas relevant in compensated compactness, and were recently connected to the Jacobian problem $Ju = f$ in L^p (see [29]). For a small sample of commutator estimates in various other key setting see e.g. [26, 27, 28, 36].

1.3. **Theorem.** *Suppose T is an n-linear m-parameter CZO in $\mathbb{R}^d = \prod_{i=1}^{m} \mathbb{R}^{d_i}$, $1 < p_1, \ldots, p_n \leq \infty$ and $1/p = \sum_{i=1}^{n} 1/p_i > 0$. Suppose also that $\|b\|_{\mathrm{bmo}} = \sup_R \frac{1}{|R|} \int_R |b - \langle b \rangle_R| < \infty$. Then for all $1 \leq k \leq n$ we have the commutator estimate*

$$\|[b, T]_k(f_1, \ldots, f_n)w\|_{L^p} \lesssim \|b\|_{\mathrm{bmo}} \prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}},$$

$$[b, T]_k(f_1, \ldots, f_n) := bT(f_1, \ldots, f_n) - T(f_1, \ldots, f_{k-1}, bf_k, f_{k+1}, \ldots, f_n),$$

for all n-linear m-parameter weights $\vec{w} = (w_1, \ldots, w_n) \in A_{\vec{p}}$. Analogous results hold for iterated commutators.

We note that we can also finally dispose of some of the sparse domination tools that restricted some of the theory of [40] to bi-parameter.

2. Preliminaries

Throughout this paper $A \lesssim B$ means that $A \leq CB$ 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 $A \leq C(\epsilon)B$.

2.A. **Dyadic notation.** Given a dyadic grid $\mathcal{D}$ in $\mathbb{R}^d$, $I \in \mathcal{D}$ and $k \in \mathbb{Z}$, $k \geq 0$, 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 $\eta \neq 0$ 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, $\eta \neq 0$. 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), $\langle f, h_I \rangle := \int f h_I$.

We make a few clarifying comments related to the use of Haar functions. In the model operators coming from the representation theorem there are Haar functions involved. There we use the just mentioned convention that h_I means some unspecified cancellative Haar function h_I^η which we do not specify. On the other hand, the square function estimates in Section 5 are formulated using martingale differences (which involve multiple Haar functions as $\Delta_I f = \sum_{\eta \neq 0} \langle f, h_I^\eta \rangle h_I^\eta$). When we estimate the model operators, we carefully consider this difference by passing from the Haar functions into martingale differences via the simple identity

(2.1) $$\langle f, h_I \rangle = \langle \Delta_I f, h_I \rangle,$$

which follows from $\Delta_I f = \sum_{\eta \neq 0} \langle f, h_I^\eta \rangle h_I^\eta$ and orthogonality.

2.B. **Multi-parameter notation.** We will be working on the m-parameter product space $\mathbb{R}^d = \prod_{i=1}^m \mathbb{R}^{d_i}$. 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 $\prod_{i=1}^m I^i, \prod_{i=1}^m J^i, \prod_{i=1}^m K^i$ etc. We usually denote the collection of dyadic rectangles by $\mathcal{D} = \prod_{i=1}^m \mathcal{D}^i$.

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 $A^1 f(x) = A(f(\cdot, x_2, \ldots, x_n))(x_1)$. 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, an m-parameter martingale difference is

$$\Delta_R f = \Delta_{I^1}^1 \cdots \Delta_{I^m}^m f, \qquad R = \prod_{i=1}^m I^i.$$

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

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

or

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

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

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

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

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

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). These notions have obvious extensions to m-parameters.

2.D. **Structure of the paper.** To avoid unnecessarily complicating the notation, we start by proving everything in the bi-parameter case $m = 2$. Importantly, we present a proof which does not exploit this in a way that would not be extendable to m-parameters (e.g., our proof for the partial paraproducts does not exploit sparse domination for the appearing one-parameter paraproducts). At the end, we demonstrate for some key model operators how the general case can be dealt with.

3. WEIGHTS

The following notions have an obvious extension to m-parameters. 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. Thus, this is the one-parameter definition but cubes are replaced by rectangles.

We have

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

and that

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

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. [28]. We say $w \in A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ if

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

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

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

We introduce the classes of multilinear Muckenhoupt weights that we will use.

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

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

almost everywhere and

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

where the supremum is over rectangles R,

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

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 $\operatorname{ess\,sup}_R w$.

3.3. *Remark.*   (1) It is important that the lower bound

(3.4) $$\langle w^p \rangle_R^{\frac{1}{p}} \prod_{i=1}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \ge 1$$

holds always. To see this recall that for $\alpha_1, \alpha_2 > 0$ we have by Hölder's inequality that

(3.5) $$1 \le \langle w^{-\alpha_1} \rangle_R^{\frac{1}{\alpha_1}} \langle w^{\alpha_2} \rangle_R^{\frac{1}{\alpha_2}}.$$

Apply this with $\alpha_2 = p$ and $\alpha_1 = \frac{1}{n - \frac{1}{p}}$. Then apply Hölder's inequality with the exponents $u_i = \left(n - \frac{1}{p} \right) p_i'$ to get $\left\langle \left(\prod_{i=1}^n w_i \right)^{-\frac{1}{n - \frac{1}{p}}} \right\rangle_R^{n - \frac{1}{p}} \le \prod_{i=1}^n \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}}$.

(2) Our definition is essentially the usual one-parameter definition [35] with the difference that cubes are replaced by rectangles. However, we are also using the renormalised definition from [38] that works better with the exponents $p_i = \infty$.

Compared to the usual formulation of [35] the relation is that $[w_1^{p_1}, \ldots, w_n^{p_n}]_{A_{\vec{p}}}^{\frac{1}{p}}$ with $A_{\vec{p}}$ defined as in [35] agrees with our $[\vec{w}]_{A_{\vec{p}}}$ when $p_i < \infty$.

(3) The case $p_1 = \cdots = p_n = \infty = p$ can be used as the starting point of extrapolation. This is rarely useful but we will find use for it when we consider the multilinear maximal function.

The following characterization of the class $A_{\vec{p}}$ is convenient. The one-parameter result with the different normalization is [35, Theorem 3.6]. We record the proof for the convenience of the reader.

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

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

and

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

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

Conversely, we have

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

Proof. We fix an arbitrary $j \in \{1, \ldots, n\}$ for which we will show $[w_j^{-p_j'}]_{A_{np_j'}} \le [\vec{w}]_{A_{\vec{p}}}^{p_j'}.$ Notice that

(3.7)
$$\frac{1}{p} + \sum_{i \ne j} \frac{1}{p_i'} = n - 1 + \frac{1}{p_j}.$$

We define q_j via the identity

$$\frac{1}{q_j} = \frac{1}{n - 1 + \frac{1}{p_j}} \cdot \frac{1}{p}$$

and for $i \ne j$ we set

$$\frac{1}{q_i} = \frac{1}{n - 1 + \frac{1}{p_j}} \cdot \frac{1}{p_i'}.$$

From (3.7) we have that $\sum_i \frac{1}{q_i} = 1$. By definition we have

(3.8)
$$[w_j^{-p_j'}]_{A_{np_j'}} = \sup_R \langle w_j^{-p_j'}\rangle_R \langle w_j^{p_j' \frac{1}{np_j'-1}}\rangle_R^{np_j'-1}.$$

Notice that

$$p_j' \frac{1}{np_j' - 1} = \frac{1}{n - \frac{1}{p_j'}} = \frac{1}{n - 1 + \frac{1}{p_j}}.$$

Using Hölder's inequality with the exponents $q_1, \ldots, q_n$ we have the desired estimate

$$\langle w_j^{-p_j'}\rangle_R^{\frac{1}{p_j'}} \langle w_j^{\frac{p}{q_j}}\rangle_R^{\frac{q_j}{p}} = \langle w_j^{-p_j'}\rangle_R^{\frac{1}{p_j'}} \langle w^{\frac{p}{q_j}} \prod_{i \ne j} w_i^{-\frac{p}{q_j}}\rangle_R^{\frac{q_j}{p}} \le \langle w^p\rangle_R^{\frac{1}{p}} \prod_i \langle w_i^{-p_i'}\rangle_R^{\frac{1}{p_i'}} \le [\vec{w}]_{A_{\vec{p}}}.$$

When $p_j = 1$ this is $\operatorname{ess\,sup}_R w_j^{-1} \langle w_j^{\frac{1}{n}}\rangle_R^n \le [\vec{w}]_{A_{\vec{p}}}$, and so $[w_j^{\frac{1}{n}}]_{A_1}^n \le [\vec{w}]_{A_{\vec{p}}}$.

We now move on to bounding $[w^p]_{A_{np}}$. Notice that by definition

$$(3.9) \qquad [w^p]_{A_{np}} = \sup_R \langle w^p \rangle_R \langle w^{-\frac{p}{np-1}} \rangle_R^{np-1}.$$

We define s_i via

$$-\frac{p}{np-1} \cdot s_i = -p_i'$$

and notice that then $\sum_i \frac{1}{s_i} = 1$. Then, by Hölder's inequality with the exponents $s_1, \ldots, s_n$ we have

$$\langle w^p \rangle_R \langle w^{-\frac{p}{np-1}} \rangle_R^{np-1} \leq \langle w^p \rangle_R \prod_i \langle w_i^{-p_i'} \rangle_R^{\left(\frac{p}{np-1}\right)\frac{1}{p_i'}(np-1)} = \left[\langle w^p \rangle_R^{\frac{1}{p}} \prod_i \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} \right]^p \leq [\vec{w}]_{A_{\vec{p}}}^p,$$

which is the desired bound for $[w^p]_{A_{np}}$. Notice that in the case $p = \infty$ we get

$$[w^{-\frac{1}{n}}]_{A_1}^n = \sup_R \langle w^{-\frac{1}{n}} \rangle_R^n \operatorname*{ess\,sup}_R w \leq \sup_R \left[\prod_i \langle w_i^{-1} \rangle_R \right] \operatorname*{ess\,sup}_R w = [\vec{w}]_{A_{\vec{p}}}.$$

We then move on to bounding $[\vec{w}]_{A_{\vec{p}}}$. It is based on the following inequality

$$(3.10) \qquad 1 \leq \langle w^{-\frac{p}{np-1}} \rangle_R^{n-\frac{1}{p}} \prod_i \left\langle w_i^{\frac{1}{n-1+\frac{1}{p_i}}} \right\rangle_R^{n-1+\frac{1}{p_i}}.$$

Before proving this, we show how it implies the desired bound. We have

$$[\vec{w}]_{A_{\vec{p}}} = \sup_R \langle w^p \rangle_R^{\frac{1}{p}} \prod_i \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}}$$

$$\leq \sup_R \left[\langle w^p \rangle_R \langle w^{-\frac{p}{np-1}} \rangle_R^{np-1} \right]^{\frac{1}{p}} \prod_i \left[\langle w_i^{-p_i'} \rangle_R \langle w_i^{p_i' \frac{1}{np_i'-1}} \rangle_R^{np_i'-1} \right]^{\frac{1}{p_i'}} \leq [w^p]_{A_{np}}^{\frac{1}{p}} \prod_i [w_i^{-p_i'}]_{A_{np_i'}}^{\frac{1}{p_i'}},$$

where in the last estimate we recalled (3.8) and (3.9).

Let us now give the details of (3.10). We apply (3.5) with $\alpha_1 = \frac{p}{np-1}$ and $\alpha_2 = \frac{1}{n(n-1)+\frac{1}{p}}$ to get

$$1 \leq \langle w^{-\frac{p}{np-1}} \rangle_R^{n-\frac{1}{p}} \left\langle w^{\frac{1}{n(n-1)+\frac{1}{p}}} \right\rangle_R^{n(n-1)+\frac{1}{p}}.$$

The first term is already as in (3.10). Define u_i via

$$\frac{1}{n(n-1)+\frac{1}{p}} u_i = \frac{1}{n-1+\frac{1}{p_i}}$$

and notice that by Hölder's inequality with these exponents $(\sum_i \frac{1}{u_i} = 1)$ we have

$$\left\langle w^{\frac{1}{n(n-1)+\frac{1}{p}}} \right\rangle_R^{n(n-1)+\frac{1}{p}} \leq \prod_i \left\langle w_i^{\frac{1}{n-1+\frac{1}{p_i}}} \right\rangle_R^{n-1+\frac{1}{p_i}},$$

which matches the second term in (3.10). $\qquad \square$

The following duality of multilinear weights is handy – see [42, Lemma 3.1]. We give the short proof for convenience.

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

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

Then we have

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

Proof. We take $i = 1$ for notational convenience. Notice that $\frac{1}{p'} + \sum_{i=2}^{n} \frac{1}{p_i} = \frac{1}{p_1'}$. Notice also that $w^{-1} \prod_{i=2}^{n} w_i = w_1^{-1}$. Therefore, we have

$$[\vec{w}^i]_{A_{\vec{p}^i}} = \langle w_1^{-p_1'} \rangle_R^{\frac{1}{p_1'}} \langle w^p \rangle_R^{\frac{1}{p}} \prod_{i=2}^{n} \langle w_i^{-p_i'} \rangle_R^{\frac{1}{p_i'}} = [\vec{w}]_{A_{\vec{p}}}.$$

$\square$

We now recall the recent extrapolation result of [38]. The previous version, which did not yet allow exponents to be ∞ appeared in [37]. For related independent work see [45]. The previous extrapolation results with the separate assumptions $w_i^{p_i} \in A_{p_i}$ appear in [17] and [15]. An even more general result than the one below appears in [38], but we will not need that generality here. Finally, we note that the proof of this extrapolation result can be made to work in m-parameters even though [38] provides the details only in the one-parameter case – we give more details later in Section 8.

3.12. **Theorem.** *Let $f_1, \ldots, f_n$ and g be given functions. Given $\vec{p} = (p_1, \ldots, p_n)$ with $1 \leq p_1, \ldots, p_n \leq \infty$ let $\frac{1}{p} = \sum_{i=1}^{n} \frac{1}{p_i}$. Assume that given any $\vec{w} = (w_1, \ldots, w_n) \in A_{\vec{p}}$ the inequality*

$$(3.13) \qquad \qquad \|gw\|_{L^p} \lesssim \prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}}$$

holds, where $w := \prod_{i=1}^{n} w_i$. Then for all exponents $\vec{q} = (q_1, \ldots, q_n)$, with $1 < q_1, \ldots, q_n \leq \infty$ and $\frac{1}{q} = \sum_{i=1}^{n} \frac{1}{q_i} > 0$, and for all weights $\vec{v} = (v_1, \ldots, v_n) \in A_{\vec{q}}$ the inequality

$$\|gv\|_{L^q} \lesssim \prod_{i=1}^{n} \|f_i v_i\|_{L^{q_i}}$$

holds, where $v := \prod_{i=1}^{n} v_i$.

Given functions $f_1^j, \ldots, f_n^j$ and g^j so that (3.13) holds uniformly on j, we have for the same family of exponents and weights as above, and for all exponents $\vec{s} = (s_1, \ldots, s_n)$ with $1 < s_1, \ldots, s_n \leq \infty$ and $\frac{1}{s} = \sum_i \frac{1}{s_i} > 0$ the inequality

$$(3.14) \qquad \qquad \|(g^j v)_j\|_{L^q(\ell^s)} \lesssim \prod_{i=1}^{n} \|(f_i^j v_i)_j\|_{L^{q_i}(\ell^{s_i})}.$$

3.15. *Remark.* Using Lemma 3.11 and extrapolation, Theorem 3.12, we see that the weighted boundedness of T transfers to the adjoints T^{j*}. Partial adjoints have to always be considered separately, though.

As a final thing in this section, we demonstrate the necessity of the $A_{\vec{p}}$ condition for the weighted boundedness of SIOs. We work in the m-parameter setting and let $\mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$. Let R_j be the following version of the n-linear one-parameter Riesz transform in $\mathbb{R}^{d_j}$:

$$R_j(f_1, \ldots, f_n) = \text{p.v.} \int_{\mathbb{R}^{d_j n}} \frac{\sum_{i=1}^n \sum_{k=1}^{d_j} (x - y_i)_k}{(\sum_{i=1}^n |x - y_i|)^{d_j n + 1}} f_1(y_1) \cdots f_n(y_n) \, \mathrm{d}y_1 \cdots \mathrm{d}y_n,$$

where $(x - y_i)_k$ is the k-th coordinate of $x - y_i \in \mathbb{R}^{d_j}$. Consider the tensor product $R_1 \otimes R_2 \otimes \cdots \otimes R_m$. Let $\vec{w} = (w_1, \ldots, w_n)$ be a multilinear weight, that is, $0 < w_i < \infty$ a.e., and denote $w = \prod_{i=1}^n w_i$. Suppose that for some exponents $1 < p_1, \ldots, p_n \leq \infty$ with $1/p = \sum_{i=1}^n 1/p_i > 0$ the estimate

$$\|R_1 \otimes R_2 \otimes \cdots \otimes R_m(f_1, \ldots, f_n)\|_{L^{p,\infty}(w^p)} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}$$

holds for all $f_i \in L_c^\infty$. We show that $\vec{w}$ is an m-parameter $A_{\vec{p}}$ weight.

Define $\sigma_i = w_i^{-p_i'}$. Let $E \subset \mathbb{R}^d$ be an arbitrary set such that $1_E \sigma_i \in L_c^\infty$ for all $i = 1, \ldots, n$. Fix an m-parameter rectangle $R = R^1 \times \cdots \times \mathbb{R}^m \subset \mathbb{R}^d$, where each R^j is a cube. Let $R^+ = (R^1)^+ \times \cdots \times (R^m)^+$, where $(R^j)^+ := R^j + (\ell(R^j), \ldots, \ell(R^j))$.

Using the kernel of $R_1 \otimes \cdots \otimes R_m$ we have for all $x \in R^+$ that

$$R_1 \otimes R_2 \otimes \cdots \otimes R_m(1_E \sigma_1 1_R, \ldots, 1_E \sigma_n 1_R)(x) \gtrsim \prod_{i=1}^n \langle 1_E \sigma_i \rangle_R.$$

Hence

$$w^p(R^+)^{\frac{1}{p}} \prod_{i=1}^n \langle 1_E \sigma_i \rangle_R \lesssim \prod_{i=1}^n \|1_E \sigma_i 1_R w_i\|_{L^{p_i}} = \prod_{i=1}^n \sigma_i(E \cap R)^{\frac{1}{p_i}},$$

which gives that $\langle w^p \rangle_{R^+}^{\frac{1}{p}} \prod_{i=1}^n \langle 1_E \sigma_i \rangle_R^{\frac{1}{p_i'}} \lesssim 1$. Since E was arbitrary this implies the estimate

(3.16) $$\langle w^p \rangle_{R^+}^{\frac{1}{p}} \prod_{i=1}^n \langle \sigma_i \rangle_R^{\frac{1}{p_i'}} \lesssim 1.$$

Similarly, we can show that

(3.17) $$\langle w^p \rangle_R^{\frac{1}{p}} \prod_{i=1}^n \langle \sigma_i \rangle_{R^+}^{\frac{1}{p_i'}} \lesssim 1.$$

By Hölder's inequality we have that

$$\langle w^{-\frac{p}{np-1}} \rangle_{R^+}^{\frac{np-1}{p}} \leq \prod_{i=1}^n \langle \sigma_i \rangle_{R^+}^{\frac{1}{p_i'}}.$$

Hence, (3.17) shows that

$$\langle w^p \rangle_R^{\frac{1}{p}} \langle w^{-\frac{p}{np-1}} \rangle_{R^+}^{\frac{np-1}{p}} \lesssim 1.$$

Therefore,

$$\frac{\langle w^p \rangle_R^{\frac{1}{p}}}{\langle w^p \rangle_{R^+}^{\frac{1}{p}}} = \frac{\langle w^p \rangle_R^{\frac{1}{p}} \langle w^{-\frac{p}{np-1}} \rangle_{R^+}^{\frac{np-1}{p}}}{\langle w^p \rangle_{R^+}^{\frac{1}{p}} \langle w^{-\frac{p}{np-1}} \rangle_{R^+}^{\frac{np-1}{p}}} \lesssim 1,$$

where the denominator in the middle term was ≥ 1. Thus, $\langle w^p \rangle_R^{\frac{1}{p}} \lesssim \langle w^p \rangle_{R^+}^{\frac{1}{p}}$, which together with (3.16) gives that $\langle w^p \rangle_R^{\frac{1}{p}} \prod_{i=1}^n \langle \sigma_i \rangle_R^{\frac{1}{p_i'}} \lesssim 1$.

4. MAXIMAL FUNCTIONS

It was proved in [16] that the multilinear bi-parameter (or multi-parameter) maximal function is bounded with respect to the genuinely multilinear bi-parameter weights. We give a new efficient proof of this. Let $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$ be a fixed lattice of dyadic rectangles and define

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

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

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

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

Proof. Our proof is based on the proof of the case $\vec{p} = (p_1, \ldots, p_n) = (\infty, \ldots, \infty)$ and extrapolation, Theorem 3.12. We have

$$\sup_R \Big[\prod_i \langle w_i^{-1} \rangle_R \Big] \cdot \operatorname*{ess\,sup}_R w = [\vec{w}]_{A_{\vec{p}}},$$

and therefore

$$\prod_i \langle w_i^{-1} \rangle_R \lesssim \frac{1}{\operatorname{ess\,sup}_R w}.$$

For every $R \in \mathcal{D}$ let $N_R \subset R$ be such that $|N_R| = 0$ and $w(x) \leq \operatorname{ess\,sup}_R w$ for all $x \in R \setminus N_R$. Let $N = \bigcup_{R \in \mathcal{D}} N_R$. Then $|N| = 0$ and for every $x \in \mathbb{R}^d \setminus N$ we have

$$\frac{1}{w(x)} \geq \sup_{R \in \mathcal{D}} \frac{1_R(x)}{\operatorname{ess\,sup}_R w}.$$

Thus, we have

$$M_{\mathcal{D}}(f_1, \ldots, f_n)(x) w(x) \leq \Big[\prod_i \|f_i w_i\|_{L^\infty} \Big] \sup_{R \in \mathcal{D}} \Big[1_R(x) \prod_i \langle w_i^{-1} \rangle_R \Big] \cdot w(x)$$

$$\lesssim \Big[\prod_i \|f_i w_i\|_{L^\infty} \Big] \sup_{R \in \mathcal{D}} \Big[\frac{1_R(x)}{\operatorname{ess\,sup}_R w} \Big] \cdot w(x) \leq \prod_i \|f_i w_i\|_{L^\infty}$$

almost everywhere, and so $\|M_{\mathcal{D}}(f_1, \ldots, f_n) w\|_{L^\infty} \lesssim \prod_i \|f_i w_i\|_{L^\infty}$ as desired.

$\square$

If an average is with respect to a different measure μ than the Lebesgue measure, we write $\langle f \rangle_R^\mu := \frac{1}{\mu(R)} \int_R f \, \mathrm{d}\mu$ and define

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

The following is a result of R. Fefferman [24]. Recently, we also recorded a proof in [39, Appendix B].

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

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

We formulate some vector-valued versions of Proposition 4.2. We state the following version with two sequence spaces – of course, a version with arbitrarily many also works. Proposition 4.3 is proved in the end of Section 8.

4.3. **Proposition.** *Let* $\mu \in A_\infty$, $w \in A_p(\mu)$ *and* $1 < p, s, t < \infty$. *Then we have*

$$\left\| \left\| \|\{M^\mu f_j^i\}\|_{\ell^s} \right\|_{\ell^t} \right\|_{L^p(w\mu)} \lesssim \left\| \left\| \|\{f_j^i\}\|_{\ell^s} \right\|_{\ell^t} \right\|_{L^p(w\mu)}.$$

In particular, we have

$$\left\| \left\| \|\{M^\mu f_j^i\}\|_{\ell^s} \right\|_{\ell^t} \right\|_{L^p(\mu)} \lesssim \left\| \left\| \|\{f_j^i\}\|_{\ell^s} \right\|_{\ell^t} \right\|_{L^p(\mu)}.$$

Finally, we point out that everything in this section works easily in the general multi-parameter situation.

5. SQUARE FUNCTIONS

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 following lower square function estimate valid for A_∞ weights is important for us. The importance comes from the fact that by Lemma 3.6 some of the key weights w^p and $w_i^{-p_i'}$ are at least A_∞ for the multilinear weights of Definition 3.2.

5.1. **Lemma.** *There holds*

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

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

For a proof of the one-parameter estimate see [48, Theorem 2.5]. The bi-parameter results can be deduced using the following extremely useful A_∞ extrapolation result [11], which will be applied several times during the paper. We also mention that square function estimates related to Lemma 5.1 also appear in [3].

5.2. 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*

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

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

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

Proof of Lemma 5.1. Let $w \in A_\infty$ be a bi-parameter weight. The first estimate in the statement follows from the one-parameter result [48, Theorem 2.5] and the fact that $w(x_1, \cdot) \in A_\infty(\mathbb{R}^{d_2})$ and $w(\cdot, x_2) \in A_\infty(\mathbb{R}^{d_1})$. Using this, we have that

$$\|f\|_{L^2(w)} \lesssim \|S_{\mathcal{D}^1}^1 f\|_{L^2(w)} = \Big(\sum_{I^1 \in \mathcal{D}^1} \|\Delta_{I^1}^1 f\|_{L^2(w)}^2 \Big)^{\frac{1}{2}}.$$

For each I^1 we again use the one-parameter estimate to get

$$\|\Delta_{I^1}^1 f\|_{L^2(w)} \lesssim \|S_{\mathcal{D}^2}^2 \Delta_{I^1}^1 f\|_{L^2(w)} = \Big(\sum_{I^2 \in \mathcal{D}^2} \|\Delta_{I^2}^2 \Delta_{I^1}^1 f\|_{L^2(w)}^2 \Big)^{\frac{1}{2}}.$$

Since $\Delta_{I^2}^2 \Delta_{I^1}^1 f = \Delta_{I^1 \times I^2} f$, inserting the last estimate into the previous one shows that

$$\|f\|_{L^2(w)} \lesssim \Big(\sum_{I^1 \times I^2 \in \mathcal{D}^1 \times \mathcal{D}^2} \|\Delta_{I^1 \times I^2} f\|_{L^2(w)}^2 \Big)^{\frac{1}{2}} = \|S_{\mathcal{D}} f\|_{L^2(w)}.$$

Since this holds for every bi-parameter weight $w \in A_\infty$, Lemma 5.2 concludes the proof. We point out that with further extrapolation we could obtain vector-valued versions analogous to Proposition 4.3, see the end of Section 8. $\qquad\square$

5.3. *Remark.* We often use the lower square function estimate with the additional observation that we e.g. have for all $k = (k_1, k_2) \in \{0, 1, \ldots\}^2$ that

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

This simply follows from disjointness.

For $k = (k_1, k_2)$ we define the following family of n-linear square functions. First, we set

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

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

$$A_{1,k}(f_1, \ldots, f_n) = \Big(\sum_{K \in \mathcal{D}} \langle |\Delta_{K^2, k_2}^2 f_1| \rangle_K^2 \langle |\Delta_{K^1, k_1}^1 f_2| \rangle_K^2 \prod_{j=3}^n \langle |f_j| \rangle_K^2 1_K \Big)^{\frac{1}{2}}.$$

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

$$A_{2,k}(f_1, \ldots, f_n)$$
$$(5.4) \qquad = \Big(\sum_{K^2 \in \mathcal{D}^2} \Big(\sum_{K^1 \in \mathcal{D}^1} \langle |\Delta^2_{K^2,k_1} f_1| \rangle_K \langle |\Delta^1_{K^1,k_2} f_2| \rangle_K \langle |\Delta^1_{K^1,k_3} f_3| \rangle_K \prod_{j=4}^n \langle |f_j| \rangle_K 1_K \Big)^2 \Big)^{\frac{1}{2}},$$

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

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

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

5.5. Theorem. *If $1 < p_1, \ldots, p_n \le \infty$ and $\frac{1}{p} = \sum_{i=1}^n \frac{1}{p_i} > 0$ we have*

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

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

Proof. The proofs of all of the cases have the same underlying idea based on an iterative use of duality and the lower square function estimate until all of the cancellation has been utilised. One can also realise that the result for $A_{3,k}$ follows using the above scheme just once if the result is first proved for $A_{1,k}$ and $A_{2,k}$.

We show the proof for $A_{2,k}$ with the explicit form (5.4). Fix some $\vec{p} = (p_1, \ldots, p_n)$ with $1 < p_i < \infty$ and $p > 1$. This is enough by extrapolation, Theorem 3.12. To estimate $\|A_{2,k}(f_1, \ldots, f_n) w\|_{L^p}$ we take a sequence $(f_{n+1,K^2})_{K^2} \subset L^{p'}(\ell^2)$ with a norm $\|(f_{n+1,K^2})_{K^2}\|_{L^{p'}(\ell^2)} \le 1$ and look at

$$(5.6) \qquad \sum_K \langle |\Delta^2_{K^2,k_1} f_1|, 1_K \rangle \langle |\Delta^1_{K^1,k_2} f_2| \rangle_K \langle |\Delta^1_{K^1,k_3} f_3| \rangle_K \prod_{j=4}^n \langle |f_j| \rangle_K \langle f_{n+1,K^2} w \rangle_K.$$

There holds that

$$(5.7) \qquad \langle |\Delta^2_{K^2,k_1} f_1|, 1_K \rangle = \langle \Delta^2_{K^2,k_1} f_1, \varphi_{K^2,f_1} \rangle = \langle f_1, \Delta^2_{K^2,k_1} \varphi_{K^2,f_1} \rangle, \qquad |\varphi_{K^2,f_1}| \le 1_K.$$

We now get that (5.6) is less than $\|f_1 w_1\|_{L^{p_1}}$ multiplied by

$$\Big\| \sum_K \langle f_{n+1,K^2} w \rangle_K \langle |\Delta^1_{K^1,k_2} f_2| \rangle_K \langle |\Delta^1_{K^1,k_3} f_3| \rangle_K \prod_{j=4}^n \langle |f_j| \rangle_K \Delta^2_{K^2,k_1} \varphi_{K^2,f_1} w_1^{-1} \Big\|_{L^{p_1'}}.$$

We will now apply the lower square function estimate $\|g w_1^{-1}\|_{L^{p_1'}} \lesssim \|S^2_{\mathcal{D}^2}(g) w_1^{-1}\|_{L^{p_1'}}$, Lemma 5.1, with the weight $w_1^{-p_1'} \in A_\infty$ (see Lemma 3.6). Here we use the block form of

Remark 5.3. Using also that $|\Delta^2_{K^2,k_1}\varphi_{K^2,f_1}| \lesssim 1_K$ we get that the last norm is dominated by

$$\left\|\left(\sum_{K^2}\left(\sum_{K^1}\langle|f_{n+1,K^2}|w\rangle_K\langle|\Delta^1_{K^1,k_2}f_2|\rangle_K\langle|\Delta^1_{K^1,k_3}f_3|\rangle_K\prod_{j=4}^n\langle|f_j|\rangle_K 1_K\right)^2\right)^{\frac{1}{2}}w_1^{-1}\right\|_{L^{p_1'}}.$$

We still have cancellation to use in the form of the other two martingale differences and will continue the process.

We repeat the argument from above – this gives that the previous term is dominated by $\|f_2 w_2\|_{L^{p_2}}$ multiplied by

$$\left\|\left(\sum_{K^1}\left(\sum_{K^2}\langle|f_{n+1,K^2}|w\rangle_K\langle|f_{1,K^2}|w_1^{-1}\rangle_K\langle|\Delta^1_{K^1,k_3}f_3|\rangle_K\prod_{j=4}^n\langle|f_j|\rangle_K 1_K\right)^2\right)^{\frac{1}{2}}w_2^{-1}\right\|_{L^{p_2'}}$$

where $\|(f_{1,K^2})_{K^2}\|_{L^{p_1}(\ell^2)} \leq 1$. Running this argument one more time finally gives us that this is dominated by $\|f_3 w_3\|_{L^{p_3}}$ multiplied by

$$\left\|\left(\sum_{K^1}\left(\sum_{K^2}\langle|f_{n+1,K^2}|w\rangle_K\langle|f_{1,K^2}|w_1^{-1}\rangle_K\langle|f_{2,K^1}|w_2^{-1}\rangle_K\prod_{j=4}^n\langle|f_j|\rangle_K 1_K\right)^2\right)^{\frac{1}{2}}w_3^{-1}\right\|_{L^{p_3'}}$$

$$\leq \left\|\left(\sum_{K^1}\left(\sum_{K^2}M_{\mathcal{D}}(f_{n+1,K^2}w, f_{1,K^2}w_1^{-1}, f_{2,K^1}w_2^{-1}, f_4, \ldots, f_n)\right)^2\right)^{\frac{1}{2}}w_3^{-1}\right\|_{L^{p_3'}},$$

where $\|(f_{2,K^1})_{K^1}\|_{L^{p_2}(\ell^2)} \leq 1$.

Using Lemma 3.11 three times (we dualized three times) shows that

$$(w^{-1}, w_1, w_2, w_4, \ldots, w_n) \in A_{(p',p_1,p_2,p_4,\ldots,p_n)}.$$

The maximal function satisfies the weighted

$$L^{p'}(\ell^\infty_{K^1}(\ell^2_{K^2})) \times L^{p_1}(\ell^\infty_{K^1}(\ell^2_{K^2})) \times L^{p_2}(\ell^2_{K^1}(\ell^\infty_{K^2})) \times L^{p_4} \times \cdots \times L^{p_n} \to L^{p_3'}(\ell^2_{K^1}(\ell^1_{K^2}))$$

estimate. This gives that the last norm above is dominated by

$$\|(f_{n+1,K^2}ww^{-1})_{K^2}\|_{L^{p'}(\ell^2)}\|(f_{1,K^2}w_1^{-1}w_1)_{K^2}\|_{L^{p_1}(\ell^2)}\|(f_{2,K^1}w_2^{-1}w_2)_{K^1}\|_{L^{p_2}(\ell^2)}\prod_{i=4}^n\|f_i w_i\|_{L^{p_i}},$$

where the first three norms are ≤ 1. This concludes the proof for $A_{2,k}$ and the rest of the cases are similar. $\qquad\square$

We also record some linear estimates. We will need these when we deal with the most complicated model operators – the partial paraproducts.

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

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

Proof. By (3.5) we have for all $n \geq 2$ that

$$1 \leq \langle u\rangle_K\left\langle u^{-\frac{1}{n-1}}\right\rangle_K^{n-1}.$$

Simply using this we reduce to

$$\left\| \left[\sum_m \left(\sum_{K \in \mathcal{D}} \langle |\Delta_{K,k} f_m| \rangle_K^2 \left\langle u^{-\frac{1}{n-1}} \right\rangle_K^{2(n-1)} 1_K \right)^{\frac{s}{2}} \right]^{\frac{1}{s}} u^{\frac{1}{p}} \right\|_{L^p}$$

$$= \left\| \left[\sum_m A_{1,k} \left(f_m, u^{-\frac{1}{n-1}}, \ldots, u^{-\frac{1}{n-1}} \right)^s \right]^{\frac{1}{s}} u^{\frac{1}{p}} \right\|_{L^p},$$

where $A_{1,k}$ is a suitable square function as in Theorem 5.5.

We then fix n large enough so that $u \in A_n$. We then notice that this implies that

$$\text{(5.9)} \qquad \left(u^{-\frac{1}{p'}}, u^{\frac{1}{n-1}}, \ldots, u^{\frac{1}{n-1}} \right) \in A_{(p,\infty,\ldots,\infty)}.$$

To see this, notice that the target weight associated with this tuple is $u^{-\frac{1}{p'}} u = u^{\frac{1}{p}}$ and that the target exponent is p, and so

$$\left[\left(u^{-\frac{1}{p'}}, u^{\frac{1}{n-1}}, \ldots, u^{\frac{1}{n-1}} \right) \right]_{A_{(p,\infty,\ldots,\infty)}} = \sup_R \langle u \rangle_R^{1/p} \langle u \rangle_R^{1/p'} \left\langle u^{-\frac{1}{n-1}} \right\rangle_R^{n-1} = [u]_{A_n} < \infty.$$

It remains to use the weighted (with the weight (5.9)) vector-valued estimate $L^p(\ell^s) \times L^\infty \times \cdots \times L^\infty \to L^p(\ell^s)$ of $A_{1,k}$, which follows by Theorem 5.5 and (3.14). $\qquad \square$

5.10. *Remark.* It is possible to prove the above proposition also directly with the duality and lower square function strategy that was used in the proof of Theorem 5.5.

6. Dyadic model operators

In this section we are working with a fixed set of dyadic rectangles $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$. All the model operators depend on this lattice, but it is not emphasised in the notation.

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

$$\langle S_k(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k_j)} = K}} a_{K,(R_j)} \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{R_j} \rangle.$$

Here $K, R_1, \ldots, R_{n+1} \in \mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, $R_j = I_j^1 \times I_j^2$, $R_j^{(k_j)} := (I_j^1)^{(k_j^1)} \times (I_j^2)^{(k_j^2)}$ and $\widetilde{h}_{R_j} = \widetilde{h}_{I_j^1} \otimes \widetilde{h}_{I_j^2}$. Here we assume that for $m \in \{1, 2\}$ there exist two indices $j_0^m, j_1^m \in \{1, \ldots, n+1\}$, $j_0^m \neq j_1^m$, so that $\widetilde{h}_{I_{j_0^m}^m} = h_{I_{j_0^m}^m}$, $\widetilde{h}_{I_{j_1^m}^m} = h_{I_{j_1^m}^m}$ and for the remaining indices $j \notin \{j_0^m, j_1^m\}$ we have $\widetilde{h}_{I_j^m} \in \{h_{I_j^m}^0, h_{I_j^m}\}$. Moreover, $a_{K,(R_j)} = a_{K,R_1,\ldots,R_{n+1}}$ is a scalar satisfying the normalization

$$\text{(6.1)} \qquad |a_{K,(R_j)}| \leq \frac{\prod_{j=1}^{n+1} |R_j|^{1/2}}{|K|^n}.$$

6.2. **Theorem.** *Suppose S_k is an n-linear bi-parameter shift, $1 < p_1, \ldots, p_n \leq \infty$ 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.

Proof. We use duality to always reduce to one of the operators of type A_3 in Theorem 5.5. Performing the proof like this has the advantage that the form of the shift really plays no role – it just affects which type of A_3 operator we get. For example, we consider the explicit case

$$S_k(f_1, \ldots, f_n) = \sum_K A_K(f_1, \ldots, f_n),$$

where

$$A_K(f_1, \ldots, f_n) = \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k_j)} = K}} a_{K,(R_j)} \langle f_1, h_{R_1} \rangle \prod_{j=2}^{n} \langle f_j, \widetilde{h}_{R_j} \rangle h_{R_{n+1}}.$$

Fix some $\vec{p} = (p_1, \ldots, p_n)$ with $1 < p_i < \infty$ and $p > 1$, which is enough by extrapolation. We will dualise using f_{n+1} with $\|f_{n+1} w^{-1}\|_{L^{p'}} \leq 1$. The normalisation of the shift coefficients gives the direct estimate

$$\sum_K |\langle A_K(f_1, \ldots, f_n), f_{n+1} \rangle|$$

$$\leq \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k_j)} = K}} \frac{\prod_{j=1}^{n+1} |R_j|^{1/2}}{|K|^n} \left| \langle \Delta_{K,k_1} f_1, h_{R_1} \rangle \prod_{j=2}^{n} \langle f_j, \widetilde{h}_{R_j} \rangle \langle \Delta_{K,k_{n+1}} f_{n+1}, h_{R_{n+1}} \rangle \right|$$

$$\leq \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k_j)} = K}} \frac{1}{|K|^n} \langle |\Delta_{K,k_1} f_1|, 1_{R_1} \rangle \prod_{j=2}^{n} \langle |f_j|, 1_{R_j} \rangle \langle |\Delta_{K,k_{n+1}} f_{n+1}|, 1_{R_{n+1}} \rangle$$

$$\leq \sum_K \langle |\Delta_{K,k_1} f_1| \rangle_K \prod_{j=2}^{n} \langle |f_j| \rangle_K \langle |\Delta_{K,k_{n+1}} f_{n+1}| \rangle_K |K|$$

$$= \left\| \sum_K \langle |\Delta_{K,k_1} f_1| \rangle_K \prod_{j=2}^{n} \langle |f_j| \rangle_K \langle |\Delta_{K,k_{n+1}} f_{n+1}| \rangle_K 1_K \right\|_{L^1},$$

where we used (2.1) in the first step in the passage from Haar functions into martingale differences. Notice that

$$(6.3) \qquad (w_1, \cdots, w_n, w^{-1}) \in A_{(p_1, \cdots, p_n, p')}, \qquad w = \prod_{i=1}^{n} w_i.$$

The target weight associated to this data is $ww^{-1} = 1$ and the target exponent is $1/p + 1/p' = 1$. By using Theorem 5.5 with a suitable $A_3(f_1, \ldots, f_{n+1})$ and the above weight we can directly dominate this by

$$\left[\prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}} \right] \cdot \|f_{n+1} w^{-1}\|_{L^{p'}} \leq \prod_{i=1}^{n} \|f_i w_i\|_{L^{p_i}}.$$

We are done. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

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

(6.4) $\qquad \langle (S\pi)_k(f_1, \ldots, f_n), f_{n+1} \rangle = \displaystyle\sum_{K = K^1 \times K^2} \sum_{\substack{I_1^1, \ldots, I_{n+1}^1 \\ (I_j^1)^{(k_j)} = K^1}} a_{K,(I_j^1)} \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{I_j^1} \otimes u_{j,K^2} \rangle,$

where the functions $\widetilde{h}_{I_j^1}$ and u_{j,K^2} satisfy the following. There are $j_0, j_1 \in \{1, \ldots, n+1\}$, $j_0 \neq j_1$, so that $\widetilde{h}_{I_{j_0}^1} = h_{I_{j_0}^1}$, $\widetilde{h}_{I_{j_1}^1} = h_{I_{j_1}^1}$ and for the remaining indices $j \notin \{j_0, j_1\}$ we have $\widetilde{h}_{I_j^1} \in \{h_{I_j^1}^0, h_{I_j^1}\}$. There is $j_2 \in \{1, \ldots, n+1\}$ so that $u_{j_2,K^2} = h_{K^2}$ and for the remaining indices $j \neq j_2$ we have $u_{j,K^2} = \frac{1_{K^2}}{|K^2|}$. Moreover, the coefficients are assumed to satisfy

(6.5) $\qquad \|(a_{K,(I_j^1)})_{K^2}\|_{\mathrm{BMO}} = \displaystyle\sup_{K_0^2 \in \mathcal{D}^2} \Big(\frac{1}{|K_0^2|} \sum_{K^2 \subset K_0^2} |a_{K,(I_j^1)}|^2 \Big)^{1/2} \leq \frac{\prod_{j=1}^{n+1} |I_j^1|^{\frac{1}{2}}}{|K^1|^n}.$

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

The following H^1-BMO duality type estimate is well-known and elementary:

(6.6) $\qquad \displaystyle\sum_{K^2} |a_{K^2}||b_{K^2}| \lesssim \|(a_{K^2})\|_{\mathrm{BMO}} \Big\| \Big(\sum_{K^2} |b_{K^2}|^2 \frac{1_{K^2}}{|K^2|} \Big)^{1/2} \Big\|_{L^1}.$

Such estimates have natural multi-parameter analogues also, and the proofs in all parameters are analogous. See e.g. [44, Equation (4.1)].

Our result for the partial paraproducts has a significantly more difficult proof than for the other model operators. It is also more inefficient in that is produces an exponential – although crucially with an arbitrarily small exponent – dependence on the complexity. This has some significance for the required kernel regularity of CZOs, but a standard $t \mapsto t^\alpha$ type continuity modulus will still suffice.

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

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

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

Proof. Recall that $(S\pi)_k$ is of the form (6.4). Recall also the indices j_0 and j_1, which say that $\widetilde{h}_{I_j^1} = h_{I_j^1}$ at least for $j \in \{j_0, j_1\}$, and the index j_2, which specifies the place of h_{K^2} in the second parameter. It makes no difference for the argument what the indices j_0 and j_1 are, so we assume that $j_0 = 1$ and $j_1 = 2$. It makes a small difference whether $j_2 \in \{j_0, j_1\}$ or $j_2 \notin \{j_0, j_1\}$, so we do not specify j_2 yet. To make the following formulae shorter we write $\widetilde{h}_{I_j^1}$ for every j but keep in mind that these are cancellative at least for $j \in \{1, 2\}$. We define

$$A_{K^2}(g_1, \ldots, g_{n+1}) = \prod_{j=1}^{n+1} \langle g_j, u_{j,K^2} \rangle.$$

and write $(S\pi)_k$ in the form

$$\langle (S\pi)_k(f_1,\ldots,f_n),f_{n+1}\rangle = \sum_{K=K^1\times K^2} \sum_{\substack{I_1^1,\ldots,I_{n+1}^1 \\ (I_j^1)^{(k_j)}=K^1}} a_{K,(I_j^1)} A_{K^2}(\langle f_1,\widetilde{h}_{I_1^1}\rangle_1,\ldots,\langle f_{n+1},\widetilde{h}_{I_{n+1}^1}\rangle_1).$$

Fix some $\vec{p}=(p_1,\ldots,p_n)$ with $1<p_i<\infty$ and $p>1$, which is enough by extrapolation. We will dualise using f_{n+1} with $\|f_{n+1}w^{-1}\|_{L^{p'}}\le 1$. We may assume $f_j\in L_c^\infty$. The H^1-BMO duality (6.6) gives that

(6.8)
$$|\langle (S\pi)_k(f_1,\ldots,f_n),f_{n+1}\rangle| \lesssim \sum_{K^1}\sum_{\substack{I_1^1,\ldots,I_{n+1}^1 \\ (I_j^1)^{(k_j)}=K^1}} \left[\frac{\prod_{j=1}^{n+1}|I_j^1|^{\frac{1}{2}}}{|K^1|^n}\right.$$
$$\left.\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}|A_{K^2}(\langle f_1,\widetilde{h}_{I_1^1}\rangle_1,\ldots,\langle f_{n+1},\widetilde{h}_{I_{n+1}^1}\rangle_1)|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac{1}{2}}\right].$$

Suppose $j\in\{3,\ldots,n+1\}$ is such that $\widetilde{h}_{I_j^1}=h_{I_j^1}^0$ and $k_j>0$, that is, we have non-cancellative Haar functions and non-zero complexity. We expand

$$|I_j^1|^{-\frac{1}{2}}\langle f_j,h_{I_j^1}^0\rangle_1 = \langle f\rangle_{I_j^1,1} = \langle f_j\rangle_{K^1,1} + \sum_{i_j=1}^{k_j}\langle \Delta_{(I_j^1)^{(i_j)}}^1 f_j\rangle_{(I_j^1)^{(i_j-1)},1}.$$

For convenience, we further write that

$$\langle \Delta_{(I_j^1)^{(i_j)}}^1 f_j\rangle_{(I_j^1)^{(i_j-1)},1} = \langle h_{(I_j^1)^{(i_j)}}\rangle_{(I_j^1)^{(i_j-1)}}\langle f_j,h_{(I_j^1)^{(i_j)}}\rangle_1,$$

where we are suppressing the summation over the $2^{d_1}-1$ different Haar functions. We perform these expansions inside the operators A_{K^2}, and take the sums out of the $\ell_{K^2}^2$ norm. This gives that the right hand side of (6.8) is less than a sum of at most $\prod_{j=3}^n(1+k_j)$ terms of the form

$$\sum_{K^1}\sum_{\substack{I_1^1,\ldots,I_{n+1}^1 \\ (I_j^1)^{(k_j)}=K^1}} \left[\frac{\prod_{j=1}^{n+1}|I_j^1||(I_j^1)^{(i_j)}|^{-\frac{1}{2}}}{|K^1|^n}\right.$$
$$\left.\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}|A_{K^2}(\langle f_1,\widetilde{h}_{(I_1^1)^{(i_1)}}\rangle_1,\ldots,\langle f_{n+1},\widetilde{h}_{(I_{n+1}^1)^{(i_{n+1})}}\rangle_1)|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac{1}{2}}\right].$$

Here we have the following properties. If j in an index such that we did not do the expansion related to j, then $i_j=0$. Thus, at least $i_1=i_2=0$. We also remind that $\widetilde{h}_{(I_j^1)^{(i_j)}}=h_{(I_j^1)^{(i_j)}}$ for $j=1,2$. If $i_j<k_j$, then $\widetilde{h}_{(I_j^1)^{(i_j)}}=h_{(I_j^1)^{(i_j)}}$. If $i_j=k_j$, then $\widetilde{h}_{(I_j^1)^{(i_j)}}\in\{h_{K^1},h_{K^1}^0\}$. We can further rewrite this as

(6.9) $$\sum_{K^1}\sum_{\substack{L_1^1,\ldots,L_{n+1}^1 \\ (L_j^1)^{(l_j)}=K^1}} \frac{\prod_{j=1}^{n+1}|L_j^1|^{\frac{1}{2}}}{|K^1|^n}\int_{\mathbb{R}^{d_2}}\Big(\sum_{K^2}|A_{K^2}(\langle f_1,\widetilde{h}_{L_1}\rangle_1,\ldots,\langle f_{n+1},\widetilde{h}_{L_{n+1}}\rangle_1)|^2\frac{1_{K^2}}{|K^2|}\Big)^{\frac{1}{2}}.$$

This is otherwise analogous to the right hand side of (6.8) except for the key difference that if a non-cancellative Haar function appears, then the related complexity is zero.

We turn to estimate (6.9). We show that

$$(6.10) \qquad (6.9) \lesssim_\beta 2^{\max_j k_j \frac{\beta}{2}} \Big[\prod_{j=1}^n \|f_j w_j\|_{L^{p_j}} \Big] \|f_{n+1} w^{-1}\|_{L^{p'}}.$$

Recalling that $\|f_{n+1} w^{-1}\|_{L^{p'}} \leq 1$ this implies that the left hand side of (6.8) satisfies

$$LHS(6.8) \lesssim_\beta (1 + \max_j k_j)^{n-1} 2^{\max_j k_j \frac{\beta}{2}} \prod_{j=1}^n \|f_j w_j\|_{L^{p_j}} \lesssim_\beta 2^{\max_j k_j \beta} \prod_{j=1}^n \|f_j w_j\|_{L^{p_j}},$$

which proves the theorem.

Let $(v_1, \ldots, v_{n+1}) \in A_{(2,\ldots,2)}$ and $v = \prod_{j=1}^{n+1} v_j$. We will prove the $(n+1)$-linear estimate

$$(6.11)$$
$$\left\| \sum_{K^1} \sum_{\substack{L_1^1, \ldots, L_{n+1}^1 \\ (L_j^1)^{(l_j)} = K^1}} \left[\frac{\prod_{j=1}^{n+1} |L_j^1|^{\frac{1}{2}}}{|K^1|^n} \frac{1_{K^1}}{|K^1|} \right. \right.$$
$$\left. \left. \left(\sum_{K^2} |A_{K^2}(\langle f_1, \widetilde{h}_{L_1} \rangle_1, \ldots, \langle f_{n+1}, \widetilde{h}_{L_{n+1}} \rangle_1)|^2 \frac{1_{K^2}}{|K^2|} \right)^{\frac{1}{2}} \right] v \right\|_{L^{\frac{2}{n+1}}} \lesssim 2^{\max_j k_j \frac{\beta}{2}} \prod_{j=1}^{n+1} \|f_j v_j\|_{L^2}.$$

Extrapolation, Theorem 3.12, then gives that

$$\left\| \sum_{K^1} \sum_{\substack{L_1^1, \ldots, L_{n+1}^1 \\ (L_j^1)^{(l_j)} = K^1}} \frac{\prod_{j=1}^{n+1} |L_j^1|^{\frac{1}{2}}}{|K^1|^n} \frac{1_{K^1}}{|K^1|} \left(\sum_{K^2} |A_{K^2}(\langle f_1, \widetilde{h}_{L_1} \rangle_1, \ldots, \langle f_{n+1}, \widetilde{h}_{L_{n+1}} \rangle_1)|^2 \frac{1_{K^2}}{|K^2|} \right)^{\frac{1}{2}} v \right\|_{L^q}$$

$$\lesssim 2^{\max_j k_j \frac{\beta}{2}} \prod_{j=1}^{n+1} \|f_j v_j\|_{L^{q_j}}$$

for all $q_1, \ldots, q_{n+1} \in (1, \infty]$ such that $\frac{1}{q} = \sum_{j=1}^{n+1} \frac{1}{q_j} > 0$ and for all $(v_1, \ldots, v_{n+1}) \in A_{(q_1, \ldots, q_{n+1})}$. Applying this with the exponent tuple $(p_1, \ldots, p_n, p')$ and the weight tuple $(w_1, \ldots, w_n, w^{-1}) \in A_{(p_1, \ldots, p_n, p')}$ gives (6.10).

It remains to prove (6.11). We denote $\sigma_j = v_j^{-2}$. The $A_{(2,\ldots,2)}$ condition gives that

$$\langle v^{\frac{2}{n+1}} \rangle_K^{n+1} \prod_{j=1}^{n+1} \langle \sigma_j \rangle_K \lesssim 1.$$

Using this we have

$$|A_{K^2}(\langle f_1, \widetilde{h}_{L_1^1} \rangle_1, \ldots, \langle f_{n+1}, \widetilde{h}_{L_{n+1}^1} \rangle_1)| \lesssim \frac{1}{\langle v^{\frac{2}{n+1}} \rangle_K^{n+1}} \left| A_{K^2} \left(\frac{\langle f_1, \widetilde{h}_{L_1^1} \rangle_1}{\langle \sigma_1 \rangle_K}, \ldots, \frac{\langle f_{n+1}, \widetilde{h}_{L_{n+1}^1} \rangle_1}{\langle \sigma_{n+1} \rangle_K} \right) \right|.$$

For the moment we abbreviate the last $|A_{K^2}(\cdots)|$ as $c_{K,(L_j^1)}$. There holds that

$$
\frac{1}{\langle v^{\frac{2}{n+1}} \rangle_K^{n+1}} c_{K,(L_j^1)} = \left[\frac{1}{\langle v^{\frac{2}{n+1}} \rangle_K} \left\langle c_{K,(L_j^1)}^{\frac{1}{n+1}} 1_K v^{-\frac{2}{n+1}} v^{\frac{2}{n+1}} \right\rangle_K \right]^{n+1}
$$

$$
\leq \left(M_{\mathcal{D}}^{v^{\frac{2}{n+1}}} \left(c_{K,(L_j^1)}^{\frac{1}{n+1}} 1_K v^{-\frac{2}{n+1}} \right)(x) \right)^{n+1}
$$

for all $x \in K$.

We substitute this into the left hand side of (6.11). This gives that the term there is dominated by

$$
\left\| \sum_{K^1} \sum_{\substack{L_1^1,\ldots,L_{n+1}^1 \\ (L_j^1)^{(l_j)}=K^1}} \frac{\prod_{j=1}^{n+1} |L_j^1|^{\frac{1}{2}}}{|K^1|^{n+1}} \left(\sum_{K^2} M_{\mathcal{D}}^{v^{\frac{2}{n+1}}} \left(c_{K,(L_j^1)}^{\frac{1}{n+1}} 1_K v^{-\frac{2}{n+1}} \right)^{2(n+1)} \frac{1}{|K^2|} \right)^{\frac{1}{2}} v \right\|_{L^{\frac{2}{n+1}}}.
$$

We use the $L^2(\ell_{K^1,(L_j^1)}^{n+1}(\ell_{K^2}^{2(n+1)}))$ boundedness of the maximal function $M_{\mathcal{D}}^{v^{\frac{2}{n+1}}}$, see Proposition 4.3. This gives that the last norm is dominated by

$$
(6.12) \qquad \left\| \sum_{K^1} \sum_{\substack{L_1^1,\ldots,L_{n+1}^1 \\ (L_j^1)^{(l_j)}=K^1}} \frac{\prod_{j=1}^{n+1} |L_j^1|^{\frac{1}{2}}}{|K^1|^{n+1}} 1_{K^1} \left(\sum_{K^2} c_{K,(L_j^1)}^2 \frac{1_{K^2}}{|K^2|} \right)^{\frac{1}{2}} v^{-1} \right\|_{L^{\frac{2}{n+1}}}.
$$

Now, we recall what the numbers $c_{K,(L_j^1)}$ are. At this point it becomes relevant which of the Haar functions $\widetilde{h}_{L_j^1}$ are cancellative and what is the form of the operators A_{K^2}. We assume that $\widetilde{h}_{L_j^1} = h_{L_j^1}$ for $j = 1,\ldots,n$ and $\widetilde{h}_{L_{n+1}^1} = h_{L_{n+1}^1}^0 = h_{K_{n+1}^1}^0$, which is a good representative of the general case. First, we assume that the index j_2, which specifies the place of h_{K^2} in A_{K^2}, satisfies $j_2 \in \{1,\ldots,n\}$. The point is that then $\widetilde{h}_{L_{j_2}^1} = h_{L_{j_2}^1}$. For convenience of notation we assume that $j_2 = 1$. With these assumptions there holds that

$$
(6.13) \qquad c_{K,(L_j^1)} = \left| \frac{\langle f_1, h_{L_1^1} \otimes h_{K^2} \rangle}{\langle \sigma_1 \rangle_K} \prod_{j=2}^{n} \frac{\left\langle f_j, h_{L_j^1} \otimes \frac{1_{K^2}}{|K^2|} \right\rangle}{\langle \sigma_j \rangle_K} \cdot \frac{\left\langle f_{n+1}, h_{K^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle}{\langle \sigma_{n+1} \rangle_K} \right|.
$$

For $j = 2,\ldots,n$ we estimate that

$$
(6.14) \qquad \frac{\left| \left\langle f_j, h_{L_j^1} \otimes \frac{1_{K^2}}{|K^2|} \right\rangle \right|}{\langle \sigma_j \rangle_K} = \frac{\left| \left\langle \langle f_j, h_{L_j^1} \rangle_1 \langle \sigma_j \rangle_{K^1,1}^{-1} \langle \sigma_j \rangle_{K^1,1}, \frac{1_{K^2}}{|K^2|} \right\rangle \right|}{\langle \langle \sigma_j \rangle_{K^1,1} \rangle_{K^2}}
$$

$$
\leq M_{\mathcal{D}^2}^{\langle \sigma_j \rangle_{K^1,1}} (\langle f_j, h_{L_j^1} \rangle_1 \langle \sigma_j \rangle_{K^1,1}^{-1})(x_2)
$$

for all $x_2 \in K^2$. Also, there holds that

$$
\frac{\left| \left\langle f_{n+1}, h_{K^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle \right|}{\langle \sigma_{n+1} \rangle_K} \leq |K^1|^{\frac{1}{2}} M_{\mathcal{D}}^{\sigma_{n+1}} (f_{n+1} \sigma_{n+1}^{-1})(x)
$$

for all $x \in K$. These give (recall that $L^1_{n+1} = K^1$) that

$$\sum_{\substack{L^1_1,\ldots,L^1_{n+1} \\ (L^1_j)^{(l_j)}=K^1}} \frac{\prod_{j=1}^{n+1} |L^1_j|^{\frac{1}{2}}}{|K^1|^{n+1}} 1_{K^1} \Big(\sum_{K^2} c^2_{K,(L^1_j)} \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}} \leq \prod_{j=1}^{n} F_{j,K^1} \cdot M^{\sigma_{n+1}}_{\mathcal{D}}(f_{n+1}\sigma^{-1}_{n+1}),$$

where

$$(6.15) \qquad F_{1,K^1} = 1_{K^1} \sum_{(L^1_1)^{(l_1)}=K^1} \frac{|L^1_1|^{\frac{1}{2}}}{|K^1|} \Big(\sum_{K^2} \frac{|\langle f_1, h_{L^1_1} \otimes h_{K^2} \rangle|^2}{\langle \sigma_1 \rangle^2_K} \frac{1_{K^2}}{|K^2|} \Big)^{\frac{1}{2}}$$

and

$$(6.16) \qquad F_{j,K^1} = 1_{K^1} \sum_{(L^1_j)^{(l_j)}=K^1} \frac{|L^1_j|^{\frac{1}{2}}}{|K^1|} M^{\langle \sigma_j \rangle_{K^{1,1}}}_{\mathcal{D}^2}(\langle f_j, h_{L^1_j} \rangle_1 \langle \sigma_j \rangle^{-1}_{K^1,1})$$

for $j = 2, \ldots, n$.

We will now continue from (6.12) using the above pointwise estimates. Notice that

$$\sum_{K^1} \prod_{j=1}^{n} F_{j,K^1} \leq \prod_{j=1}^{2} \Big(\sum_{K^1} (F_{j,K^1})^2 \Big)^{1/2} \prod_{j=3}^{n} \sup_{K^1} F_{j,K^1} \leq \prod_{j=1}^{n} \Big(\sum_{K^1} (F_{j,K^1})^2 \Big)^{1/2}.$$

Since $v^{-1} = \prod_{j=1}^{n+1} v^{-1}_j$, we have that

$$(6.12) \lesssim \prod_{j=1}^{n} \Big\| \Big(\sum_{K^1} F^2_{j,K^1} \Big)^{\frac{1}{2}} v^{-1}_j \Big\|_{L^2} \big\| M^{\sigma_{n+1}}_{\mathcal{D}}(f_{n+1}\sigma^{-1}_{n+1}) v^{-1}_{n+1} \big\|_{L^2}.$$

Since $\sigma_j = v^{-2}_j$ there holds by Proposition 4.2 that

$$\| M^{\sigma_{n+1}}_{\mathcal{D}}(f_{n+1}\sigma^{-1}_{n+1}) v^{-1}_{n+1} \|_{L^2} = \| M^{\sigma_{n+1}}_{\mathcal{D}}(f_{n+1}\sigma^{-1}_{n+1}) \|_{L^2(\sigma_{n+1})} \lesssim \| f_{n+1} v_{n+1} \|_{L^2}.$$

It remains to estimate the norms for $j = 1, \ldots, n$.

We begin with $j = 1$. If $l_1 = 0$, then we directly have that

$$\Big(\sum_{K^1} F^2_{1,K^1} \Big)^{\frac{1}{2}} = \Big(\sum_{K} \frac{|\langle f_1, h_K \rangle|^2}{\langle \sigma_1 \rangle^2_K} \frac{1_K}{|K|} \Big)^{\frac{1}{2}}.$$

Since $|\langle f_1, h_K \rangle||K|^{-\frac{1}{2}} \leq \langle |\Delta_K f_1| \rangle_K$, we may use Proposition 5.8 to have that

$$\Big\| \Big(\sum_{K^1} F^2_{1,K^1} \Big)^{\frac{1}{2}} v^{-1}_1 \Big\|_{L^2} \lesssim \| f_1 \sigma^{-\frac{1}{2}}_1 \|_{L^2} = \| f_1 v_1 \|_{L^2}.$$

Suppose then $l_1 > 0$. There holds that

$$\Big\| \Big(\sum_{K^1} F^2_{1,K^1} \Big)^{\frac{1}{2}} v^{-1}_1 \Big\|_{L^2} = \Big(\sum_{K^1} \| F_{1,K^1} v^{-1}_1 \|^2_{L^2} \Big)^{\frac{1}{2}}.$$

Let $s \in (1, \infty)$ be such that $d_1/s' = \beta/(2n)$. Then

$$F_{1,K^1} \leq 2^{\frac{l_1 \beta}{2n}} 1_{K^1} \Big(\sum_{(L^1_1)^{(l_1)}=K^1} \frac{|L^1_1|^{\frac{s}{2}}}{|K^1|^s} \Big(\sum_{K^2} \frac{|\langle f_1, h_{L^1_1} \otimes h_{K^2} \rangle|^2}{\langle \sigma_1 \rangle^2_K} \frac{1_{K^2}}{|K^2|} \Big)^{\frac{s}{2}} \Big)^{\frac{1}{s}}.$$

Therefore, $\|F_{1,K^1}v_j^{-1}\|_{L^2}^2$ is less than

$$(6.17) \qquad 2^{\frac{l_1\beta}{n}}\left\|\left(\sum_{(L_1^1)^{(l_1)}=K^1}\frac{|L_1^1|^{\frac{s}{2}}}{|K^1|^s}\left(\sum_{K^2}\frac{|\langle f_1, h_{L_1^1}\otimes h_{K^2}\rangle|^2}{\langle\sigma_1\rangle_K^2}\frac{1_{K^2}}{|K^2|}\right)^{\frac{s}{2}}\right)^{\frac{1}{s}}\langle\sigma_1\rangle_{K^1,1}^{\frac{1}{2}}\right\|_{L^2}^2|K^1|.$$

Notice that $|\langle f_1, h_{L_1^1}\otimes h_{K^2}\rangle||K^2|^{-\frac{1}{2}}\leq\langle|\Delta_{K^2}\langle f_1, h_{L_1^1}\rangle_1|\rangle_{K^2}$. Therefore, the one-parameter case of Proposition 5.8 gives that

$$(6.17)\lesssim 2^{\frac{l_1\beta}{n}}\left\|\left(\sum_{(L_1^1)^{(l_1)}=K^1}\frac{|L_1^1|^{\frac{s}{2}}}{|K^1|^s}|\langle f_1, h_{L_1^1}\rangle_1|^s\right)^{\frac{1}{s}}\langle\sigma_1\rangle_{K^1,1}^{-\frac{1}{2}}\right\|_{L^2}^2|K^1|$$

$$(6.18)$$

$$\leq 2^{\frac{l_1\beta}{n}}\left\|\sum_{(L_1^1)^{(l_1)}=K^1}\frac{|L_1^1|^{\frac{1}{2}}}{|K^1|}|\langle f_1, h_{L_1^1}\rangle_1|\langle\sigma_1\rangle_{K^1,1}^{-\frac{1}{2}}\right\|_{L^2}^2|K^1|.$$

Notice that

$$\sum_{(L_1^1)^{(l_1)}=K^1}\frac{|L_1^1|^{\frac{1}{2}}}{|K^1|}|\langle f_1, h_{L_1^1}\rangle_1|\leq\langle|\Delta_{K^1,l_1}^1 f_1|\rangle_{K^1,1}.$$

Thus, summing the right hand side of (6.18) over K^1 leads to

$$2^{\frac{l_1\beta}{n}}\int_{\mathbb{R}^{d_2}}\sum_{K^1}\langle|\Delta_{K^1,l_1}^1 f_1|\rangle_{K^1,1}^2\langle\sigma_1\rangle_{K^1,1}^{-1}|K^1|=2^{\frac{l_1\beta}{n}}\int_{\mathbb{R}^d}\sum_{K^1}\langle|\Delta_{K^1,l_1}^1 f_1|\rangle_{K^1,1}^2\frac{1_{K^1}}{\langle\sigma_1\rangle_{K^1,1}^2}\sigma_1$$

$$\lesssim 2^{\frac{l_1\beta}{n}}\int_{\mathbb{R}^d}|f_1|^2 v_1^2,$$

where we used Proposition 5.8 again.

Finally, we estimate the norms related to $j=2,\ldots,n$, which are all similar. We assume that $l_j>0$. It will be clear how to do the case $l_j=0$. As above we have

$$F_{j,K^1}\leq 2^{\frac{l_j\beta}{2n}}1_{K^1}\left(\sum_{(L_j^1)^{(l_j)}=K^1}\frac{|L_j^1|^{\frac{s}{2}}}{|K^1|^s}M_{\mathcal{D}^2}^{\langle\sigma_j\rangle_{K^1,1}}(\langle f_j, h_{L_j^1}\rangle_1\langle\sigma_j\rangle_{K^1,1}^{-1})^s\right)^{\frac{1}{s}}.$$

Therefore, we get that

$$\|F_{j,K^1}v_j^{-1}\|_{L^2}^2\leq 2^{\frac{l_j\beta}{n}}\left\|\left(\sum_{(L_j^1)^{(l_j)}=K^1}\frac{|L_j^1|^{\frac{s}{2}}}{|K^1|^s}M_{\mathcal{D}^2}^{\langle\sigma_j\rangle_{K^1,1}}(\langle f_j, h_{L_j^1}\rangle_1\langle\sigma_j\rangle_{K^1,1}^{-1})^s\right)^{\frac{1}{s}}\langle\sigma_j\rangle_{K^1,1}^{\frac{1}{2}}\right\|_{L^2}^2|K^1|$$

$$\lesssim 2^{\frac{l_j\beta}{n}}\left\|\left(\sum_{(L_j^1)^{(l_j)}=K^1}\frac{|L_j^1|^{\frac{s}{2}}}{|K^1|^s}|\langle f_j, h_{L_j^1}\rangle_1\langle\sigma_j\rangle_{K^1,1}^{-1}|^s\right)^{\frac{1}{s}}\langle\sigma_j\rangle_{K^1,1}^{\frac{1}{2}}\right\|_{L^2}^2|K^1|$$

$$\leq 2^{\frac{l_j\beta}{n}}\left\|\sum_{(L_j^1)^{(l_j)}=K^1}\frac{|L_j^1|^{\frac{1}{2}}}{|K^1|}|\langle f_j, h_{L_j^1}\rangle_1|\langle\sigma_j\rangle_{K^1,1}^{-\frac{1}{2}}\right\|_{L^2}^2|K^1|,$$

where we applied the one-parameter version of Proposition 4.3. The last norm is like the last norm in (6.18), and therefore the estimate can be concluded with familiar steps.

Combining the estimates we have shown that

$$\prod_{j=1}^{n}\Big\|\Big(\sum_{K^1}F_{j,K^1}^2\Big)^{\frac{1}{2}}v_j^{-1}\Big\|_{L^2}\lesssim\prod_{j=1}^{n}2^{\frac{l_j\beta}{2n}}\|f_jv_j\|_{L^2}\leq 2^{\max k_j\frac{\beta}{2}}\prod_{j=1}^{n}\|f_jv_j\|_{L^2}.$$

Above, we assumed that the index j_2 related to the form of the paraproduct satisfied $j_2 = 1$, see the discussion before (6.13). It remains to comment on the case $j_2 = n + 1$. In this case the formula corresponding to (6.13) is

$$c_{K,(L_j^1)}=\bigg|\prod_{j=1}^{n}\frac{\big\langle f_j, h_{L_j^1}\otimes\frac{1_{K^2}}{|K^2|}\big\rangle}{\langle\sigma_j\rangle_K}\cdot\frac{\langle f_{n+1}, h_{K^1}^0\otimes h_{K^2}\rangle}{\langle\sigma_{n+1}\rangle_K}\bigg|.$$

For $j = 1,\ldots,n$ we do the estimate (6.14). Also, there holds that

$$\frac{|\langle f_{n+1}, h_{K^1}^0\otimes h_{K^2}\rangle|}{\langle\sigma_{n+1}\rangle_K}=|K^1|^{\frac{1}{2}}\frac{\big|\big\langle\langle f_{n+1}, h_{K^2}\rangle_2\langle\sigma_{n+1}\rangle_{K^2,2}^{-1}\langle\sigma_{n+1}\rangle_{K^2,2}\big\rangle_{K^1}\big|}{\langle\langle\sigma_{n+1}\rangle_{K^2,2}\rangle_{K^1}}$$

$$\leq |K^1|^{\frac{1}{2}}M_{\mathcal{D}^1}^{\langle\sigma_{n+1}\rangle_{K^2,2}}(\langle f_{n+1}, h_{K^2}\rangle_2\langle\sigma_{n+1}\rangle_{K^2,2}^{-1})(x_1)$$

for any $x_1\in K^1$. With the pointwise estimates we proceed as above. Related to f_j, $j = 1,\ldots,n$, this leads to terms which we know how to estimate. Related to f_{n+1} we get a similar term except that the parameters are in opposite roles. We are done. □

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

$$(6.19)\qquad\langle\Pi(f_1,\ldots,f_n), f_{n+1}\rangle=\sum_{K=K^1\times K^2}a_K\prod_{j=1}^{n+1}\langle f_j, u_{j,K^1}\otimes u_{j,K^2}\rangle,$$

where the functions u_{j,K^1} and u_{j,K^2} are like in (6.4). The coefficients are assumed to satisfy

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

where the supremum is over open sets $\Omega\subset\mathbb{R}^d=\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$ with $0 < |\Omega| < \infty$. As already discussed the H^1-BMO duality works also in bi-parameter (see again [44, Equation (4.1)]):

$$(6.20)\qquad\sum_{K}|a_K||b_K|\lesssim\|(a_K)\|_{\mathrm{BMO_{prod}}}\Big\|\Big(\sum_{K}|b_K|^2\frac{1_K}{|K|}\Big)^{1/2}\Big\|_{L^1}.$$

We are ready to bound the full paraproducts.

6.21. **Theorem.** *Suppose Π is an n-linear bi-parameter full paraproduct, $1 < p_1,\ldots,p_n\leq\infty$ 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 $\vec{w}\in A_{\vec{p}}$.

Proof. We use duality to always reduce to one of the operators of type A_1 in Theorem 5.5. Fix some $\vec{p} = (p_1, \ldots, p_n)$ with $1 < p_i < \infty$ and $p > 1$, which is enough by extrapolation. We will dualise using f_{n+1} with $\|f_{n+1}w^{-1}\|_{L^{p'}} \leq 1$. The particular form of Π does not matter – it only affects the form of the operator A_1 we will get. We may, for example, look at

$$\Pi(f_1, \ldots, f_n) = \sum_{K=K^1 \times K^2} a_K \Big\langle f_1, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big\langle f_2, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2} \Big\rangle \prod_{j=3}^{n} \langle f_j \rangle_K \cdot \frac{1_K}{|K|}.$$

We have

$$|\langle \Pi(f_1, \ldots, f_n), f_{n+1} \rangle| \leq \sum_K |a_K| \Big| \Big\langle f_1, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big\langle f_2, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2} \Big\rangle \Big| \prod_{j=3}^{n+1} \langle |f_j| \rangle_K.$$

We now apply the unweighted H^1-BMO duality estimate from above to bound this with

$$\Big\| \Big(\sum_K \langle |\Delta_{K^1}^1 f_1| \rangle_K^2 \langle |\Delta_{K^2}^2 f_2| \rangle_K^2 \prod_{j=3}^{n+1} \langle |f_j| \rangle_K^2 1_K \Big)^{\frac{1}{2}} \Big\|_{L^1}.$$

Recalling (6.3) it remains to apply Theorem 5.5 with a suitable $A_1(f_1, \ldots, f_{n+1})$. $\square$

7. SINGULAR INTEGRALS

Let ω be a modulus of continuity: an increasing and subadditive function with $\omega(0) = 0$. A relevant quantity is the modified Dini condition

$$(7.1) \qquad \|\omega\|_{\mathrm{Dini}_\alpha} := \int_0^1 \omega(t) \Big(1 + \log \frac{1}{t}\Big)^\alpha \frac{dt}{t}, \qquad \alpha \geq 0.$$

In practice, the quantity (7.1) arises as follows:

$$(7.2) \qquad \sum_{k=1}^\infty \omega(2^{-k}) k^\alpha = \sum_{k=1}^\infty \frac{1}{\log 2} \int_{2^{-k}}^{2^{-k+1}} \omega(2^{-k}) k^\alpha \frac{dt}{t} \lesssim \int_0^1 \omega(t) \Big(1 + \log \frac{1}{t}\Big)^\alpha \frac{dt}{t}.$$

We define what it means to be an n-linear bi-parameter SIO. Let $\mathscr{F}_{d_i}$ denote the space of finite linear combinations of indicators of cubes in $\mathbb{R}^{d_i}$, and let $\mathscr{F}$ denote the space of finite linear combinations of indicators of rectangles in $\mathbb{R}^d$. Suppose that we have n-linear operators $T_{1,2}^{j_1*, j_2*}$, $j_1, j_2 \in \{0, \ldots, n\}$, each mapping $\mathscr{F} \times \cdots \times \mathscr{F}$ into locally integrable functions. We denote $T = T^{0*, 0*}$ and assume that the operators $T_{1,2}^{j_1*, j_2*}$ satisfy the duality relations as described in Section 2.C.

Let ω_i be a modulus of continuity on $\mathbb{R}^{d_i}$. Assume $f_j = f_j^1 \otimes f_j^2$, $j = 1, \ldots, n+1$, where $f_j^i \in \mathscr{F}_{d_i}$.

Bi-parameter SIOs.

Full kernel representation. Here we assume that given $m \in \{1, 2\}$ there exist $j_1, j_2 \in \{1, \ldots, n+1\}$ so that $\mathrm{spt}\, f_{j_1}^m \cap \mathrm{spt}\, f_{j_2}^m = \emptyset$. In this case we demand that

$$\langle T(f_1, \ldots, f_n), f_{n+1} \rangle = \int_{\mathbb{R}^{(n+1)d}} K(x_{n+1}, x_1, \ldots, x_n) \prod_{j=1}^{n+1} f_j(x_j) \, dx,$$

where
$$K \colon \mathbb{R}^{(n+1)d} \setminus \{(x_1, \ldots, x_{n+1}) \in \mathbb{R}^{(n+1)d} \colon x_1^1 = \cdots = x_{n+1}^1 \text{ or } x_1^2 = \cdots = x_{n+1}^2\} \to \mathbb{C}$$
is a kernel satisfying a set of estimates which we specify next.

The kernel K is assumed to satisfy the size estimate
$$|K(x_{n+1}, x_1, \ldots, x_n)| \lesssim \prod_{m=1}^{2} \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^m - x_j^m| \right)^{d_m n}}.$$

We also require the following continuity estimates. For example, we require that we have
$$|K(x_{n+1}, x_1, \ldots, x_n) - K(x_{n+1}, x_1, \ldots, x_{n-1}, (c^1, x_n^2))$$
$$- K((x_{n+1}^1, c^2), x_1, \ldots, x_n) + K((x_{n+1}^1, c^2), x_1, \ldots, x_{n-1}, (c^1, x_n^2))|$$
$$\lesssim \omega_1 \Big(\frac{|x_n^1 - c^1|}{\sum_{j=1}^{n} |x_{n+1}^1 - x_j^1|} \Big) \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^1 - x_j^1| \right)^{d_1 n}}$$
$$\times \omega_2 \Big(\frac{|x_{n+1}^2 - c^2|}{\sum_{j=1}^{n} |x_{n+1}^2 - x_j^2|} \Big) \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^2 - x_j^2| \right)^{d_2 n}}$$

whenever $|x_n^1 - c^1| \leq 2^{-1} \max_{1 \leq i \leq n} |x_{n+1}^1 - x_i^1|$ and $|x_{n+1}^2 - c^2| \leq 2^{-1} \max_{1 \leq i \leq n} |x_{n+1}^2 - x_i^2|$. 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 $(n + 1)^2$ different estimates.

Finally, we require the following mixed continuity and size estimates. For example, we ask that
$$|K(x_{n+1}, x_1, \ldots, x_n) - K(x_{n+1}, x_1, \ldots, x_{n-1}, (c^1, x_n^2))|$$
$$\lesssim \omega_1 \Big(\frac{|x_n^1 - c^1|}{\sum_{j=1}^{n} |x_{n+1}^1 - x_j^1|} \Big) \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^1 - x_j^1| \right)^{d_1 n}} \cdot \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^2 - x_j^2| \right)^{d_2 n}}$$

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

Partial kernel representations. Suppose now only that there exist $j_1, j_2 \in \{1, \ldots, n + 1\}$ so that $\operatorname{spt} f_{j_1}^1 \cap \operatorname{spt} f_{j_2}^1 = \emptyset$. Then we assume that
$$\langle T(f_1, \ldots, f_n), f_{n+1} \rangle = \int_{\mathbb{R}^{(n+1)d_1}} K_{(f_j^2)}(x_{n+1}^1, x_1^1, \ldots, x_n^1) \prod_{j=1}^{n+1} f_j^1(x_j^1) \, \mathrm{d}x^1,$$

where $K_{(f_j^2)}$ is a one-parameter ω_1-Calderón–Zygmund kernel but with a constant depending on the fixed functions $f_1^2, \ldots, f_{n+1}^2$. For example, this means that the size estimate takes the form
$$|K_{(f_j^2)}(x_{n+1}^1, x_1^1, \ldots, x_n^1)| \leq C(f_1^2, \ldots, f_{n+1}^2) \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^1 - x_j^1| \right)^{d_1 n}}.$$

The continuity estimates are analogous.

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

$$C(1_{I^2}, \ldots, 1_{I^2}) \lesssim |I^2|$$

and

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

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

Analogous partial kernel representation on the second parameter is assumed when $\operatorname{spt} f_{j_1}^2 \cap \operatorname{spt} f_{j_2}^2 = \emptyset$ for some j_1, j_2.

7.4. **Definition.** If T is an n-linear operator with full and partial kernel representations as defined above, we call T an n-linear bi-parameter (ω_1, ω_2)-SIO.

Bi-parameter CZOs. We say that T satisfies the weak boundedness property if

$$|\langle T(1_R, \ldots, 1_R), 1_R \rangle| \lesssim |R|$$

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 $a_{I^i} \in \mathscr{F}_{d_i}$ with $a_{I^i} = 1_{I^i} a_{I^i}$, $|a_{I^i}| \leq 1$ and $\int a_{I^i} = 0$ we have

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

and

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

An SIO T satisfies the product BMO assumption if it holds

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

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

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

where 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 $\langle S(1, \ldots, 1), h_R \rangle$ can be defined, in a natural way, using the kernel representations.

7.5. **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 $((\omega_1, \omega_2)$-CZO$)$.

Dyadic representation theorem. In Section 6 we have introduced the three different dyadic model operators (DMOs). In this section we explain how and why the DMOs are linked to the CZOs. Before stating the known representation theorem, to aid the reader, we first outline the main structure and idea of representation theorems – for the lenghty details in this generality see [2].

Step 1. There is a natural probability space $\Omega = \Omega_1 \times \Omega_2$, the details of which are not relevant for us here (but see [31]), 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 starting point is the martingale difference decomposition

$$\langle T(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_{j_1, j_2 = 1}^{n+1} \mathbb{E}_\sigma \Sigma_{j_1, j_2, \sigma} + \mathbb{E}_\sigma \operatorname{Rem}_\sigma,$$

where

$$\Sigma_{j_1, j_2, \sigma} = \sum_{\substack{R_1, \ldots, R_{n+1} \\ \ell(I_{i_1}^1) > \ell(I_{j_1}^1) \text{ for } i_1 \neq j_1 \\ \ell(I_{i_2}^2) > \ell(I_{j_2}^2) \text{ for } i_2 \neq j_2}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle$$

and $R_1 = I_1^1 \times I_1^2, \ldots, R_{n+1} = I_{n+1}^1 \times I_{n+1}^2 \in \mathcal{D}_\sigma = \mathcal{D}_{\sigma_1} \times \mathcal{D}_{\sigma_2}$. Notice how we have already started the proof working parameter by parameter. At this point the randomization is not yet important: it is used at a later point in the proof to find suitable common parents for dyadic cubes. Looking at the definition of shifts this is clearly critical: everything is organised under the common parent K and they cannot be arbitrarily large.

Step 2. There are $(n+1)^2$ main terms $\Sigma_{j_1, j_2, \sigma}$ – these are similar to each other and all of them produce shifts, partial paraproducts and *exactly one* full paraproduct. For example, a further parameter by parameter $T1$ style decomposition of $\Sigma_{n, n+1, \sigma}$ produces the full paraproduct

$$\sum_{R = K^1 \times K^2} \langle T(1, \ldots, 1, h_{K^1} \otimes 1), \quad 1 \otimes h_{K^2} \rangle \prod_{j=1}^{n-1} \langle f_j \rangle_R \Big\langle f_n, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big\langle f_{n+1}, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2} \Big\rangle,$$

where

$$\langle T(1, \ldots, 1, h_{K^1} \otimes 1), 1 \otimes h_{K^2} \rangle = \langle T_1^{n*}(1, \ldots, 1), h_R \rangle.$$

For the quantitative part the product BMO assumption of T_1^{n*} is critical here, and the remaining product BMO assumptions are used to control the full paraproducts coming from the other main terms.

The shifts and partial paraproducts structurally arise from the $T1$ decomposition combined with probability. Again, the randomization is simply used to find suitably sized common parents. After this completely structural part (for full details see [2]), the focus is on providing estimates for the coefficients, like the coefficient $a_{K, (R_j)}$ of the shifts. *As in the full paraproduct case above, it is important to understand that the coefficients always have a concrete form in terms of pairings involving T and various Haar functions.* These pairings are estimated in various ways:

- The shift coefficients are handled with kernel estimates only (various size and continuity estimates). Only in the part $\operatorname{Rem}_\sigma$, which we have not yet discussed, also the weak boundedness is used to handle the diagonal case, where kernel estimates are not valid.
- In the partial paraproduct case a size estimate does not suffice, as there is not enough cancellation. A more refined BMO estimate needs to be proved – this is done via a duality argument. This duality is the source of the atoms a_I appearing in some of the assumptions – e.g. in (7.3).

Step 3. The final step is to deal with the remainder $\operatorname{Rem}_\sigma$. This only produces shifts and partial paraproducts. Another difference to the main terms is that all the diagonal

parts of the summation are here – to deal with them we need to assume the weak bound-edness property and the diagonal BMO assumptions.

7.6. *Remark.* An m-parameter representation theorem is structurally identical: the pairing $\langle T(f_1, \ldots, f_n), f_{n+1} \rangle$ is split into $(n+1)^m$ main terms and the remainder. These are then further split into shifts, partial paraproducts and full paraproducts. The full paraproduct is produced parameter by parameter, as it is in the bi-parameter case, and this produces partial paraproducts, where the paraproduct component can vary from being 1-parameter to being $(m-1)$-parameter. The definition of a CZO is adjusted so that all of the appearing coefficients of the appearing model operators involving T and Haar functions can be estimated. For the linear m-parameter representation theorem see Ou [47] – this establishes the appropriate definition of a multi-parameter CZO. We discuss the m-parameter case in more detail in Section 9. The point there is the following: while the representation theorem itself is straightforward, some of the estimates of Section 6 are harder in m-parameter.

In the paper [2], among other things, a dyadic representation theorem for n-linear bi-parameter CZOs was proved. The minimal regularity required is $\omega_i \in \text{Dini}_{\frac{1}{2}}$, but then the dyadic representation is in terms of certain modified versions of the model operators we have presented, and bounded, in this paper. It appears to be difficult to prove weighted bounds for the modified operators with the optimal dependency on the complexity. Instead, we will rely on a lemma, which says that all of the modified operators can be written as suitable sums of the standard model operators. This step essentially outright loses $\frac{1}{2}$ of kernel regularity, and puts us in competition to obtain our weighted bounds with $\omega_i \in \text{Dini}_1$. The bilinear bi-parameter representation theorem with the usual Hölder type kernel regularity $w_i(t) = t^{\alpha_i}$ appeared first in [40]. We now state a representation theorem that we will rely on.

A consequence of [2, Theorem 5.35] and [2, Lemma 5.12] is the following.

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

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

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:

$$(7.8) \qquad\qquad\qquad U_u = C \sum_{i_1=0}^{u_1-1} \sum_{i_2=0}^{u_2-1} V_{i_1,i_2},$$

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

$$k_{j,V}^m \leq u_m.$$

7.9. *Remark.* We assumed that the operator T and its adjoints are initially well-defined for finite linear combinations of indicators of rectangles. However, a careful proof of the representation theorem [40] shows that this implies the boundedness of T (for related details see also [21] and [30]). Therefore, we do not need to worry about this detail any more at this point and we can work with general functions. Moreover, we do not need

to work with the CZOs directly – after the representation theorem we only need to work with the dyadic model operators.

Weighted estimates for CZOs. In this paper we were able to prove a complexity free weighted estimate for the shifts. On the contrary, the weighted estimate for the partial paraproducts is even exponential, however, with an arbitrarily small power. For these reasons, we can prove a weighted estimate with mild kernel regularity for paraproduct free T, and otherwise we will deal with the standard kernel regularity $\omega_i(t) = t^{\alpha_i}$. 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 (only the partial paraproducts, and not the full paraproducts, are problematic in terms of kernel regularity, of course). In the paraproduct free case the reader can think of convolution form SIOs.

7.10. **Theorem.** *Suppose T is an n-linear bi-parameter (ω_1, ω_2)-CZO. For $1 < p_1, \ldots, p_n \le \infty$ and $1/p = \sum_i 1/p_i > 0$ we have*

$$\|T(f_1, \ldots, f_n)w\|_{L^p} \lesssim \prod_i \|f_i w_i\|_{L^{p_i}}$$

for all multilinear bi-parameter weights $\vec{w} \in A_{\vec{p}}$, if one of the following conditions hold.

(1) T is paraproduct free and $\omega_i \in \mathrm{Dini}_1$.
(2) We have $\omega_i(t) = t^{\alpha_i}$ for some $\alpha_i \in (0, 1]$.

Proof. Notice that in the paraproduct free case (1) by our results for the shifts we always have

$$\|U_{u,\sigma}(f_1, \ldots, f_n)w\|_{L^p} \lesssim (1 + u_1)(1 + u_2) \prod_i \|f_i w_i\|_{L^{p_i}},$$

where the complexity dependency comes only from the decomposition (7.8). We then take some $1 < p_1, \ldots, p_n < \infty$ with $p \in (1, \infty)$, use the dyadic representation theorem and conclude that T satisfies the weighted bound with these fixed exponents – recall (7.2) and that $\omega_i \in \mathrm{Dini}_1$. Finally, we extrapolate using Theorem 3.12.

The case of a completely general CZO with the standard kernel regularity is proved completely analogously. Just choose the exponent β in the exponential complexity dependency of the partial paraproducts to be small enough compared to α_1 and α_2. $\square$

8. EXTRAPOLATION

This section is devoted to providing more details about Theorem 3.12 in the multi-parameter setting. We also obtain the proof of the vector-valued Proposition 4.3. We give the details in the bi-parameter case with the general case being similar.

We begin with the following definitions. Given $\mu \in A_\infty(\mathbb{R}^{n+m})$, we say $w \in A_p(\mu)$ if $w > 0$ a.e. and

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

And we say $w \in A_1(\mu)$ if $w > 0$ a.e. and

$$[w]_{A_1(\mu)} = \sup_R \langle w \rangle_R^\mu \operatorname{ess\,sup}_R w^{-1} < \infty.$$

We begin with the following auxiliary result needed to build the required machinery. This is an extension of Proposition 4.2.

8.1. *Remark.* The so-called three lattice theorem states that there are lattices $\mathcal{D}_j^m$ in $\mathbb{R}^{d_m}$, $m \in \{1, 2\}$, $j \in \{1, \ldots, 3^{d_m}\}$, such that for every cube $Q^m \subset \mathbb{R}^{d_m}$ there exists a j and $I^m \in \mathcal{D}_j^m$ so that $Q^m \subset I^m$ and $|I^m| \sim |Q^m|$. Given $\lambda \in A_\infty$ we have in particular that λ is doubling: $\lambda(2R) \lesssim \lambda(R)$ for all rectangles R. It then follows that also the non-dyadic variant M^λ satisfies Proposition 4.2.

8.2. **Lemma.** *Let $\mu \in A_\infty$ and $w \in A_p(\mu)$, $1 < p < \infty$. Then we have*
$$\|M^\mu f\|_{L^p(w\mu)} \lesssim \|f\|_{L^p(w\mu)}.$$

Proof. Fix x and $f \geq 0$ and denote $\sigma = w^{-\frac{1}{p-1}}$. For an arbitrary rectangle $R \subset \mathbb{R}^d$ with $x \in R$ we have
$$\langle f \rangle_R^\mu = \langle \sigma \rangle_R^\mu \left(\langle w \rangle_R^\mu \right)^{\frac{1}{p-1}} \left(\langle w \rangle_R^\mu \right)^{-\frac{1}{p-1}} \frac{1}{\sigma\mu(R)} \int_R f\mu$$
$$\leq [w]_{A_p(\mu)}^{\frac{1}{p-1}} \left(M^{w\mu} \big([M^{\sigma\mu}(f\sigma^{-1})]^{p-1} w^{-1} \big)(x) \right)^{\frac{1}{p-1}}.$$

The idea of the above pointwise estimate is from [34]. If $w\mu, \sigma\mu \in A_\infty$, then by (the non-dyadic version of) Proposition 4.2 we have
$$\|M^\mu f\|_{L^p(w\mu)} \lesssim \left\| \left(M^{w\mu} \big([M^{\sigma\mu}(f\sigma^{-1})]^{p-1} w^{-1} \big) \right)^{\frac{1}{p-1}} \right\|_{L^p(w\mu)}$$
$$\lesssim \left\| \left([M^{\sigma\mu}(f\sigma^{-1})]^{p-1} w^{-1} \right)^{\frac{1}{p-1}} \right\|_{L^p(w\mu)}$$
$$= \|M^{\sigma\mu}(f\sigma^{-1})\|_{L^p(\sigma\mu)} \lesssim \|f\|_{L^p(w\mu)}.$$

Therefore, it remains to check that $w\mu, \sigma\mu \in A_\infty$. We only explicitly prove that $w\mu \in A_\infty$, since the other one is symmetric. First of all, write
$$\langle w \rangle_R^\mu \left(\langle w^{-\frac{1}{p-1}} \rangle_R^\mu \right)^{p-1} \leq [w]_{A_p(\mu)}$$
in the form
$$\langle w\mu \rangle_R \langle w^{-\frac{1}{p-1}} \mu \rangle_R^{p-1} \leq [w]_{A_p(\mu)} \langle \mu \rangle_R^p.$$
Then by the Lebesgue differentiation theorem, we have for all cubes $I^1 \subset \mathbb{R}^{d_1}$ that
$$\langle w\mu \rangle_{I^1,1}(x_2) \langle w^{-\frac{1}{p-1}} \mu \rangle_{I^1,1}^{p-1}(x_2) \leq [w]_{A_p(\mu)} \langle \mu \rangle_{I^1,1}^p(x_2), \quad x_2 \in \mathbb{R}^{d_2} \setminus N_{I^1},$$
where $|N_{I^1}| = 0$. By standard considerations there exists N so that $|N| = 0$ and for all cubes $I^1 \subset \mathbb{R}^{d_1}$ we have
$$\langle w\mu \rangle_{I^1,1}(x_2) \langle w^{-\frac{1}{p-1}} \mu \rangle_{I^1,1}^{p-1}(x_2) \leq [w]_{A_p(\mu)} \langle \mu \rangle_{I^1,1}^p(x_2), \quad x_2 \in \mathbb{R}^{d_2} \setminus N.$$
In other words, $w(\cdot, x_2) \in A_p(\mu(\cdot, x_2))$ (uniformly) for all $x_2 \in \mathbb{R}^m \setminus N$. As $\mu \in A_\infty$ there exists $s < \infty$ so that $\mu \in A_s$. Then for all cubes $I^1 \subset \mathbb{R}^{d_1}$ and arbitrary $E \subset I^1$ we have
$$\frac{|E|}{|I^1|} \lesssim \left(\frac{\mu(\cdot, x_2)(E)}{\mu(\cdot, x_2)(I^1)} \right)^{\frac{1}{s}} \lesssim \left(\frac{w\mu(\cdot, x_2)(E)}{w\mu(\cdot, x_2)(I^1)} \right)^{\frac{1}{ps}}, \quad \text{a.e. } x_2 \in \mathbb{R}^{d_2},$$
where the implicit constant is independent from x_2. This means $w\mu(\cdot, x_2) \in A_\infty(\mathbb{R}^{d_1})$ uniformly for a.e. $x_2 \in \mathbb{R}^{d_2}$. Likewise we can show that $w\mu(x_1, \cdot) \in A_\infty(\mathbb{R}^{d_2})$ uniformly for a.e. $x_1 \in \mathbb{R}^{d_1}$. This completes the proof and we are done. $\square$

Now we are ready to formulate the following version of Rubio de Francia algorithm.

8.3. Lemma. *Let $\mu \in A_\infty$ and $p \in (1, \infty)$. Let f be a non-negative function in $L^p(w\mu)$ for some $w \in A_p(\mu)$. Let M_k^μ be the k-th iterate of M^μ, $M_0^\mu f = f$, and $\|M^\mu\|_{L^p(w\mu)} := \|M^\mu\|_{L^p(w\mu) \to L^p(w\mu)}$ be the norm of M^μ as a bounded operator on $L^p(w\mu)$. Define*

$$Rf(x) = \sum_{k=0}^{\infty} \frac{M_k^\mu f}{(2\|M^\mu\|_{L^p(w\mu)})^k}.$$

Then $f(x) \le Rf(x)$, $\|Rf\|_{L^p(w\mu)} \le 2\|f\|_{L^p(w\mu)}$, and Rf is an $A_1(\mu)$ weight with constant $[Rf]_{A_1(\mu)} \le 2\|M^\mu\|_{L^p(w\mu)}$.

Proof. The statements $f(x) \le Rf(x)$ and $\|Rf\|_{L^p(w\mu)} \le 2\|f\|_{L^p(w\mu)}$ are obvious. Since

$$M^\mu(Rf) \le \sum_{k=0}^{\infty} \frac{M_{k+1}^\mu f}{(2\|M^\mu\|_{L^p(w\mu)})^k} \le 2\|M^\mu\|_{L^p(w\mu)} Rf,$$

we have

$$[Rf]_{A_1(\mu)} \le \sup_R \left(\inf_R M^\mu(Rf) \right) \left(\operatorname{ess\,inf}_R Rf \right)^{-1} \le 2\|M^\mu\|_{L^p(w\mu)}.$$

We are done. $\qquad\square$

With the above Rubio de Francia algorithm at hand, we are able to prove the bi-parameter version of [37, Theorem 3.1] and the corresponding endpoint cases similarly as in [38, Theorem 2.3]. On the other hand, the key technical lemma [38, Lemma 2.14] can be extended to the bi-parameter setting very easily. Using these as in [38] we obtain Theorem 3.12.

The above Rubio de Francia algorithm, of course, also yields the following standard linear extrapolation. Let $\mu \in A_\infty$ and assume that

(8.4) $$\|g\|_{L^{p_0}(w\mu)} \lesssim \|f\|_{L^{p_0}(w\mu)}$$

for all $w \in A_{p_0}(\mu)$. Then the same inequality holds for all $p \in (1, \infty)$ and $w \in A_p(\mu)$. Using this and Lemma 8.2 we obtain Proposition 4.3 via the following standard argument.

Proof of Proposition 4.3. From Lemma 8.2 we directly have that

$$\left\| \left(\sum_i (M^\mu f_j^i)^s \right)^{\frac{1}{s}} \right\|_{L^s(w\mu)} \lesssim \left\| \left(\sum_i |f_j^i|^s \right)^{\frac{1}{s}} \right\|_{L^s(w\mu)}, \quad w \in A_s(\mu).$$

Then, the extrapolation described around (8.4) gives that

$$\left\| \left(\sum_i (M^\mu f_j^i)^s \right)^{\frac{1}{s}} \right\|_{L^t(w\mu)} \lesssim \left\| \left(\sum_i |f_j^i|^s \right)^{\frac{1}{s}} \right\|_{L^t(w\mu)}, \quad w \in A_t(\mu).$$

This in turn gives

$$\left\| \left(\sum_j \left(\sum_i (M^\mu f_j^i)^s \right)^{\frac{t}{s}} \right)^{\frac{1}{t}} \right\|_{L^t(w\mu)} \lesssim \left\| \left(\sum_j \left(\sum_i |f_j^i|^s \right)^{\frac{t}{s}} \right)^{\frac{1}{t}} \right\|_{L^t(w\mu)}, \quad w \in A_t(\mu).$$

Extrapolating once more concludes the proof. $\qquad\square$

9. THE MULTI-PARAMETER CASE

One can approach the multi-parameter case as follows.

(1) What is the definition of an SIO/CZO? The important base case is the linear multi-parameter definition given in [47]. That can be straightforwardly extended to the multilinear situation as in Section 7. The definition becomes extremely lengthy due to the large number of different partial kernel representations, and for this reason we do not write it down explicitly. However, there is no complication in combining the linear multi-parameter definition [47] and our multilinear bi-parameter definition.

   We mention that another way to define the operators would be to adapt a Journé [32] style definition – this kind of vector-valued definition would be shorter to state. In this paper we do not use the Journé style formulation. However, for the equivalence of the Journé style definitions and the style we use here see [20, 40, 47].

(2) Is there a representation theorem in this generality? Yes – see Remark 7.6. The linear multi-parameter representation theorem is proved in [47]. The multilinear representation theorems [2, 40] are stated only in the bi-parameter setting for convenience. However, using the multilinear methods from [2, 40] the multi-parameter theorem [47] can easily be generalised to the multilinear setting.

(3) How do the model operators look like? Studying the above presented bi-parameter model operators, one realises that the philosophies in each parameter are independent of each other – for example, if one has a shift type of philosophy in a given parameter, one needs at least two cancellative Haar functions in that parameter. With this logic it is clear how to define the m-parameter analogues just by working parameter by parameter. Alternatively, one can take all possible m-fold tensor products of one-parameter n-linear model operators, and then just replace the appearing product coefficients by general coefficients. This yields the form of the model operators.

   We demonstrate this with an example of a bilinear tri-parameter partial paraproduct. Tri-parameter partial paraproducts have the shift structure in one or two of the parameters. In the remaining parameters there is a paraproduct structure. The following is an example of a partial paraproduct with the shift structure in the first parameter and the paraproduct structure in the second and third parameters:

$$\sum_{K=K^1\times K^2\times K^3\in\mathcal{D}} \sum_{\substack{I_1^1,I_2^1,I_3^1\in\mathcal{D}^1 \\ (I_j^1)^{(k_j)}=K^1}} \Big[a_{K,(I_j^1)}$$

$$\Big\langle f_1, h_{I_1^1}\otimes\frac{1_{K^2}}{|K^2|}\otimes\frac{1_{K^3}}{|K^3|}\Big\rangle\Big\langle f_2, h_{I_2^1}^0\otimes\frac{1_{K^2}}{|K^2|}\otimes h_{K^3}\Big\rangle\Big\langle f_3, h_{I_3^1}\otimes h_{K^2}\otimes\frac{1_{K^3}}{|K^3|}\Big\rangle\Big].$$

   Here $\mathcal{D}=\mathcal{D}^1\times\mathcal{D}^2\times\mathcal{D}^3$. The assumption on the coefficients is that when K^1, I_1^1, I_2^1 and I_3^1 are fixed, then

$$\|(a_{K,(I_j^1)})_{K^2\times K^3\in\mathcal{D}^2\times\mathcal{D}^3}\|_{\mathrm{BMO}_{\mathrm{prod}}}\leq\frac{|I_1^1|^{\frac{1}{2}}|I_2^1|^{\frac{1}{2}}|I_3^1|^{\frac{1}{2}}}{|K^1|^2}.$$

In the shift parameter there is at least two cancellative Haar functions and in the paraproduct parameters there is exactly one cancellative Haar and the remaining functions are normalised indicators. Thus, this is a generalization of $S \otimes \pi \otimes \pi$, where S is a one-parameter shift and π is a one-parameter paraproduct – and all model operators arise like this.

(4) Finally, is it more difficult to show the genuinely multilinear weighted estimates for the m-parameter, $m \geq 3$, model operators compared to the bi-parameter model operators? When it comes to shifts and full paraproducts, there is no essential difference – their boundedness always reduces to Theorem 5.5, which has an obvious m-parameter version. With out current proof, the answer for the partial paraproducts is more complicated. Thus, we will elaborate on how to prove the weighted estimates for m-parameter partial paraproducts. Notice that previously e.g. in [40] we could only handle bi-parameter partial paraproducts as our proof exploited the one-parameter nature of the paraproducts by sparse domination. Here we have already disposed of sparse domination, but the proof is still complicated and leads to some new philosophies in higher parameters.

We now discuss how to prove a tri-parameter analogue of Theorem 6.7. We can have a partial paraproduct with a bi-parameter paraproduct component and a one-parameter shift component, or the other way around. Regardless of the form, the initial stages of the proof of Theorem 6.7 can be used to reduce to estimating the weighted L^2 norms of certain functions which are analogous to (6.15) and (6.16). Most of these norms can be estimated with similar steps as in the bi-parameter case. However, also a new type of variant appears. An example of such a variant is given by

$$(9.1) \quad F_{j,K^1} = 1_{K^1} \sum_{(L_j^1)^{(l_j)} = K^1} \frac{|L_j^1|^{\frac{1}{2}}}{|K^1|} \Big(\sum_{K^2} \frac{1_{K^2}}{|K^2|} M_{\mathcal{D}^3}^{\langle \sigma_j \rangle_{K^{1,2}}} \big(\langle f_j, h_{L_j^1} \otimes h_{K^2} \rangle \langle \sigma_j \rangle_{K^{1,2}}^{-1} \big)^2 \Big)^{\frac{1}{2}},$$

where $f_j \colon \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3} \to \mathbb{C}$. The goal is to estimate $\sum_{K^1} \|F_{j,K^1}\|_{L^2(\sigma_j)}^2$. Here we are denoting $K^{1,2} = K^1 \times K^2$ the original tri-parameter rectangle being $K = K^1 \times K^2 \times K^3$, and for brevity we only write $\langle \sigma_j \rangle_{K^{1,2}}$ instead of $\langle \sigma_j \rangle_{K^{1,2},1,2}$.

Comparing with (6.16), the key difference is that in (6.16) the measure of the maximal function depended only on K^1. Here, it depends also on K^2, and therefore we have maximal functions with respect to different measures inside the norms. We will use the following lemma and the appearing new type of extrapolation trick to overcome this.

9.2. **Lemma.** *Let $\mu \in A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ be a bi-parameter weight. Let $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$ be a grid of bi-parameter dyadic rectangles in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Suppose that for each $m \in \mathbb{Z}$ and $K^1 \in \mathcal{D}^1$ we have a function $f_{m,K^1} \colon \mathbb{R}^{d_2} \to \mathbb{C}$. Then, for all $p, s, t \in (1, \infty)$, the estimate*

$$\Big\| \Big(\sum_m \Big(\sum_{K^1} 1_{K^1} M_{\mathcal{D}^2}^{\langle \mu \rangle_{K^1}} (f_{m,K^1})^t \Big)^{\frac{s}{t}} \Big)^{\frac{1}{s}} \Big\|_{L^p(w\mu)} \lesssim \Big\| \Big(\sum_m \Big(\sum_{K^1} 1_{K^1} |f_{m,K^1}|^t \Big)^{\frac{s}{t}} \Big)^{\frac{1}{s}} \Big\|_{L^p(w\mu)}$$

holds for all $w \in A_p(\mu)$.

Proof. By extrapolation, see the discussion around (8.4), it suffices to take a function $f \colon \mathbb{R}^{d_2} \to \mathbb{C}$ and show that

$$\big\| 1_{K^1} M_{\mathcal{D}^2}^{\langle \mu \rangle_{K^1}} f \big\|_{L^q(w\mu)} \lesssim \|1_{K^1} f\|_{L^q(w\mu)}$$

for some $q \in (1, \infty)$ and for all $w \in A_q(\mu)$. We fix some $w \in A_q(\mu)$. The above estimate can be rewritten as

$$\big\| M_{\mathcal{D}^2}^{\langle\mu\rangle_{K^1}} f \big\|_{L^q(\langle w\mu\rangle_{K^1})} |K^1|^{\frac{1}{q}} \lesssim \|f\|_{L^q(\langle w\mu\rangle_{K^1})} |K^1|^{\frac{1}{q}}.$$

We have the identity

(9.3) $$\langle w\mu\rangle_{K^1}(x_2) = \frac{\langle w\mu\rangle_{K^1}(x_2)}{\langle\mu\rangle_{K^1}(x_2)} \langle\mu\rangle_{K^1}(x_2) = \langle w(\cdot, x_2)\rangle_{K^1}^{\mu(\cdot, x_2)} \langle\mu\rangle_{K^1}(x_2).$$

Define $v(x_2) = \langle w(\cdot, x_2)\rangle_{K^1}^{\mu(\cdot, x_2)}$. We show that $v \in A_q(\langle\mu\rangle_{K^1})$. Let I^2 be a cube in $\mathbb{R}^{d_2}$. First, we have that

$$\int_{I^2} v\langle\mu\rangle_{K^1} = \int_{I^2} \langle w(\cdot, x_2)\rangle_{K^1}^{\mu(\cdot, x_2)} \langle\mu(\cdot, x_2)\rangle_{K^1} \, \mathrm{d}x_2 = \int_{K^1 \times I^2} w\mu |K^1|^{-1}.$$

Therefore, $\langle v\rangle_{I^2}^{\langle\mu\rangle_{K^1}} = \langle w\rangle_{K^1 \times I^2}^{\mu}$. Hölder's inequality gives that

$$\big(\langle w(\cdot, x_2)\rangle_{K^1}^{\mu(\cdot, x_2)}\big)^{-\frac{1}{q-1}} \leq \big\langle w(\cdot, x_2)^{-\frac{1}{q-1}}\big\rangle_{K^1}^{\mu(\cdot, x_2)},$$

which shows that

$$\int_{I^2} v^{-\frac{1}{q-1}} \langle\mu\rangle_{K^1} \leq \int_{I^2} \big\langle w(\cdot, x_2)^{-\frac{1}{q-1}}\big\rangle_{K^1}^{\mu(\cdot, x_2)} \langle\mu(\cdot, x_2)\rangle_{K^1} \, \mathrm{d}x_2 = \int_{K^1 \times I^2} w^{-\frac{1}{q-1}} \mu |K^1|^{-1}.$$

Thus, we have that

$$\big(\langle v^{-\frac{1}{q-1}}\rangle_{I^2}^{\langle\mu\rangle_{K^1}}\big)^{q-1} \leq \big(\langle w^{-\frac{1}{q-1}}\rangle_{K^1 \times I^2}^{\mu}\big)^{q-1}.$$

These estimates yield that $[v]_{A_q(\langle\mu\rangle_{K^1})} \leq [w]_{A_q(\mu)}$.

Recall the identity (9.3). Since $\langle\mu\rangle_{K^1} \in A_\infty$, we have that

$$\big\| M_{\mathcal{D}^2}^{\langle\mu\rangle_{K^1}} f \big\|_{L^q(\langle w\mu\rangle_{K^1})} = \big\| M_{\mathcal{D}^2}^{\langle\mu\rangle_{K^1}} f \big\|_{L^q(v\langle\mu\rangle_{K^1})} \lesssim \|f\|_{L^q(v\langle\mu\rangle_{K^1})} = \|f\|_{L^q(\langle w\mu\rangle_{K^1})},$$

where we used Lemma 8.2. This concludes the proof. $\qquad\square$

We now show how to estimate (9.1). First, we have that $\|F_{j,K^1}\|_{L^2(\sigma_j)}^2$ is less than $2^{\frac{2l_j^1 d_1}{s'}}$ multiplied by

$$\left\| \bigg[\sum_{(L_j^1)^{(l_j)} = K^1} \bigg[\frac{|L_j^1|^{\frac{1}{2}}}{|K^1|} \bigg(\sum_{K^2} \frac{1_{K^2}}{|K^2|} M_{\mathcal{D}^3}^{\langle\sigma_j\rangle_{K^{1,2}}} \big(\langle f_j, h_{L_j^1} \otimes h_{K^2}\rangle \langle\sigma_j\rangle_{K^{1,2}}^{-1} \big)^2 \bigg)^{\frac{1}{2}} \bigg]^s \bigg]^{\frac{1}{s}} \right\|_{L^2(\langle\sigma_j\rangle_{K^1})}^2 |K^1|.$$

The exponent $s \in (1, \infty)$ is chosen small enough so that we get a suitable dependence on the complexity through $2^{\frac{2l_j^1 d_1}{s'}}$, see the corresponding step in the bi-parameter case. Since $\langle\sigma_j\rangle_{K^1} \in A_\infty(\mathbb{R}^{d_2} \times \mathbb{R}^{d_3})$, we can use Lemma 9.2 to have that the last term is dominated by

$$\left\| \bigg[\sum_{(L_j^1)^{(l_j)} = K^1} \bigg[\frac{|L_j^1|^{\frac{1}{2}}}{|K^1|} \bigg(\sum_{K^2} \frac{1_{K^2}}{|K^2|} \big| \langle f_j, h_{L_j^1} \otimes h_{K^2}\rangle \langle\sigma_j\rangle_{K^{1,2}}^{-1} \big|^2 \bigg)^{\frac{1}{2}} \bigg]^s \bigg]^{\frac{1}{s}} \right\|_{L^2(\langle\sigma_j\rangle_{K^1})}^2 |K^1|.$$

After these key steps it only remains to use Proposition 5.8 twice in a very similar way as in the bi-parameter proof. We are done.

10. Applications

10.A. Mixed-norm estimates. With our main result, Theorem 1.2, and extrapolation, Theorem 3.12, the following result becomes immediate.

10.1. Theorem. *Let T be an n-linear m-parameter Calderón-Zygmund operator. Let $1 < p_i^j \leq \infty$, $i = 1, \ldots, n$, with $\frac{1}{p^j} = \sum_i \frac{1}{p_i^j} > 0$, $j = 1, \ldots, m$. Then we have that*

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

10.2. Remark. We understand this as an a priori estimate with $f_i \in L_c^\infty$ – this is only a concern when some p_i^j is ∞. In [38], which concerned the bilinear bi-parameter case with *tensor* form CZOs, we went to great lengths to check that this restriction can always be removed. We do not want to get into such considerations here, and prefer this a priori interpretation at least when $n \geq 3$. See also [40] for some previous results for bilinear bi-parameter CZOs that are not of tensor form, but where, compared to [38], the range of exponents had some limitations in the ∞ cases. See also [14].

We also mention that mixed-norm estimates for multilinear bi-parameter Coifman-Meyer operators have been previously obtained in [4] and [5]. Related to this, bi-parameter mixed norm Leibniz rules were proved in [46].

The proof is immediate by extrapolating with tensor form weights. For the general idea see [38, Theorem 4.5] – here the major simplification is that everything can be done with extrapolation and the operator-valued analysis is not needed. This is because the weighted estimate, Theorem 1.2, is now with the genuinely multilinear weights unlike in [38, 40].

10.B. Commutators. We will state these applications in the bi-parameter case $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. The m-parameter versions are obvious. We define

$$[b, T]_k(f_1, \ldots, f_n) := bT(f_1, \ldots, f_n) - T(f_1, \ldots, f_{k-1}, bf_k, f_{k+1}, \ldots, f_n).$$

One can also define the iterated commutators as usual. We say that $b \in \mathrm{bmo}$ if

$$\|b\|_{\mathrm{bmo}} = \sup_R \frac{1}{|R|} \int_R |b - \langle b \rangle_R| < \infty,$$

where the supremum is over rectangles. Recall that given $b \in \mathrm{bmo}$, we have

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

See e.g. [28]. In the one-parameter case the following was proved in [12, Lemma 5.6].

10.4. Proposition. *Let $\vec{p} = (p_1, \ldots, p_n)$ with $1 < p_1, \ldots, p_n < \infty$ and $\frac{1}{p} = \sum_i \frac{1}{p_i} < 1$. Let $\vec{w} = (w_1, \ldots, w_n) \in A_{\vec{p}}$. Then for any $1 \leq j \leq n$ we have*

$$\vec{w}_{b,z} := (w_1, \ldots, w_j e^{\mathrm{Re}(bz)}, \ldots, w_n) \in A_{\vec{p}}$$

with $[\vec{w}_{b,z}]_{A_{\vec{p}}} \lesssim [\vec{w}]_{A_{\vec{p}}}$ provided that

$$|z| \leq \frac{\epsilon}{\max([w^p]_{A_\infty}, \max_i [w_i^{-p_i'}]_{A_\infty}) \|b\|_{\mathrm{BMO}}},$$

where ϵ depends on $\vec{p}$ and the dimension of the underlying space.

If $1 < p_1, \ldots, p_n < \infty$, then there holds that $[\vec{w}]_{A_{\vec{p}}(\mathbb{R}^d)} < \infty$ if and only if

$$\max\big(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{d_1}}[\vec{w}(x_1, \cdot)]_{A_{\vec{p}}(\mathbb{R}^{d_2})}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{d_2}}[\vec{w}(\cdot, x_2)]_{A_{\vec{p}}(\mathbb{R}^{d_1})}\big) < \infty.$$

Moreover, we have that the above maximum satisfies $\max(\cdot, \cdot) \le [\vec{w}]_{A_{\vec{p}}(\mathbb{R}^d)} \lesssim \max(\cdot, \cdot)^\gamma$ where γ is allowed to depend on $\vec{p}$ and d. The first estimate follows from the Lebesgue differentiation theorem. The second estimate can be proved by using Lemma 3.6 and the corresponding linear statement, see (3.1). Using this, (10.3) and Proposition 10.4 gives a bi-parameter version of Proposition 10.4 – the statement is obtained by replacing BMO with bmo, and the quantitative estimate is of the form $[w_{b,z}]_{A_{\vec{p}}} \lesssim [\vec{w}]_{A_{\vec{p}}}^\gamma$. Now, we have everything ready to prove the following commutator estimate.

10.5. Theorem. *Suppose T is an n-linear bi-parameter CZO in $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $1 < p_1, \ldots, p_n \le \infty$ and $1/p = \sum_i 1/p_i > 0$. Suppose also that $b \in \mathrm{bmo}$. Then for all $1 \le k \le n$ we have the commutator estimate*

$$\|[b, T]_k(f_1, \ldots, f_n)w\|_{L^p} \lesssim \|b\|_{\mathrm{bmo}} \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}$$

for all n-linear bi-parameter weights $\vec{w} = (w_1, \ldots, w_n) \in A_{\vec{p}}$. Analogous results hold for iterated commutators.

Proof. We assume $\|b\|_{\mathrm{bmo}} = 1$. It suffices to study $[b, T]_1$, and in fact we shall prove the following principle. Once we have

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

for some $\vec{p}$ in the Banach range, then

$$\|[b, T]_1(f_1, \ldots, f_n)w\|_{L^p} \lesssim \prod_{i=1}^n \|f_i w_i\|_{L^{p_i}}.$$

In this principle the form of the n-linear operator plays no role (T does not need to be a CZO). The iterated cases follow immediately from this principle and the full range then follows from extrapolation.

Define

$$T_z^1(f_1, \ldots, f_n) = e^{zb}T(e^{-zb}f_1, f_2, \ldots, f_n).$$

Then, by the Cauchy integral theorem, we get for nice functions $f_1, \ldots, f_n$, that

$$[b, T]_1(f_1, \ldots, f_n) = \frac{\mathrm{d}}{\mathrm{d}z}T_z^1(f_1, \ldots, f_n)\Big|_{z=0} = \frac{-1}{2\pi i} \int_{|z|=\delta} \frac{T_z^1(f_1, \ldots, f_n)}{z^2}\,\mathrm{d}z, \quad \delta > 0.$$

Since $p \ge 1$, by Minkowski's inequality

$$\|[b, T]_1(f_1, \ldots, f_n)w\|_{L^p} \le \frac{1}{2\pi\delta^2} \int_{|z|=\delta} \|T_z^1(f_1, \ldots, f_n)w\|_{L^p}|\,\mathrm{d}z|.$$

We choose

$$\delta \sim \frac{1}{\max([w^p]_{A_\infty}, \max_i[w_i^{-p_i'}]_{A_\infty})}.$$

This allows to use the bi-parameter version of Proposition 10.4 to have that

$$\|T_z^1(f_1, \ldots, f_n)w\|_{L^p} = \|T(e^{-zb}f_1, f_2, \ldots, f_n)we^{\mathrm{Re}(bz)}\|_{L^p}$$
$$\lesssim \|e^{-zb}f_1w_1e^{\mathrm{Re}(bz)}\|_{L^{p_1}}\prod_{i=2}^n \|f_iw_i\|_{L^{p_i}} = \prod_{i=1}^n \|f_iw_i\|_{L^{p_i}}.$$

The claim follows. $\qquad\square$

REFERENCES

[1] E. Airta, K. Li, H. Martikainen, E. Vuorinen, Some new weighted estimates on product spaces, Indiana Univ. Math. J., to appear, arXiv:1910.12546, 2019. 2

[2] E. Airta, H. Martikainen, E. Vuorinen, Modern singular integral theory with mild kernel regularity, arXiv:2006.05807, 2020. 3, 28, 29, 30, 34

[3] C. Benea, C. Muscalu, Multiple vector-valued, mixed norm estimates for Littlewood-Paley square functions, arXiv:1808.03248v2, 2018. 13

[4] C. Benea, C. Muscalu, Quasi-Banach valued inequalities via the helicoidal method, J. Funct. Anal. 273 (2017) 1295–1353. 37

[5] C. Benea, C. Muscalu, Multiple vector-valued inequalities via the helicoidal method, Anal. PDE 9 (2016) 1931–1988. 37

[6] A. Culiuc, F. Di Plinio, Y. Ou, Domination of multilinear singular integrals by positive sparse forms, J. Lond. Math. Soc. 98 (2) (2018) 369–392. 3

[7] S.-Y. A. Chang, R. Fefferman, A continuous version of duality of H^1 with BMO on the Bidisc, Ann. of Math. 112 (1980) 179–201. 2

[8] S.-Y. A. Chang, R. Fefferman, Some recent developments in Fourier analysis and H^p theory on product domains, Bull. Amer. Math. Soc. 12 (1985) 1–43. 2

[9] R. Coifman, Y. Meyer, Au delà des opérateurs pseudo-différentiels, Astérisque 57 (1978) 1–185. 2

[10] R. Coifman, R. Rochberg, G. Weiss, Factorization theorems for Hardy spaces in several variables, Ann. of Math. (2) 103 (1976) 611–635. 4

[11] D. Cruz-Uribe, J.M. Martell, C. Pérez, Extrapolation from A_∞ weights and applications, J. Funct. Anal. 213 (2004) 412–439. 13

[12] W. Damián, M. Hormozi, K. Li, New bounds for bilinear Calderón-Zygmund operators and applications. Rev. Mat. Iberoam. 34 (2018), no. 3, 1177–1210. 37

[13] F. Di Plinio, K. Li, H. Martikainen, E. Vuorinen, Multilinear operator-valued Calderón-Zygmund theory, J. Funct. Anal. 279 (8) 108666 (2020). 3

[14] F. Di Plinio, Y. Ou, Banach-valued multilinear singular integrals, Indiana Univ. Math. J. 67 (2018) 1711–1763. 3, 37

[15] J. Duoandikoetxea, Extrapolation of weights revisited: New proofs and sharp bounds, J. Funct. Anal. 260 (2011) 1886–1901. 3, 10

[16] L. Grafakos, L. Liu, R. Torres, C. Pérez, The multilinear strong maximal function, J. Geom. Anal. 21 (2011) 118–149. 2, 3, 12

[17] L. Grafakos, J.M. Martell, Extrapolation of weighted norm inequalities for multivariable operators and applications, J. Geom. Anal. 14 (2004) 19–46. 3, 10

[18] L. Grafakos, S. Oh, The Kato-Ponce inequality, Comm. Partial Differential Equations 39 (2014) 1128–1157. 1

[19] L. Grafakos, R. Torres, Multilinear Calderón–Zygmund theory, Adv. Math. 165 (2002) 124–164. 1

[20] A. Grau de la Herrán, Comparison of $T1$ conditions for multi-parameter operators, Proc. Amer. Math. Soc. 144 (2016) 2437–2443. 34

[21] A. Grau de la Herrán, T. Hytönen, Dyadic representation and boundedness of non-homogeneous Calderón–Zygmund operators with mild kernel regularity, Michigan Math. J. 67 (2018) 757–786. 30

[22] R. Fefferman, Harmonic analysis on product spaces, Ann. of Math. 126 (1987) 109–130. 2

[23] R. Fefferman, A^p weights and singular integrals, Amer. J. Math. 110 (1988) 975–987. 2

[24] R. Fefferman, Strong differentiation with respect to measures. Amer. J. Math. 103 (1981) 33–40. 2, 13

[25] R. Fefferman, E. Stein, Singular integrals on product spaces, Adv. Math. 45 (1982) 117–143. 2

[26] S. H. Ferguson, M. T. Lacey, A characterization of product BMO by commutators, Acta Math. 189 (2002) 143–160. 2, 4

[27] I. Holmes, M. Lacey, B. Wick, Commutators in the two-weight setting, Math. Ann. 367 (2017) 51–80. 4

[28] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators. Anal. PDE 11 (2018) 1693–1740. 2, 4, 7, 37

[29] T. Hytönen, The L^p-to-L^q boundedness of commutators with applications to the Jacobian operator, preprint, arXiv:1804.11167, 2018. 4

[30] T. Hytönen, The two-weight inequality for the Hilbert transform with general measures, Proc. Lond. Math. Soc. (3) 117 (2018) 483–526. 30

[31] T. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, Ann. of Math. 175 (2012) 1473–1506. 3, 28

[32] J.-L. Journé, Calderón-Zygmund operators on product spaces, Rev. Mat. Iberoam. 1 (1985) 55–91. 2, 34

[33] T. Kato, G. Ponce, Commutator estimates and the Euler and Navier-Stokes equations, Comm. Pure Appl. Math. 41 (1988) 891–907. 1

[34] A. Lerner, An elementary approach to several results on the Hardy–Littlewood maximal operator, Proc. Amer. Math. Soc. 136 (2008) 2829–2833. 32

[35] A. Lerner, S. Ombrosi, C. Pérez, R.H. Torres, R. Trujillo-González, New maximal functions and multiple weights for the multilinear Calderón–Zygmund theory, Adv. Math. 220 (2009) 1222–1264. 1, 2, 7, 8

[36] A. Lerner, S. Ombrosi, I. Rivera-Ríos, On pointwise and weighted estimates for commutators of Calderón-Zygmund operators, Adv. Math. 319 (2017) 153–181. 4

[37] K. Li, J. M. Martell, S. Ombrosi, Extrapolation for multilinear Muckenhoupt classes and applications, Adv. Math. 373 107286 (2020). 3, 10, 33

[38] K. Li, J.M. Martell, H. Martikainen, S. Ombrosi, E. Vuorinen, End-point estimates, extrapolation for multilinear Muckenhoupt classes, and applications, Trans. Amer. Math. Soc. 374 (2021) 97–135. 3, 7, 10, 33, 37

[39] K. Li, H. Martikainen, E. Vuorinen, Bloom type inequality for bi-parameter singular integrals: efficient proof and iterated commutators, Int. Math. Res. Not. IMRN 2021 8153–8187. 13

[40] K. Li, H. Martikainen, E. Vuorinen, Bilinear Calderón-Zygmund theory on product spaces, J. Math. Pures Appl. 138 (2020) 356–412. 2, 3, 4, 30, 34, 35, 37

[41] K. Li, H. Martikainen, Y. Ou, E. Vuorinen, Bilinear representation theorem, Trans. Amer. Math. Soc. 371 (6) (2019) 4193–4214. 3

[42] K. Li, K. Moen, W. Sun, The Sharp Weighted Bound for Multilinear Maximal Functions and Calderón–Zygmund Operators, J. Fourier Anal. Appl. 20 (2014) 751–765. 9

[43] H. Martikainen, Representation of bi-parameter singular integrals by dyadic operators, Adv. Math. 229 (2012) 1734–1761. 3

[44] H. Martikainen, T. Orponen, Some obstacles in characterising the boundedness of bi-parameter singular integrals, Math. Z. 282 (2016) 535–545. 19, 25

[45] B. Nieraeth, Quantitative estimates and extrapolation for multilinear weight classes, Math. Ann. 375 (2019) 453–507. 3, 10

[46] S. Oh, X. Wu, On L^1 endpoint Kato-Ponce inequality, Math. Res. Lett. 27 (2020) 1129–1164. 37

[47] Y. Ou, Multi-parameter singular integral operators and representation theorem, Rev. Mat. Iberoam. 33 (2017) 325–350. 3, 30, 34

[48] M. Wilson, Weighted Littlewood-Paley theory and exponential-square integrability, Lecture Notes in Mathematics, 1924, Springer, Berlin, 2008. 13, 14

(K.L.) CENTER FOR APPLIED MATHEMATICS, TIANJIN UNIVERSITY, WEIJIN ROAD 92, 300072 TIANJIN, CHINA

Email address: kli@tju.edu.cn

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

Email address: henri.martikainen@helsinki.fi

Email address: emil.vuorinen@helsinki.fi