

OFF-DIAGONAL ESTIMATES FOR BI-COMMUTATORS

EMIL AIRTA, TUOMAS HYTÖNEN, KANGWEI LI, HENRI MARTIKAINEN, AND TUOMAS OIKARI

ABSTRACT. We study the bi-commutators $[T_1, [b, T_2]]$ of pointwise multiplication and Calderón–Zygmund operators, and characterize their $L^{p_1} L^{p_2} \rightarrow L^{q_1} L^{q_2}$ boundedness for several off-diagonal regimes of the mixed-norm integrability exponents $(p_1, p_2) \neq (q_1, q_2)$. The strategy is based on a bi-parameter version of the recent approximate weak factorization method.

1. INTRODUCTION

In this paper we characterise many new estimates for **bi**-commutators. The classical commutators have the form $[b, T]: f \mapsto bTf - T(bf)$, where T is a singular integral operator

$$(1.1) \quad Tf(x) = \int_{\mathbb{R}^d} K(x, y) f(y) \, dy.$$

The expression (1.1) represents a broad class of linear transformations of functions arising across analysis. Distinguished special cases are the Hilbert transform H in dimension $d = 1$, which has the kernel $K(x, y) = \frac{1}{x-y}$, and the Riesz transforms R_j in dimensions $d \geq 2$, which have the kernel $K_j(x, y) = \frac{x_j - y_j}{|x-y|^{d+1}}$, $j = 1, \dots, d$.

The Hilbert transform lies in the scope of complex analysis, and the Hilbert commutator $[b, H]$ is connected to Hankel operators. A classical theorem of Nehari [42] characterises the boundedness of the Hilbert commutators $[b, H]$ via this link. On the other hand, the result of Coifman–Rochberg–Weiss [7] brings the commutators to the heart of harmonic analysis by showing that

$$(1.2) \quad \|b\|_{\text{BMO}} \lesssim \|[b, T]\|_{L^p(\mathbb{R}^d) \rightarrow L^p(\mathbb{R}^d)} \lesssim \|b\|_{\text{BMO}}, \quad p \in (1, \infty),$$

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

$$\|b\|_{\text{BMO}} := \sup_I \int_I |b - \langle b \rangle_I|,$$

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

The off-diagonal situation $[b, T]: L^p \rightarrow L^q$, $p \neq q$, is also completely understood. In the case $1 < p < q < \infty$, a two-sided estimate like (1.2) with BMO replaced by the

2010 *Mathematics Subject Classification.* 42B20.

Key words and phrases. Calderón–Zygmund operators, singular integrals, multi-parameter analysis, commutators.

homogeneous Hölder space $\dot{C}^{0,\alpha}$,

$$\|b\|_{\dot{C}^{0,\alpha}} := \sup_{x \neq y} \frac{|b(x) - b(y)|}{|x - y|^\alpha}, \quad \alpha := d \left(\frac{1}{p} - \frac{1}{q} \right),$$

was obtained by Janson [30]. The remaining range with $1 < q < p < \infty$ was characterised only very recently in [24]:

$$\|[b, T]\|_{L^p \rightarrow L^q} \sim \|b\|_{\dot{L}^r} := \inf_{c \in \mathbb{C}} \|b - c\|_{L^r}, \quad \frac{1}{r} := \frac{1}{q} - \frac{1}{p}.$$

As the commutator annihilates constants, all of the spaces above also have the feature that they do not see constants. This philosophy is more complicated in the bi-parameter setting where the bi-commutator annihilates all functions that depend only on one of the variables $(x_1, x_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

The motivation for commutator estimates stems from their many applications and connections to modern harmonic analysis. For example, the $L^p \rightarrow L^p$ characterization yields factorizations for Hardy spaces [7] and implies various div-curl lemmas relevant for compensated compactness [8]. In [24], the off-diagonal $L^p \rightarrow L^q$ boundedness for $q < p$ is connected to a conjecture of Iwaniec [29] about the prescribed Jacobian problem.

In this paper, we address the question of the off-diagonal $L^p(\mathbb{R}^d) \rightarrow L^q(\mathbb{R}^d)$ boundedness of the bi-commutators $[T_1, [b, T_2]]$, where each T_i is a Calderón–Zygmund operator on \mathbb{R}^{d_i} and \mathbb{R}^d is viewed as the bi-parameter product space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Due to the product space nature of the problem, it is natural to introduce an additional level of generality by allowing different integrability exponents in the x_1 and x_2 coordinates, thereby leading to the question of $L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}$ boundedness. In contrast to the three qualitatively different regimes $p < q$, $p = q$ and $p > q$ for the commutator, there will now be nine different situations depending on the relative size of both p_1, q_1 on the one hand, and p_2, q_2 on the other hand.

Our question is the natural bi-parameter analogue of [24]. In fact, [24] already suggests a naive conjecture, where in each regime of the exponents, we should have the corresponding natural vector-valued space. For example, in the case

$$p_1 < q_1 \text{ and } p_2 = q_2$$

we could expect the space $\dot{C}_{x_1}^{0,\beta_1}(\text{BMO}_{x_2})$ and in the case

$$p_1 = q_1 \text{ and } p_2 < q_2$$

we could expect the space $\text{BMO}_{x_1}(\dot{C}_{x_2}^{0,\beta_2})$, where $\beta_i := d_i \left(\frac{1}{p_i} - \frac{1}{q_i} \right)$. Somewhat strikingly, we show that the naive conjecture is not always the right one. In the second case, the space $\dot{C}_{x_2}^{0,\beta_2}(\text{BMO}_{x_1})$, where the order of the spaces $\dot{C}_{x_2}^{0,\beta_2}$ and BMO_{x_1} has been switched to an unexpected order, provides the correct sufficient and necessary condition.

The following Theorem 1.3 is our main result. It verifies the naive conjecture in some cases and proves an unexpected necessary and sufficient condition in some other cases. Some characterisations are left for future work as they do not appear to be amenable to our current methods. The stated theorem is a simplified and shortened version of the obtained new estimates – for example, the symmetry assumption is used here purely for convenience in order not to have to include partial adjoints $[T_1^*, [b, T_2]]$ in the estimates. In the three lower-right cases of the table of Theorem 1.3, where only an upper bound is stated, we also obtain certain related lower bounds, but these do not admit a simple

formulation in terms of classical function spaces, and are therefore omitted in this Introduction.

1.3. Theorem. *Let T_1 and T_2 be two symmetrically non-degenerate CZOs on \mathbb{R}^{d_1} and \mathbb{R}^{d_2} , respectively, $b \in L^2_{\text{loc}}(\mathbb{R}^{d_1+d_2})$ and $p_1, p_2, q_1, q_2 \in (1, \infty)$. Let*

$$\beta_i := d_i \left(\frac{1}{p_i} - \frac{1}{q_i} \right), \quad \text{if } p_i < q_i; \quad \frac{1}{r_i} := \frac{1}{q_i} - \frac{1}{p_i}, \quad \text{if } p_i > q_i.$$

Then $\|[T_1, [b, T_2]]\|_{L^{p_1}_{x_1} L^{p_2}_{x_2} \rightarrow L^{q_1}_{x_1} L^{q_2}_{x_2}}$ has upper and lower bounds according to the following table:

	$p_1 < q_1$	$p_1 = q_1$	$p_1 > q_1$
$p_2 < q_2$	$\sim \ b\ _{\dot{C}_{x_1}^{0, \beta_1}(\dot{C}_{x_2}^{0, \beta_2})}$	$\sim \ b\ _{\dot{C}_{x_2}^{0, \beta_2}(\text{BMO}_{x_1})}$	$\lesssim \ b\ _{\dot{L}_{x_1}^{r_1}(\dot{C}_{x_2}^{0, \beta_2})}$ $\gtrsim \ b\ _{\dot{C}_{x_2}^{0, \beta_2}(\dot{L}_{x_1}^{r_1})}$
$p_2 = q_2$	$\sim \ b\ _{\dot{C}_{x_1}^{0, \beta_1}(\text{BMO}_{x_2})}$	$\lesssim \ b\ _{\text{BMO}_{\text{prod}}}$ $\gtrsim \ b\ _{\text{BMO}_{\text{rect}}}$	$\lesssim \ b\ _{\dot{L}_{x_1}^{r_1}(\text{BMO}_{x_2})}$
$p_2 > q_2$	$\sim \ b\ _{\dot{C}_{x_1}^{0, \beta_1}(\dot{L}_{x_2}^{r_2})}$	$\lesssim \ b\ _{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})}$	$\lesssim \ b\ _{\dot{L}_{x_1}^{r_1}(\dot{L}_{x_2}^{r_2})}$

As one can see from the table, all of the four cases, where we obtain a complete characterization, are about off-diagonal mapping properties in the sense that $(p_1, p_2) \neq (q_1, q_2)$. The diagonal case, involving the product BMO space of Chang and Fefferman [5, 6], has been studied by Ferguson–Sadosky [14] and Ferguson–Lacey [15], with further extensions appearing in [32, 33, 44]. Very recently, the celebrated characterization in [15] has been found to contain a gap, see [21], which remains open at the time of writing, and reappears in several subsequent works on the topic. The original counterexample to a key step in [15] is presented in [21], and we provide another counterexample in Appendix A. It should be observed that these are only counterexamples to the existing proofs, not to the Ferguson–Lacey characterization itself, so that the truth of the actual result remains open. The ongoing quest to repair the Ferguson–Lacey characterization significantly adds to the interest in finding alternative approaches to the difficult lower bounds for bi-commutators, and it might be worthwhile investigating whether our new bi-parameter weak factorization methods could also shed some light on the diagonal case and the product BMO space.

While it is clear that the rectangular oscillations appearing in this paper cannot, as such, directly characterize the product BMO norm involving general open sets, even the product BMO norm can be obtained by suitably dualizing against product Hardy space atoms using the product space theory of Chang and Fefferman. It is not unthinkable that one could directly construct suitable weak factorizations even for these complicated atoms; presumably, this should involve considerations in the spirit of Journé’s covering lemma for product spaces [31] or its variants [4, 46]. Even if the gap in [15] is eventually fixed by modifying the original proof, a shift of paradigm from the Fourier-analytic methods of [15, 32, 33, 44] to real-variable methods of our approach would have other positive effects – for example, it could aid the development of related commutator bounds in other settings like those of [11, 12, 13]. For the time being, our result concerning the rectangular BMO lower bound is in fact the best that is available in the diagonal case.

It is important to realize that commutator upper bounds are expected for all bounded singular integrals, while the lower bounds obviously require some non-degeneracy. All our lower bounds hold under weak non-degeneracy assumptions on the operators, and neither the homogeneity nor the translation-invariance of the operators, crucially used in [15, 32], is needed here.

There has been a vast amount of other recent activity regarding both one-parameter and multi-parameter commutators. For example, the two-weight commutator estimates, which include [18, 19, 20, 24, 34, 35, 36, 37], have been one of the main lines of development. Commutators are also actively studied in other settings: see for example [11] for the flag setting, [12] for the Zygmund dilation setting and [13] for the Bessel setting.

The motivation for this paper is not just the results in Theorem 1.3 but also the associated methodology. We develop some interesting bi-parameter versions of the flexible one-parameter commutator methods [24]. In more detail, our strategy requires us to develop new bi-parameter methods that exploit the following known interplay between commutator bounds and weak factorizations. The logic of this connection is that while Nehari's theorem [42] may be seen as a corollary of the classical factorization of $H^1(\mathbb{D})$, Coifman, Rochberg and Weiss [7] reversed the reasoning and deduced from the commutator estimate (1.2), using duality, a *weak* factorization (involving sums of products, rather than just products) of the real-variable Hardy space $H^1(\mathbb{R}^d)$. This connection between commutators estimates and factorizations of functions stems from identities like

$$(1.4) \quad \langle [b, T]f, g \rangle = \langle b, Tf \cdot g - f \cdot T^*g \rangle.$$

Recently, in [24] the idea was to reverse the reasoning again, and to first directly prove a suitable factorization, and then use it to prove the desired commutator estimate. This leads to the approximate weak factorization method, where a function h is expanded in the form

$$(1.5) \quad h = \frac{h}{T^*g} T^*g =: -f \cdot T^*g = -f \cdot T^*g + Tf \cdot g + h', \quad h' := -Tf \cdot g,$$

and a suitable g above allows to absorb the error $\|h'\| \ll \|h\|$.

Instead of (1.4), in the bi-parameter setting we have

$$\langle [T_1, [b, T_2]]f, g \rangle = \langle b, T_2f \cdot T_1^*g - f \cdot T_1^*T_2^*g - T_1T_2f \cdot g + T_1f \cdot T_2^*g \rangle.$$

We develop our new weak factorizations in Section 4. In Lemma 4.4 we expand a function f supported on a rectangle $R = I \times J$ and satisfying $\int_I f = \int_J f = 0$ using the approximate weak factorization logic. This produces more error terms, with more complicated supports, when compared to the one-parameter analogue (1.5), and makes the repeated use of the factorization more tricky. Nevertheless, we manage to use our bi-parameter approximate weak factorization to e.g. prove Theorem 4.18, which is the key to many commutator lower bounds.

Here is an outline of the paper. After the short preliminaries of Section 2 we move on to the oscillatory characterizations of various function spaces in Section 3. These function space characterizations will be combined with the weak factorizations of Section 4 to prove our necessary conditions for commutator boundedness in Section 5. Section 6 collects all the sufficient conditions. Sections 5 and 6 combined give Theorem 1.3. Section 7 records some additional vector-valued estimates for commutators that are of independent interest – see e.g. Theorem 7.12.

Acknowledgements. E. Airta, T. Hytönen, H. Martikainen and T. Oikari were supported by the Academy of Finland through project Nos. 294840 (Martikainen), 327271 (Airta, Martikainen), 306901 (Airta, Martikainen, Oikari), and 314829 (Hytönen), and through the Finnish Centre of Excellence in Analysis and Dynamics Research (project No. 307333, all four). Martikainen and Oikari were also supported by the three-year research grant 75160010 of the University of Helsinki.

2. PRELIMINARIES

2.A. Basic notation. We denote $A \lesssim B$ if $A \leq CB$ for some absolute constant C . The constant C can at least depend on the dimensions of the appearing Euclidean spaces, on integration exponents, and on various Banach space constants. It can also depend on various other fixed constants, like those related to singular integrals, and so on. We denote $A \sim B$ if $B \lesssim A \lesssim B$.

When we consider \mathbb{R}^d as a bi-parameter product space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ we often denote the mixed-norm space $L^{p_1}(\mathbb{R}^{d_1}; L^{p_2}(\mathbb{R}^{d_2}))$ by $L_{x_1}^{p_1} L_{x_2}^{p_2}$ – this is suggested by the notation $x = (x_1, x_2) \in \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We also always identify $f: \mathbb{R}^d \rightarrow \mathbb{C}$ satisfying

$$\left(\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} |f(x_1, x_2)|^{p_2} dx_2 \right)^{p_1/p_2} dx_1 \right)^{1/p_1} < \infty$$

with the function $\phi_f \in L^{p_1}(\mathbb{R}^{d_1}; L^{p_2}(\mathbb{R}^{d_2}))$, $\phi_f(x_1) = f(x_1, \cdot)$.

Often integral pairings – denoted with the bracked notation $\langle f, g \rangle = \int fg$ – need to be taken with respect to one of the variables only. For example, if $f: \mathbb{R}^d \rightarrow \mathbb{C}$ and $h: \mathbb{R}^{d_1} \rightarrow \mathbb{C}$, then $\langle f, h \rangle_1: \mathbb{R}^{d_2} \rightarrow \mathbb{C}$ is defined by

$$\langle f, h \rangle_1(x_2) = \int_{\mathbb{R}^{d_1}} f(y_1, x_2) h(y_1) dy_1.$$

We denote averages by

$$\langle f \rangle_A = \int_A f := \frac{1}{|A|} \int_A f,$$

where $|A|$ is the Lebesgue measure of the set A . The indicator function of a set A is denoted by 1_A . We try to denote cubes in \mathbb{R}^{d_i} by I_i, J_i, L_i and so on – that is, the dimension of the cube can be read from the subscript. Various rectangles then take the form $I_1 \times I_2, J_1 \times J_2$, etc. The side length of a cube I_i is denoted by $\ell(I_i)$.

2.B. Singular integrals and commutators. We call

$$K_i: \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \setminus \{(x_i, y_i) \in \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} : x_i = y_i\} \rightarrow \mathbb{C}$$

a standard Calderón-Zygmund kernel on \mathbb{R}^{d_i} if we have

$$|K(x_i, y_i)| \leq \frac{C}{|x_i - y_i|^{d_i}}$$

and, for some $\alpha_i \in (0, 1]$, we have

$$(2.1) \quad |K(x_i, y_i) - K(x'_i, y_i)| + |K(y_i, x_i) - K(y_i, x'_i)| \leq C \frac{|x_i - x'_i|^{\alpha_i}}{|x_i - y_i|^{d_i + \alpha_i}}$$

whenever $|x_i - x'_i| \leq |x_i - y_i|/2$.

Many of our results hold with (2.1) replaced by a significantly weaker assumption – see Remark 4.6. However, we do not emphasise this too much as it is not a novelty of this paper (see e.g. [24]).

A singular integral operator (SIO) is a linear operator T_i on \mathbb{R}^{d_i} (initially defined, for example, on bounded and compactly supported functions) so that there is a standard kernel K_i for which

$$\langle T_i f, g \rangle = \iint_{\mathbb{R}^{d_i} \times \mathbb{R}^{d_i}} K_i(x_i, y_i) f(y_i) g(x_i) dy_i dx_i$$

whenever the functions f and g are nice and have disjoint supports. A Calderón–Zygmund operator (CZO) is an SIO T_i , which is bounded from $L^p(\mathbb{R}^{d_i}) \rightarrow L^p(\mathbb{R}^{d_i})$ for all (equivalently for some) $p \in (1, \infty)$. The $T1$ theorem [10] says that an SIO is a CZO if and only if

$$\int_{I_i} |T_i 1_{I_i}| + \int_{I_i} |T_i^* 1_{I_i}| \lesssim |I_i|$$

for all cubes $I_i \subset \mathbb{R}^{d_i}$. Here T_i^* is the linear adjoint of T_i . We know a lot about the structure of a CZO T_i : we can represent T_i with certain dyadic model operators (DMOs) – see [22, 23]. We will have use for this later.

If $b \in L_{\text{loc}}^s(\mathbb{R}^d)$ for some $s \in (1, \infty)$ and T is a CZO on \mathbb{R}^d , then the pairing

$$\langle [b, T]f, g \rangle = \langle Tf, bg \rangle - \langle T(bf), g \rangle$$

is well-defined for $f, g \in L_c^\infty$. If $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and T_i is a CZO on \mathbb{R}^{d_i} , we can similarly define $\langle [T_1, [b, T_2]]f, g \rangle$. Then we can ask, via duality, if $[T_1, [b, T_2]]$ maps $L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}$. However, when a commutator lower bound is proved, the full norm

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}}$$

is not actually needed. We will define so-called off-support versions of the norm, which can be defined even if we only have $b \in L_{\text{loc}}^1$. In fact, these off-support constants depend only on the kernels K_1 and K_2 and not on the CZOs themselves.

When we are given a CZO T_i we always write K_i for its kernel without explicit mention.

3. FUNCTION SPACES AND OSCILLATORY CHARACTERIZATIONS

In our goal of linking the boundedness of the commutator $[T_1, [b, T_2]]$ with the membership of b in a suitable function space, a useful intermediate notion is provided by various oscillatory characterizations of the different function space norms. In this section we specifically go through all the oscillatory conditions that appear in our commutator lower bounds.

3.A. One-parameter spaces. We begin by recalling the relevant characterizations in the one-parameter situation, as this will motivate the necessarily more complicated expressions in the two parameter case.

The case of BMO is most immediate, as the norm is directly given by the oscillatory quantity

$$\|b\|_{\text{BMO}(\mathbb{R}^d)} := \sup_I \int_I |b - \langle b \rangle_I|.$$

For the homogeneous Hölder norms

$$\|b\|_{\dot{C}^{0,\alpha}(\mathbb{R}^d)} := \sup_{x \neq y} \frac{|b(x) - b(y)|}{|x - y|^\alpha},$$

there is a well-known analogous equivalent norm:

3.1. Proposition. *We have*

$$\|b\|_{\dot{C}^{0,\alpha}(\mathbb{R}^d)} \sim \sup_I \frac{1}{\ell(I)^\alpha} \int_I |b - \langle b \rangle_I|,$$

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

Proof. “ \gtrsim ” is immediate, since $|b(x) - \langle b \rangle_I| \leq \int_I |b(x) - b(y)| dy$, where $|b(x) - b(y)| \lesssim \|b\|_{\dot{C}^{0,\alpha}(\mathbb{R}^d)} \ell(I)^\alpha$ for all $x, y \in I$.

For “ \lesssim ”, denote the right-hand side of the claim by C_0 , and fix $x \neq y \in \mathbb{R}^d$. Define $x_k := (1 - 2^{-k})x + 2^{-k}y$ and $y_k := (1 - 2^{-k})y + 2^{-k}x$, and note that $x_1 = y_1 = \frac{1}{2}(x + y)$. If $Q_k(x) := Q(x_k, 2^{-k}|x - y|)$ is the cube of centre x_k and side-length $2 \cdot 2^{-k}|x - y|$, we easily check that $Q_{k+1}(x) \subset Q_k(x)$ and $Q_k(x) \subset B(x, C2^{-k}|x - y|)$. Thus, we have

$$b(x) = \sum_{k=1}^{\infty} \left(\int_{Q_{k+1}(x)} b - \int_{Q_k(x)} b \right) + \int_{Q(\frac{1}{2}(x+y), \frac{1}{2}|x-y|)} b$$

and hence

$$\begin{aligned} \left| b(x) - \int_{Q(\frac{1}{2}(x+y), \frac{1}{2}|x-y|)} b \right| &\leq \sum_{k=1}^{\infty} \int_{Q_{k+1}(x)} |b - \langle b \rangle_{Q_k(x)}| \lesssim \sum_{k=1}^{\infty} \int_{Q_k(x)} |b - \langle b \rangle_{Q_k(x)}| \\ &\lesssim C_0 \sum_{k=1}^{\infty} \ell(Q_k(x))^\alpha \lesssim C_0 |x - y|^\alpha \sum_{k=1}^{\infty} 2^{-k\alpha} \lesssim C_0 |x - y|^\alpha. \end{aligned}$$

A similar bound for y in place of x and the triangle inequality show that $|b(x) - b(y)| \lesssim C_0 |x - y|^\alpha$. \square

The final space of interest in the one-parameter case is

$$\|b\|_{\dot{L}^r(\mathbb{R}^d)} := \inf_c \|b - c\|_{L^r},$$

where the infimum is taken over all constants. For this space we have the following characterisation – similar estimates already appeared in [24], but we single them out here as a separate proposition. A collection of cubes \mathcal{S} is called γ -sparse if there are pairwise disjoint subsets $E(S) \subset S$, $S \in \mathcal{S}$, with $|E(S)| \geq \gamma|S|$. We can often work with $\gamma = 1/2$ and simply talk about sparseness.

3.2. Proposition. *For $r \in (1, \infty)$ we have*

$$\begin{aligned} \|b\|_{\dot{L}^r(\mathbb{R}^d)} &\sim \sup_Q \|b - \langle b \rangle_Q\|_{L^r(Q)} \sim \|M^\# b\|_{L^r(\mathbb{R}^d)} \\ &\sim \sup \left\{ \sum_{S \in \mathcal{S}} \lambda_S \int_S |b - \langle b \rangle_S| : \mathcal{S} \text{ is sparse, } \sum_{S \in \mathcal{S}} |S| \lambda_S^{r'} \leq 1 \right\}, \end{aligned}$$

where $M^\# b$ is the sharp maximal function $M^\# b(x) = \sup_{Q \ni x} \int_Q |b - \langle b \rangle_Q|$ and the supremum is taken over all cubes $Q \subset \mathbb{R}^d$.

Proof. We prove a chain of upper bounds both starting and finishing with $\|b\|_{\dot{L}^r(\mathbb{R}^d)}$, and covering all other expressions as intermediate steps. To begin with, we have

$$\|b\|_{\dot{L}^r(\mathbb{R}^d)} \lesssim \sup_Q \|b - \langle b \rangle_Q\|_{L^r(Q)} =: C_0.$$

To see this, fix an increasing sequence of cubes $(Q_i)_i$ that exhaust \mathbb{R}^d . Let $i \leq j$ and write

$$\begin{aligned} |\langle b \rangle_{Q_j} - \langle b \rangle_{Q_i}| &= \left(\int_{Q_i} |[\langle b \rangle_{Q_j} - b(x)] + [b(x) - \langle b \rangle_{Q_i}]|^r dx \right)^{1/r} \\ &\leq |Q_i|^{-1/r} (\|b - \langle b \rangle_{Q_j}\|_{L^r(Q_i)} + \|b - \langle b \rangle_{Q_i}\|_{L^r(Q_i)}) \\ &\leq |Q_i|^{-1/r} (\|b - \langle b \rangle_{Q_j}\|_{L^r(Q_j)} + \|b - \langle b \rangle_{Q_i}\|_{L^r(Q_i)}) \leq 2|Q_i|^{-1/r} C_0. \end{aligned}$$

Thus, $(\langle b \rangle_{Q_j})_j$ is a Cauchy sequence and the limit $c := \lim_{j \rightarrow \infty} \langle b \rangle_{Q_j}$ exists. Fatou's lemma yields the desired estimate:

$$\|b\|_{\dot{L}^r(\mathbb{R}^d)}^r \leq \int_{\mathbb{R}^d} |b - c|^r = \int \lim_{j \rightarrow \infty} 1_{Q_j} |b - \langle b \rangle_{Q_j}|^r \leq \liminf_{j \rightarrow \infty} \int_{Q_j} |b - \langle b \rangle_{Q_j}|^r \leq C_0^r.$$

For a fixed cube $Q \subset \mathbb{R}^d$, we have

$$1_Q |b - \langle b \rangle_Q| \lesssim \sum_{S \in \mathcal{S}} 1_S \int_S |b - \langle b \rangle_S|$$

for a suitable sparse subcollection $\mathcal{S} \subset \mathcal{D}(Q)$. Here $\mathcal{D}(Q)$ consists of the dyadic subcubes of Q obtained by dividing Q in the natural way. For this elementary variant of Lerner's oscillation formula see e.g. [24, Lemma 3.4]. Using this we now have

$$\|b - \langle b \rangle_Q\|_{L^r(Q)} \lesssim \left\| \sum_{S \in \mathcal{S}} 1_S \int_S |b - \langle b \rangle_S| \right\|_{L^r(\mathbb{R}^d)} \lesssim \left(\sum_{S \in \mathcal{S}} |S| \left[\int_S |b - \langle b \rangle_S| \right]^r \right)^{1/r},$$

where the last step is easily verified by dualising with $\phi \in L^{r'}$ and using the definition of sparseness:

$$\begin{aligned} (3.3) \quad \int \left(\sum_{S \in \mathcal{S}} 1_S c_S \right) \phi &\lesssim \sum_{S \in \mathcal{S}} c_S \int_{E_S} \langle \phi \rangle_S \leq \int \left(\sum_{S \in \mathcal{S}} 1_{E(S)} c_S \right) M\phi \\ &\leq \left\| \sum_{S \in \mathcal{S}} 1_{E(S)} c_S \right\|_{L^r} \|M\phi\|_{L^{r'}} \lesssim \left(\sum_{S \in \mathcal{S}} |E(S)| c_S^r \right)^{1/r} \|\phi\|_{L^{r'}}. \end{aligned}$$

Dualizing the ℓ^r norm with $\ell^{r'}$, we find that

$$\left(\sum_{S \in \mathcal{S}} \left[|S|^{1/r} \int_S |b - \langle b \rangle_S| \right]^r \right)^{1/r} = \sum_{S \in \mathcal{S}} \tilde{\lambda}_S |S|^{1/r} \int_S |b - \langle b \rangle_S| = \sum_{S \in \mathcal{S}} \lambda_S \int_S |b - \langle b \rangle_S|$$

for suitable coefficients λ_S with $\sum_{S \in \mathcal{S}} |S| \lambda_S^{r'} = \sum_{S \in \mathcal{S}} \tilde{\lambda}_S^{r'} \leq 1$.

Next, using sparseness and the definition of the sharp maximal operator, we observe that given an arbitrary sparse collection \mathcal{S} and coefficients λ_S with $\sum_{S \in \mathcal{S}} |S| \lambda_S^{r'} \leq 1$, we

have

$$\begin{aligned} \sum_{S \in \mathcal{S}} |S| \lambda_S \int_S |b - \langle b \rangle_S| &\lesssim \sum_{S \in \mathcal{S}} |E(S)| \lambda_S \inf_{z \in S} M^\# b(z) \leq \int \left(\sum_{S \in \mathcal{S}} 1_{E(S)} \lambda_S \right) M^\# b \\ &\leq \left\| \sum_{S \in \mathcal{S}} 1_{E(S)} \lambda_S \right\|_{L^{r'}} \|M^\# b\|_{L^r} = \left(\sum_{S \in \mathcal{S}} |E(S)| \lambda_S^{r'} \right)^{1/r'} \|M^\# b\|_{L^r}, \end{aligned}$$

where the first factor is bounded by 1. Lastly, we have

$$\|M^\# b\|_{L^r} = \|M^\#(b - c)\|_{L^r} \lesssim \|b - c\|_{L^r}$$

for every constant c , and hence $\|M^\# b\|_{L^r} \lesssim \|b\|_{L^r}$. \square

3.B. Bi-parameter spaces. For $b \in L^1_{\text{loc}}$ and a rectangle $R = I_1 \times I_2$ we denote

$$(3.4) \quad \text{osc}^{v_1, v_2}(b, R) := \|b - \langle b \rangle_{I_1, 1} - \langle b \rangle_{I_2, 2} + \langle b \rangle_R\|_{L^{v_1}_{x_1} L^{v_2}_{x_2}(R)}, \quad 1 \leq v_i \leq \infty.$$

Homogeneous Hölder spaces. If X is a Banach space with norm $|\cdot|_X$ and $b: \mathbb{R}^d \rightarrow X$, we define

$$\|b\|_{\dot{C}^{0, \alpha}(X)} = \sup_{x \neq y} \frac{|b(x) - b(y)|_X}{|x - y|^\alpha}.$$

In $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ the natural bi-parameter homogeneous Hölder norm is

$$\begin{aligned} \|b\|_{\dot{C}^{0, \alpha, \beta}(\mathbb{R}^d)} &:= \|b\|_{\dot{C}^{0, \alpha}_{x_1}(\dot{C}^{0, \beta}_{x_2})} = \|b\|_{\dot{C}^{0, \beta}_{x_2}(\dot{C}^{0, \alpha}_{x_1})} \\ &= \sup_{x_1 \neq y_1} \left\| \frac{b(x_1, \cdot) - b(y_1, \cdot)}{|x_1 - y_1|^\alpha} \right\|_{\dot{C}^{0, \beta}} = \sup_{x_2 \neq y_2} \left\| \frac{b(\cdot, x_2) - b(\cdot, y_2)}{|x_2 - y_2|^\beta} \right\|_{\dot{C}^{0, \alpha}} \\ &= \sup_{\substack{x_1 \neq y_1 \\ x_2 \neq y_2}} \frac{|b(x_1, x_2) - b(x_1, y_2) - b(y_1, x_2) + b(y_1, y_2)|}{|x_1 - y_1|^\alpha |x_2 - y_2|^\beta}. \end{aligned}$$

The following oscillatory characterization holds for this norm.

3.5. Proposition. *We have*

$$\|b\|_{\dot{C}^{0, \alpha, \beta}(\mathbb{R}^d)} \sim \sup \left\{ \frac{1}{\ell(I_1)^\alpha \ell(I_2)^\beta} \frac{\text{osc}^{1, 1}(b, R)}{|R|} : R = I_1 \times I_2 \text{ rectangle} \right\}.$$

Proof. We denote the supremum on the right hand side by C_0 .

" \gtrsim " is immediate, since for all $x_i \in I_i$ we have

$$\begin{aligned} &|b(x_1, x_2) - \langle b \rangle_{I_1, 1}(x_2) - \langle b \rangle_{I_2, 2}(x_1) + \langle b \rangle_{I_1 \times I_2}| \\ &\leq \int_{I_1 \times I_2} |b(x_1, x_2) - b(x_1, y_2) - b(y_1, x_2) + b(y_1, y_2)| dy_1 dy_2 \\ &\lesssim \|b\|_{\dot{C}^{0, \alpha, \beta}(\mathbb{R}^d)} \ell(I_1)^\alpha \ell(I_2)^\beta. \end{aligned}$$

For " \lesssim ", notice that by Proposition 3.1 we have

$$\begin{aligned} \|b\|_{\dot{C}^{0, \alpha, \beta}(\mathbb{R}^d)} &= \sup_{x_2 \neq y_2} \left\| \frac{b(\cdot, x_2) - b(\cdot, y_2)}{|x_2 - y_2|^\beta} \right\|_{\dot{C}^{0, \alpha}} \\ &\sim \sup_{x_2 \neq y_2} \frac{1}{|x_2 - y_2|^\beta} \sup_{I_1} \frac{1}{\ell(I_1)^\alpha} \int_{I_1} |b_{x_2, y_2} - \langle b_{x_2, y_2} \rangle_{I_1}|, \end{aligned}$$

where

$$b_{x_2, y_2} = b(\cdot, x_2) - b(\cdot, y_2).$$

So we fix $x_2 \neq y_2$ and a cube $I_1 \subset \mathbb{R}^{d_1}$. By expanding $b(\cdot, x_2)$ as in the proof of Proposition 3.1 (using the same notation as there) we get

$$b(\cdot, x_2) = \sum_{k=1}^{\infty} \left(\int_{Q_{k+1}(x_2)} b(\cdot, z_2) \, dz_2 - \int_{Q_k(x_2)} b(\cdot, z_2) \, dz_2 \right) + \int_{Q(\frac{1}{2}(x_2+y_2), \frac{1}{2}|x_2-y_2|)} b(\cdot, z_2) \, dz_2.$$

Hence

$$b_{x_2, y_2} = \sum_{k=1}^{\infty} [\langle b \rangle_{Q_{k+1}(x_2), 2} - \langle b \rangle_{Q_k(x_2), 2}] - \sum_{k=1}^{\infty} [\langle b \rangle_{Q_{k+1}(y_2), 2} - \langle b \rangle_{Q_k(y_2), 2}].$$

We first consider the contribution of the first summand. We have

$$\begin{aligned} & \int_{I_1} \left| [\langle b \rangle_{Q_{k+1}(x_2), 2} - \langle b \rangle_{Q_k(x_2), 2}] - \langle [\langle b \rangle_{Q_{k+1}(x_2), 2} - \langle b \rangle_{Q_k(x_2), 2}] \rangle_{I_1} \right| \\ &= \int_{I_1} |\langle b - \langle b \rangle_{I_1, 1} \rangle_{Q_{k+1}(x_2), 2} - \langle b - \langle b \rangle_{I_1, 1} \rangle_{Q_k(x_2), 2}| \\ &\lesssim \int_{I_1 \times Q_k(x_2)} |b - \langle b \rangle_{I_1, 1} - \langle b - \langle b \rangle_{I_1, 1} \rangle_{Q_k(x_2), 2}| \\ &\lesssim C_0 \ell(I_1)^\alpha (2^{-k}|x_2 - y_2|)^\beta. \end{aligned}$$

Thus, the contribution of the first summand to the average $\int_{I_1} |b_{x_2, y_2} - \langle b_{x_2, y_2} \rangle_{I_1}|$, can be dominated by $C_0 \ell(I_1)^\alpha |x_2 - y_2|^\beta$. Likewise, the second summand contributes the same. Thus, we have established

$$\int_{I_1} |b_{x_2, y_2} - \langle b_{x_2, y_2} \rangle_{I_1}| \lesssim C_0 \ell(I_1)^\alpha |x_2 - y_2|^\beta$$

proving that $\|b\|_{\dot{C}^{0, \alpha, \beta}(\mathbb{R}^d)} \lesssim C_0$ as desired. \square

For the norm

$$\|b\|_{\dot{C}_{x_1}^{0, \alpha}(\text{BMO}_{x_2})} = \sup_{x_1 \neq y_1} \left\| \frac{b(x_1, \cdot) - b(y_1, \cdot)}{|x_1 - y_1|^{\beta_1}} \right\|_{\text{BMO}(\mathbb{R}^{d_2})}$$

the following oscillatory characterization holds.

3.6. Proposition. *We have*

$$\|b\|_{\dot{C}_{x_1}^{0, \alpha}(\text{BMO}_{x_2})} \sim \sup \left\{ \frac{1}{\ell(I_1)^\alpha} \frac{\text{osc}^{1, 1}(b, R)}{|R|} : R = I_1 \times I_2 \text{ rectangle} \right\}.$$

Proof. We denote the supremum on the right hand side by C_0 .

We again first prove the “ \gtrsim ” direction. For all $x_i \in I_i$ we have

$$\begin{aligned} & |b(x_1, x_2) - \langle b \rangle_{I_1, 1}(x_2) - \langle b \rangle_{I_2, 2}(x_1) + \langle b \rangle_{I_1 \times I_2}| \\ &= \left| \int_{I_1 \times I_2} (b(x_1, x_2) - b(x_1, y_2) - b(y_1, x_2) + b(y_1, y_2)) \, dy_1 \, dy_2 \right| \\ &= \left| \int_{I_1} |x_1 - y_1|^\alpha \left(\frac{b(x_1, x_2) - b(y_1, x_2)}{|x_1 - y_1|^\alpha} - \left\langle \frac{b(x_1, \cdot) - b(y_1, \cdot)}{|x_1 - y_1|^\alpha} \right\rangle_{I_2} \right) \, dy_1 \right| \end{aligned}$$

$$\lesssim \ell(I_1)^\alpha \int_{I_1} \left| \frac{b(x_1, x_2) - b(y_1, x_2)}{|x_1 - y_1|^\alpha} - \left\langle \frac{b(x_1, \cdot) - b(y_1, \cdot)}{|x_1 - y_1|^\alpha} \right\rangle_{I_2} \right| dy_1.$$

Then taking the average over $I_1 \times I_2$ and using Fubini we obtain the estimate as desired.

For “ \lesssim ” we fix $x_1, y_1 \in \mathbb{R}^{d_1}$ with $x_1 \neq y_1$. Similarly as Proposition 3.5 (just let $\alpha = 0$ there) we see that for all cubes $I_2 \subset \mathbb{R}^{d_2}$

$$\int_{I_2} |b_{x_1, y_1} - \langle b_{x_1, y_1} \rangle_{I_2}| \lesssim C_0 |x_1 - y_1|^\alpha,$$

where $b_{x_1, y_1} = b(x_1, \cdot) - b(y_1, \cdot)$. This proves the claim. \square

The final $\dot{C}^{0, \alpha}(X)$ type space of interest to us is the bi-parameter space $\dot{C}_{x_1}^{0, \alpha}(\dot{L}_{x_2}^r)$.

3.7. Proposition. *We have*

$$\|b\|_{\dot{C}_{x_1}^{0, \alpha}(\dot{L}_{x_2}^r)} \sim \sup \left\{ \frac{1}{|I_1| \ell(I_1)^\alpha} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \operatorname{osc}^{1,1}(b, I_1 \times S_2) : \right. \\ \left. I_1 \subset \mathbb{R}^{d_1}, \mathcal{S}_2 \text{ sparse}, \sum_{S_2 \in \mathcal{S}_2} |S_2| \lambda_{S_2}' \leq 1 \right\},$$

where $I_1 \subset \mathbb{R}^{d_1}$ is a cube and \mathcal{S}_2 is a sparse collection of cubes in \mathbb{R}^{d_2} .

Proof. “ \gtrsim ”: We apply Proposition 3.2 to the function $b(x_1, \cdot) - \langle b \rangle_{I_1, 1}$ for each $x_1 \in I_1$. This shows that

$$\sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \int_{S_2} |b(x_1, \cdot) - \langle b \rangle_{I_1, 1} - \langle b \rangle_{S_2, 2}(x_1) + \langle b \rangle_{I_1 \times S_2}| \\ \lesssim \|b(x_1, \cdot) - \langle b \rangle_{I_1, 1}\|_{\dot{L}_{x_2}^r} \lesssim \int_{I_1} \|b(x_1, \cdot) - b(y_1, \cdot)\|_{\dot{L}_{x_2}^r} dy_1 \lesssim \|b\|_{\dot{C}_{x_1}^{0, \alpha}(\dot{L}_{x_2}^r)} \ell(I_1)^\alpha.$$

The penultimate estimate is easy to see using Proposition 3.2. Integrating over $x_1 \in I_1$, we deduce the claimed bound.

“ \lesssim ”: Denoting by C_0 the right-hand side, we need to prove that $\|b(x_1, \cdot) - b(y_1, \cdot)\|_{\dot{L}_{x_2}^r} \lesssim C_0 |x_1 - y_1|^\alpha$ for all $x_1 \neq y_1 \in \mathbb{R}^{d_1}$. By Proposition 3.2, this is equivalent to

$$(3.8) \quad \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \int_{S_2} |b(x_1, \cdot) - b(y_1, \cdot) - \langle b(x_1, \cdot) - b(y_1, \cdot) \rangle_{S_2}| \lesssim C_0 |x_1 - y_1|^\alpha,$$

where \mathcal{S}_2 and λ_{S_2} are as in the claim. We then expand $b(x_1, \cdot)$ as in the proof of Proposition 3.1 (using the same notation as there):

$$b(x_1, \cdot) = \sum_{k=1}^{\infty} \left(\int_{Q_{k+1}(x_1)} b(z_1, \cdot) dz_1 - \int_{Q_k(x_1)} b(z_1, \cdot) dz_1 \right) + \int_{Q(\frac{1}{2}(x_1+y_1), \frac{1}{2}|x_1-y_1|)} b(z_1, \cdot) dz_1.$$

Doing the same with $b(y_1, \cdot)$ and using the triangle inequality, we find that

$$\begin{aligned}
& LHS(3.8) \\
& \leq \sum_{u_1=x_1, y_1} \sum_{k=1}^{\infty} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \int_{S_2} |\langle b \rangle_{Q_{k+1}(u_1), 1} - \langle b \rangle_{Q_k(u_1), 1} - (\langle b \rangle_{Q_{k+1}(u_1) \times S_2} - \langle b \rangle_{Q_k(u_1) \times S_2})| \\
& \leq \sum_{u_1=x_1, y_1} \sum_{k=1}^{\infty} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \int_{S_2} \int_{Q_{k+1}(u_1)} |b - \langle b \rangle_{Q_k(u_1), 1} - (\langle b \rangle_{S_2, 2} - \langle b \rangle_{Q_k(u_1) \times S_2})| \\
& \lesssim \sum_{u_1=x_1, y_1} \sum_{k=1}^{\infty} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \int_{S_2} \int_{Q_k(u_1)} |b - \langle b \rangle_{Q_k(u_1), 1} - \langle b \rangle_{S_2, 2} + \langle b \rangle_{Q_k(u_1) \times S_2}| \\
& \leq \sum_{u_1=x_1, y_1} \sum_{k=1}^{\infty} C_0 \ell(Q_k(u_1))^\alpha \lesssim \sum_{u_1=x_1, y_1} \sum_{k=1}^{\infty} C_0 (2^{-k} |x_1 - y_1|)^\alpha \lesssim C_0 |x_1 - y_1|^\alpha,
\end{aligned}$$

and this proves (3.8). \square

3.9. Remark. Note that it seems somewhat important that in the above proof we apply Proposition 3.2 directly to the defining condition $\|b(x_1, \cdot) - b(y_1, \cdot)\|_{\dot{L}_{x_2}^r} \lesssim |x_1 - y_1|^\alpha$, so that the extremizing \mathcal{S}_2 and λ_{S_2} depend on x_1, y_1 only, and only then apply the (considerations of) Proposition 3.1. Alternatively, it might occur to one to start with (a vector-valued version of) Proposition 3.1, reducing the proof to estimating $\int_{I_1} \|b(z_1, \cdot) - \langle b \rangle_{I_1, 1}\|_{\dot{L}_{x_2}^r} dz_1 \lesssim \ell(I_1)^\alpha$. If we now tried to continue with Proposition 3.2, applied to $b(z_1, \cdot) - \langle b \rangle_{I_1, 1}$ for each $z_1 \in I_i$, the resulting \mathcal{S}_2 and λ_{S_2} would in general depend on the integration parameter $z_1 \in I_1$.

Bounded mean oscillation. We define, for $1 \leq p_i < \infty$, the rectangular BMO norm

$$\|b\|_{\text{BMO}_{\text{rect}}(p_1, p_2)} = \sup_{R=I_1 \times I_2} \frac{\text{osc}^{p_1, p_2}(b, R)}{|I_1|^{1/p_1} |I_2|^{1/p_2}}.$$

This space is directly in the oscillatory form. It does not appear to enjoy the John–Nirenberg property, and so the choice of the exponents can matter. Usually in the literature

$$\|b\|_{\text{BMO}_{\text{rect}}} := \|b\|_{\text{BMO}_{\text{rect}}(2, 2)}.$$

For many purposes this is not the correct or optimal bi-parameter BMO space.

We will later have use for the smaller product BMO space, but it does not enjoy an oscillatory characterisation, and we will not discuss its definition now.

If X is a Banach space, we say that a locally integrable $b: \mathbb{R}^d \rightarrow X$ belongs to the X -valued BMO space $\text{BMO}(p, X)$, $0 < p < \infty$, if

$$\|b\|_{\text{BMO}(p, X)} := \sup_I \left(\int_I |b - \langle b \rangle_I|_X^p \right)^{1/p} < \infty.$$

By the Banach-valued John–Nirenberg theorem (see e.g. [26, Theorem 3.2.30]) we have that all these norms are equivalent when p varies, and we may again set

$$\|b\|_{\text{BMO}(X)} := \sup_I \int_I |b - \langle b \rangle_I|_X.$$

From this point on, we do not obtain oscillatory characterization for our spaces, but we can record one-sided estimates of oscillatory quantities by function space norms. These are of philosophical use. Later, we may obtain commutator lower bounds with respect to these potentially smaller oscillatory quantities. We start with the bi-parameter space $\text{BMO}_{x_1}(\dot{L}_{x_2}^r)$.

3.10. Lemma. *We have*

$$\sup \left\{ \frac{1}{|I_1|} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \text{osc}^{1,1}(b, I_1 \times S_2) : I_1 \subset \mathbb{R}^{d_1}, \mathcal{S}_2 \text{ sparse}, \sum_{S_2 \in \mathcal{S}_2} |S_2| \lambda_{S_2}^{r'} \leq 1 \right\} \\ \lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^r)},$$

where $I_1 \subset \mathbb{R}^{d_1}$ is a cube and \mathcal{S}_2 is a sparse collection of cubes in \mathbb{R}^{d_2} .

Proof. As in Proposition 3.7 we have by applying Proposition 3.2 to $b(x_1, \cdot) - \langle b \rangle_{I_1,1}$ that

$$\int_{I_1} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \int_{S_2} |b(x_1, \cdot) - \langle b \rangle_{I_1,1} - \langle b \rangle_{S_2,2}(x_1) + \langle b \rangle_{I_1 \times S_2}| \\ \lesssim \int_{I_1} \|b(x_1, \cdot) - \langle b \rangle_{I_1,1}\|_{\dot{L}_{x_2}^r} \lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^r)}. \quad \square$$

The space $\dot{L}^{r_1} \dot{L}^{r_2}$. The final condition that comes up in our commutator lower bounds is

$$C_0 := \sup \sum_{S_1 \in \mathcal{S}_1} \sum_{S_2 \in \mathcal{S}_2} \lambda_{1,S_1} \lambda_{2,S_2} \text{osc}^{1,1}(b, S_1 \times S_2),$$

where $r_i \in (1, \infty)$ and the supremum is taken over all sparse collections \mathcal{S}_i of cubes in \mathbb{R}^{d_i} and non-negative coefficients λ_{i,S_i} satisfying $\sum_{S_i \in \mathcal{S}_i} \lambda_{i,S_i}^{r'_i} |S_i| \leq 1$. We again only have the direction that

$$(3.11) \quad C_0 \lesssim \|b\|_{\dot{L}_{x_1}^{r_1} \dot{L}_{x_2}^{r_2}},$$

which is primarily of philosophical use. This can be seen by utilising Proposition 3.2 and its proof.

4. WEAK FACTORIZATION AND COROLLARIES

4.1. Definition (Non-degenerate kernels). Let K_i be a Calderón-Zygmund kernel on \mathbb{R}^{d_i} . We say that K_i is (symmetrically) non-degenerate if there is a constant $c_0 > 0$ such that for every $y_i \in \mathbb{R}^{d_i}$ and $r > 0$ there exists $x_i \in B(y_i, r)^c$ so that there holds

$$|K_i(x_i, y_i)| \geq \frac{1}{c_0 r^{d_i}} \quad \left(\text{and} \quad |K_i(y_i, x_i)| \geq \frac{1}{c_0 r^{d_i}} \right).$$

Notice that the x_i in the above definition necessarily satisfies $|x_i - y_i| \sim r$. If T_i is a CZO whose kernel K_i is a non-degenerate Calderón-Zygmund kernel, then we say that T_i is a non-degenerate CZO. However, in what follows we do not really need CZOs, we only need the kernels, and T_1 and T_1^* will just be notation for the integrals

$$T_1 h(x_1) = \int_{\mathbb{R}^{d_1}} K_1(x_1, y_1) h(y_1) dy_1, \quad x_1 \in \mathbb{R}^{d_1} \setminus \text{spt } h, \\ T_1^* h(x_1) = \int_{\mathbb{R}^{d_1}} K_1(y_1, x_1) h(y_1) dy_1, \quad x_1 \in \mathbb{R}^{d_1} \setminus \text{spt } h,$$

and similarly for T_2 .

4.2. Definition (The “reflected” cube \tilde{I}_i). Let K_i be a fixed non-degenerate Calderón–Zygmund kernel on \mathbb{R}^{d_i} , and fix a large constant $A \geq 3$. For each cube $I_i \subset \mathbb{R}^{d_i}$ with center c_{I_i} and sidelength $\ell(I_i)$, we define another cube $\tilde{I}_i \subset \mathbb{R}^{d_i}$ of the same size by choosing a center $c_{\tilde{I}_i}$, guaranteed by the non-degeneracy of K_i , so that

$$|c_{I_i} - c_{\tilde{I}_i}| \sim A\ell(I_i), \quad |K_i(c_{\tilde{I}_i}, c_{I_i})| \sim (A\ell(I_i))^{-d_i}.$$

Notice that $\text{dist}(I_i, \tilde{I}_i) \sim A\ell(I_i)$. If K_i is symmetrically non-degenerate, we require in addition that $|K_i(c_{I_i}, c_{\tilde{I}_i})| \sim (A\ell(I_i))^{-d_i}$. While the choice of \tilde{I}_i may not be unique in general, in the symmetric case we can and will make it in such a way that $\tilde{\tilde{I}}_i = I_i$.

By (the proof of) Proposition 2.2 in [24] we have for all $x_i \in \tilde{I}_i$ and $y_i \in I_i$ that

$$(4.3) \quad |K_i(x_i, y_i) - K_i(c_{\tilde{I}_i}, c_{I_i})| \leq \frac{CA^{-\alpha_i}}{(A\ell(I_i))^{d_i}}.$$

If K_1 and K_2 are two (symmetrically) non-degenerate Calderón–Zygmund kernels on \mathbb{R}^{d_1} and \mathbb{R}^{d_2} , and $I_i \subset \mathbb{R}^{d_i}$ are cubes, we of course define \tilde{I}_i with respect to K_i . For each rectangle $R = I_1 \times I_2 \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, we define $\tilde{R} := \tilde{I}_1 \times \tilde{I}_2$.

For a rectangle $R = I_1 \times I_2 \subset \mathbb{R}^{d_1+d_2} = \mathbb{R}^d$ and $p_1, p_2 \in [1, \infty]$ we denote

$$L_{x_1,0}^{p_1} L_{x_2,0}^{p_2}(R) = \left\{ f \in L_{x_1}^{p_1} L_{x_2}^{p_2}(R) : \int_{I_1} f = \int_{I_2} f = 0 \right\}.$$

We record and prove the following bi-parameter weak factorization theorem. A simpler one-parameter version is Lemma 2.5 of [24]. This will have important corollaries that we use to deduce the commutator lower bounds.

4.4. Lemma. *Let K_i be a non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$. If $R = I_1 \times I_2 \subset \mathbb{R}^d$ is a rectangle and $f \in L_{x_1,0}^1 L_{x_2,0}^1(R)$, then we can decompose*

$$f = 1_{\tilde{R}} T_1 T_2 h - T_2 h \cdot T_1^* 1_{\tilde{R}} - T_1 h \cdot T_2^* 1_{\tilde{R}} + h T_1^* T_2^* 1_{\tilde{R}} + \sum_{j=1}^3 \tilde{f}_j,$$

where each $\tilde{f}_j \in L_{x_1,0}^1 L_{x_2,0}^1(R_j)$ for $R_1 = \tilde{I}_1 \times I_2$, $R_2 = I_1 \times \tilde{I}_2$ and $R_3 = \tilde{R}$, and they satisfy

$$|\tilde{f}_1(x)| \lesssim A^{-\alpha_1} 1_{\tilde{I}_1}(x_1) \langle |f| \rangle_{I_1,1}(x_2), \quad |\tilde{f}_2(x)| \lesssim A^{-\alpha_2} \langle |f| \rangle_{I_2,2}(x_1) 1_{\tilde{I}_2}(x_2)$$

and

$$|\tilde{f}_3(x)| \lesssim A^{-\alpha_1} A^{-\alpha_2} 1_{\tilde{R}}(x) \langle |f| \rangle_R.$$

Moreover, $h \in L^1(R)$ satisfies

$$|h(x)| \lesssim A^d |f(x)|.$$

In particular, if $f \in L_{x_1,0}^{u_1} L_{x_2,0}^{u_2}(R)$, $1 \leq u_i \leq \infty$, then also

$$\|h\|_{L_{x_1}^{u_1} L_{x_2}^{u_2}} \lesssim A^d \|f\|_{L_{x_1}^{u_1} L_{x_2}^{u_2}} \quad \text{and} \quad \|\tilde{f}_j\|_{L_{x_1}^{u_1} L_{x_2}^{u_2}} \lesssim A^{-\min(\alpha_1, \alpha_2)} \|f\|_{L_{x_1}^{u_1} L_{x_2}^{u_2}},$$

where the implicit constants depend only on the dimensions and the kernel constants.

Proof. Denote $\phi = 1_{\tilde{R}} = 1_{\tilde{I}_1} \otimes 1_{\tilde{I}_2} =: \phi_1 \otimes \phi_2$. We write

$$f = \frac{f}{T_1^* T_2^* \phi} T_1^* T_2^* \phi = \phi T_1 T_2 h - T_2 h \cdot T_1^* \phi - T_1 h \cdot T_2^* \phi + h T_1^* T_2^* \phi + \tilde{f},$$

where

$$h := \frac{f}{T_1^* T_2^* \phi}$$

and

$$\begin{aligned} \tilde{f} &= T_1 h \cdot T_2^* \phi + T_2 h \cdot T_1^* \phi - \phi T_1 T_2 h \\ &= \phi_1 T_1 \left(\frac{f}{T_1^* \phi_1} \right) + \phi_2 T_2 \left(\frac{f}{T_2^* \phi_2} \right) - \phi T_1 \left(\frac{1}{T_1^* \phi_1} T_2 \left(\frac{f}{T_2^* \phi_2} \right) \right) \\ &= \tilde{f}_1 + \tilde{f}_2 + \tilde{f}_3. \end{aligned}$$

Everything is well-defined, since in fact we have the estimates $T_i^* 1_{\tilde{I}_i}(x_i) \gtrsim A^{-d_i}$ for $x \in I_i$ (remember that f is supported on R). To see this, fix $x_1 \in I_1$ and write

$$\begin{aligned} T_1^* 1_{\tilde{I}_1}(x_1) &= \int_{\tilde{I}_1} K_1(y_1, x_1) dy_1 \\ &= K_1(c_{\tilde{I}_1}, c_{I_1}) |I_1| + \int_{\tilde{I}_1} (K_1(y_1, x_1) - K_1(c_{\tilde{I}_1}, c_{I_1})) dy_1. \end{aligned}$$

Using (4.3) it is then easy to see that $|T_1^* 1_{\tilde{I}_1}(x_1)| \gtrsim A^{-d_1}$. The other estimate is symmetric. This also gives that

$$|h(x)| \lesssim A^d |f(x)|.$$

It is also immediate that $\int_{I_2} \tilde{f}_1 = 0$ and

$$\int_{\tilde{I}_1} \tilde{f}_1 = \int_{\mathbb{R}^{d_1}} \phi_1 T_1 \left(\frac{f}{T_1^* \phi_1} \right) = \int_{\mathbb{R}^{d_1}} T_1^* \phi_1 \frac{f}{T_1^* \phi_1} = \int_{I_1} f = 0.$$

Similarly, we can see that $\int_{I_1} \tilde{f}_2 = \int_{\tilde{I}_2} \tilde{f}_2 = \int_{\tilde{I}_1} \tilde{f}_3 = \int_{\tilde{I}_2} \tilde{f}_3 = 0$.

We now prove the pointwise estimates of f_i . Let $g: \mathbb{R}^d \rightarrow \mathbb{C}$ satisfy $g(x) = 0$ if $x_1 \notin I_1$ and $\int_{I_1} g = 0$. For $x = (x_1, x_2) \in \tilde{I}_1 \times \mathbb{R}^{d_2}$ we write

$$T_1 \left(\frac{g}{T_1^* \phi_1} \right) (x) = T_1 \left(\frac{g(\cdot, x_2)}{T_1^* \phi_1} - \frac{g(\cdot, x_2)}{|I_1| K_1(c_{\tilde{I}_1}, c_{I_1})} \right) (x_1) + \frac{1}{|I_1| K_1(c_{\tilde{I}_1}, c_{I_1})} T_1(g(\cdot, x_2))(x_1).$$

We have using (4.3) that

$$\begin{aligned} \left| \frac{1}{|I_1| K_1(c_{\tilde{I}_1}, c_{I_1})} T_1(g(\cdot, x_2))(x_1) \right| &\sim A^{d_1} \left| \int_{I_1} [K_1(x_1, y_1) - K_1(c_{\tilde{I}_1}, c_{I_1})] g(y_1, x_2) dy_1 \right| \\ &\lesssim A^{-\alpha_1} \int_{I_1} |g(\cdot, x_2)|. \end{aligned}$$

It is similarly straightforward that

$$\left| T_1 \left(\frac{g(\cdot, x_2)}{T_1^* \phi_1} - \frac{g(\cdot, x_2)}{|I_1| K_1(c_{\tilde{I}_1}, c_{I_1})} \right) (x_1) \right| \lesssim A^{-\alpha_1} \int_{I_1} |g(\cdot, x_2)|.$$

It follows that

$$(4.5) \quad \left| T_1 \left(\frac{g}{T_1^* \phi_1} \right) (x) \right| \lesssim A^{-\alpha_1} \int_{I_1} |g(\cdot, x_2)|, \quad x \in \tilde{I}_1 \times \mathbb{R}^{d_2}.$$

Applying this to $g = f$ we get

$$|\tilde{f}_1(x)| \lesssim A^{-\alpha_1} 1_{\tilde{I}_1}(x_1) \langle |f| \rangle_{I_1,1}(x_2).$$

The estimate

$$|\tilde{f}_2(x)| \lesssim A^{-\alpha_2} \langle |f| \rangle_{I_2,2}(x_1) 1_{\tilde{I}_2}(x_2)$$

is symmetric. Lastly, notice that

$$\tilde{f}_3 = \phi_1 T_1 \left(\frac{1}{T_1^* \phi_1} \phi_2 T_2 \left(\frac{f}{T_2^* \phi_2} \right) \right) = \phi_1 T_1 \left(\frac{\tilde{f}_2}{T_1^* \phi_1} \right)$$

so that applying (4.5) to $g = \tilde{f}_2$ (which is a legitimate choice) and the estimate of \tilde{f}_2 , we get that for $x \in \tilde{I}_1 \times \tilde{I}_2$ there holds

$$|\tilde{f}_3(x)| \lesssim A^{-\alpha_1} \int_{I_1} |\tilde{f}_2(y_1, x_2)| dy_1 \lesssim A^{-\alpha_1} A^{-\alpha_2} \int_R |f|.$$

The claimed $L^{u_1} L^{u_2}$ estimates are clear from the pointwise estimates. \square

4.6. *Remark.* A much weaker modulus of continuity like in [24] for the kernels K_1, K_2 would also work, but we prefer to use the standard estimates here for simplicity. The weaker assumption for K_1 would be

$$|K_1(x_1, y_1) - K_1(x'_1, y_1)| + |K_1(y_1, x_1) - K_1(y_1, x'_1)| \leq \frac{1}{|x_1 - y_1|^{d_1}} \omega \left(\frac{|x_1 - x'_1|}{|x_1 - y_1|} \right)$$

whenever $|x_1 - x'_1| \leq |x_1 - y_1|/2$. In these lower bound arguments it is only needed that the increasing modulus of continuity $\omega: [0, 1) \rightarrow [0, \infty)$ satisfies $\omega(t) \rightarrow 0$ when $t \rightarrow 0$.

The previous lemma will be used to control the oscillations (3.4) via the following basic observation. The idea is to decompose the dualizing f like above.

4.7. **Lemma.** For $v_1, v_2 \in [1, \infty]$, we have

$$\text{osc}^{v_1, v_2}(b, R) \sim \sup \left\{ \left| \int_R b f \right| : \|f\|_{L_{x_1,0}^{v_1} L_{x_2,0}^{v_2}(R)} \leq 1 \right\}.$$

Proof. By standard dualities, we have

$$\begin{aligned} \text{osc}^{v_1, v_2}(b, R) &= \sup \left\{ \left| \iint_{I_1 \times I_2} [b - \langle b \rangle_{I_1,1} - \langle b \rangle_{I_2,2} + \langle b \rangle_R] g \right| : \|g\|_{L_{x_1}^{v_1} L_{x_2}^{v_2}(R)} \leq 1 \right\} \\ &= \sup \left\{ \left| \iint_{I_1 \times I_2} b [g - \langle g \rangle_{I_1,1} - \langle g \rangle_{I_2,2} + \langle g \rangle_R] \right| : \|g\|_{L_{x_1}^{v_1} L_{x_2}^{v_2}(R)} \leq 1 \right\} \\ &\leq \sup \left\{ \left| \iint_{I_1 \times I_2} b g \right| : \|g\|_{L_0^{v_1} L_0^{v_2}(R)} \leq 4 \right\} \\ &= \sup \left\{ \left| \iint_{I_1 \times I_2} [b - \langle b \rangle_{I_1,1} - \langle b \rangle_{I_2,2} + \langle b \rangle_R] g \right| : \|g\|_{L_0^{v_1} L_0^{v_2}(R)} \leq 4 \right\} \\ &\leq 4 \text{osc}^{v_1, v_2}(b, R), \end{aligned}$$

so that all these quantities are actually comparable. \square

Motivated by this we define the following truncated oscillation, which is finite even when $b \in L^1_{\text{loc}}$. This is important for technical reasons – the absorption argument below requires a finite quantity. For $b \in L^1_{\text{loc}}$ and $N \in \mathbb{N}$ we define

$$\text{osc}_N^{v_1, v_2}(b, R) := \sup \left\{ \left| \iint_R bf \right| : \|f\|_{L^{v'_1} L^{v'_2}(R)} \leq 1 \text{ and } \|f\|_{L^\infty} \leq N \right\} < \infty.$$

4.8. Remark. When $v_1 = v_2 = 1$ – which is the case of most interest – we can directly work with $\text{osc}^{1,1}(b, R) \sim \text{osc}_N^{1,1}(b, R)$.

4.9. Lemma. Let K_i be a non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$. If we fix $A \geq 3$ in Definition 4.2 large enough (depending on the dimensions and kernel constants), then for each $b \in L^1_{\text{loc}}$, $1 \leq v_i \leq \infty$, $N \in \mathbb{N}$ and rectangle $R = I_1 \times I_2 \subset \mathbb{R}^d$ we have

$$\begin{aligned} \text{osc}_N^{v_1, v_2}(b, R) &\lesssim |\langle [T_1, [b, T_2]]h_0, 1_{R_3} \rangle| + |\langle [T_1^*, [b, T_2]]h_1, 1_{R_2} \rangle| \\ &\quad + |\langle [T_1, [b, T_2^*]]h_2, 1_{R_1} \rangle| + |\langle [T_1^*, [b, T_2^*]]h_3, 1_{R_0} \rangle| \end{aligned}$$

for some functions h_j with $\|h_j\|_{L^{v'_1} L^{v'_2}(R_j)} \lesssim 1$ and $\|h_j\|_{L^\infty(R_j)} \lesssim N$, where $R_0 = R$, $R_1 = \tilde{I}_1 \times I_2$, $R_2 = I_1 \times \tilde{I}_2$, $R_3 = \tilde{R}$.

If K_1 is symmetrically non-degenerate, we get

$$\begin{aligned} \text{osc}_N^{v_1, v_2}(b, R) &\lesssim |\langle [T_1, [b, T_2]]h_0, 1_{R_3} \rangle| + |\langle [T_1, [b, T_2]]h_1, 1_{R_2} \rangle| \\ &\quad + |\langle [T_1^*, [b, T_2^*]]h_2, 1_{R_1} \rangle| + |\langle [T_1^*, [b, T_2^*]]h_3, 1_{R_0} \rangle|. \end{aligned}$$

A symmetric estimate holds if K_2 is symmetrically non-degenerate. If both K_1 and K_2 are symmetrically non-degenerate, then

$$(4.10) \quad \text{osc}_N^{v_1, v_2}(b, R) \lesssim \sum_{j=0}^3 |\langle [T_1, [b, T_2]]h_j, 1_{\tilde{R}_j} \rangle|.$$

Proof. Let N be fixed. Fix $f \in L^{v'_1} L^{v'_2}(R)$ so that $\|f\|_{L^{v'_1} L^{v'_2}(R)} \leq 1$, $\|f\|_{L^\infty} \leq N$ and $\text{osc}_N^{v_1, v_2}(b, R) \lesssim |\langle b, f \rangle|$. We expand this f according to Lemma 4.4 to arrive at

$$\begin{aligned} \langle b, f \rangle &= \left\langle b, 1_{\tilde{R}} T_1 T_2 h_0 - T_2 h_0 \cdot T_1^* 1_{\tilde{R}} - T_1 h_0 \cdot T_2^* 1_{\tilde{R}} + h_0 T_1^* T_2^* 1_{\tilde{R}} \right\rangle + \sum_{j=1}^3 \langle b, \tilde{f}_j \rangle \\ &= \left\langle b T_1 T_2 h_0 - T_1 (b T_2 h_0) - T_2 (b T_1 h_0) + T_1 T_2 (b h_0), 1_{\tilde{R}} \right\rangle + \sum_{j=1}^3 \langle b, \tilde{f}_j \rangle \\ &= -\langle [T_1, [b, T_2]]h_0, 1_{\tilde{R}} \rangle + \sum_{j=1}^3 \langle b, \tilde{f}_j \rangle. \end{aligned}$$

Hence, we have

$$\text{osc}_N^{v_1, v_2}(b, R) \lesssim |\langle b, f \rangle| \leq |\langle [T_1, [b, T_2]]h_0, 1_{\tilde{R}} \rangle| + \sum_{j=1}^3 |\langle b, \tilde{f}_j \rangle|,$$

where, by Lemma 4.4,

$$|\langle b, \tilde{f}_j \rangle| \lesssim \text{osc}_N^{v_1, v_2}(b, R_j) A^{-\min(\alpha_1, \alpha_2)},$$

and thus

$$(4.11) \quad \text{osc}_N^{v_1, v_2}(b, R_0) \lesssim |\langle [T_1, [b, T_2]]h_0, 1_{R_3} \rangle| + A^{-\min(\alpha_1, \alpha_2)} \sum_{j \in \{1, 2, 3\}} \text{osc}_N^{v_1, v_2}(b, R_j)$$

where $\|h_0\|_{L^{v_1} L^{v_2}(R)} \lesssim A^d$ and $\|h_0\|_{L^\infty(R)} \lesssim A^d N$.

Notice that if \tilde{Q} is a reflected cube of Q with respect to a non-degenerate kernel K , then Q is a reflected cube of \tilde{Q} with respect to K^* . Therefore, we also get

$$(4.12) \quad \text{osc}_N^{v_1, v_2}(b, R_1) \lesssim |\langle [T_1^*, [b, T_2]]h_1, 1_{R_2} \rangle| + A^{-\min(\alpha_1, \alpha_2)} \sum_{j \in \{0, 2, 3\}} \text{osc}_N^{v_1, v_2}(b, R_j),$$

$$(4.13) \quad \text{osc}_N^{v_1, v_2}(b, R_2) \lesssim |\langle [T_1, [b, T_2^*]]h_2, 1_{R_1} \rangle| + A^{-\min(\alpha_1, \alpha_2)} \sum_{j \in \{0, 1, 3\}} \text{osc}_N^{v_1, v_2}(b, R_j)$$

and

$$(4.14) \quad \text{osc}_N^{v_1, v_2}(b, R_3) \lesssim |\langle [T_1^*, [b, T_2^*]]h_3, 1_{R_0} \rangle| + A^{-\min(\alpha_1, \alpha_2)} \sum_{j \in \{0, 1, 2\}} \text{osc}_N^{v_1, v_2}(b, R_j).$$

Combining (4.11), (4.12), (4.13) and (4.14), and fixing A large enough, we obtain by an absorption argument that

$$\begin{aligned} \text{osc}_N^{v_1, v_2}(b, R) &\leq \sum_{j=0}^3 \text{osc}_N^{v_1, v_2}(b, R_j) \\ &\lesssim |\langle [T_1, [b, T_2]]h_0, 1_{R_3} \rangle| + |\langle [T_1^*, [b, T_2]]h_1, 1_{R_2} \rangle| \\ &\quad + |\langle [T_1, [b, T_2^*]]h_2, 1_{R_1} \rangle| + |\langle [T_1^*, [b, T_2^*]]h_3, 1_{R_0} \rangle|, \end{aligned}$$

where $\|h_j\|_{L^{v_1} L^{v_2}(R_j)} \lesssim 1$ and $\|h_j\|_{L^\infty(R_j)} \lesssim N$.

Next, if K is *symmetrically* non-degenerate and \tilde{Q} is a reflected cube of Q with respect to K , then Q is a reflected cube of \tilde{Q} with respect to K and \tilde{Q} is also a reflected cube of Q with respect to K^* . Suppose K_1 is symmetrically non-degenerate. Then we can replace (4.12) by

$$(4.15) \quad \text{osc}_N^{v_1, v_2}(b, R_1) \lesssim |\langle [T_1, [b, T_2]]h_1, 1_{R_2} \rangle| + A^{-\min(\alpha_1, \alpha_2)} \sum_{j \in \{0, 2, 3\}} \text{osc}_N^{v_1, v_2}(b, R_j),$$

and (4.13) by

$$(4.16) \quad \text{osc}_N^{v_1, v_2}(b, R_2) \lesssim |\langle [T_1^*, [b, T_2^*]]h_2, 1_{R_1} \rangle| + A^{-\min(\alpha_1, \alpha_2)} \sum_{j \in \{0, 1, 3\}} \text{osc}_N^{v_1, v_2}(b, R_j)$$

The other symmetry statements are similar. This ends our proof. \square

From this point on, given non-degenerate kernels K_i , we always consider A fixed large enough so that the conclusion of Lemma 4.9 hold.

4.17. *Remark.* We have that $[T_1, [b, T_2]]^* = [T_1^*, [b, T_2^*]]$, but a partial adjoint like $[T_1^*, [b, T_2]]$ can more easily be unbounded even if $[T_1, [b, T_2]]$ is bounded. Thus, it can be useful to be able to eliminate these partial adjoint terms if at least one of the kernels is symmetrically non-degenerate.

4.A. Off-support constants. Notice that if we take a pairing, such as $|\langle [T_1, [b, T_2]]h_0, 1_{\tilde{R}} \rangle|$, from the conclusion of Lemma 4.9, then h_0 is supported on $I_1 \times I_2$ and $\tilde{R} = \tilde{I}_1 \times \tilde{I}_2$ satisfies $\text{dist}(I_i, \tilde{I}_i) \sim \ell(I_i)$. Moreover, the case $v_1 = v_2 = 1$ is a bit special, as then $\|h_j\|_{L^\infty(R_j)} \lesssim 1$. Otherwise, we only have the estimate $\|h_j\|_{L^{v'_1}L^{v'_2}(R_j)} \lesssim 1$ at our disposal.

We fix $b \in L^1_{\text{loc}}$ and non-degenerate kernels K_i – the off-support constants will depend on these but this will not be denoted. We first define off-support constants that are useful in the $v_1 = v_2 = 1$ situation – which is a case that is encountered often. For $s_1, s_2, t_1, t_2 \in (1, \infty)$ and $B(x, y) = b(x_1, x_2) - b(x_1, y_2) - b(y_1, x_2) + b(y_1, y_2)$ define

$$\text{Off}_{(s_1, s_2)}^{(t_1, t_2)} = \sup \frac{1}{|J_1|^{1+1/s_1-1/t_1}} \frac{1}{|J_2|^{1+1/s_2-1/t_2}} \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} B(x, y) K_1(x_1, y_1) K_2(x_2, y_2) f_1(y) f_2(x) dy dx \right|,$$

where the supremum is taken over rectangles $P_1 = J_1 \times J_2$ and $P_2 = L_1 \times L_2$ with

$$\ell(J_i) = \ell(L_i) \quad \text{and} \quad \text{dist}(J_i, L_i) \sim \ell(J_i)$$

and over functions $f_i \in L^\infty(P_i)$ with $\|f_j\|_{L^\infty} \leq 1$. We also denote $\text{Off}_{(s_1, s_2)}^{(t_1, t_2), i^*}$ the constant where K_i is replaced by K_i^* .

If T_i is a non-degenerate CZO with kernel K_i , then

$$\text{Off}_{(s_1, s_2)}^{(t_1, t_2)} \leq \|[T_1, [b, T_2]]\|_{L^{s_1}_{x_1} L^{s_2}_{x_2} \rightarrow L^{t_1}_{x_1} L^{t_2}_{x_2}}$$

whenever the commutator exists as a bounded linear mapping. This explains why these weaker off-support constants make sense.

4.18. Theorem. *Let K_i be a non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$, and let $b \in L^1_{\text{loc}}$. Let $p_1, p_2, q_1, q_2 \in (1, \infty)$. Then we have for all rectangles $R = I_1 \times I_2$ that*

$$\frac{\text{osc}^{1,1}(b, R)}{|R|} \lesssim \left(\text{Off}_{(p_1, p_2)}^{(q_1, q_2)} + \text{Off}_{(s_1, s_2)}^{(t_1, t_2), 1^*} \right) |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}}$$

with any $s_i, t_i \in (1, \infty)$ satisfying $1/p_i - 1/q_i = 1/s_i - 1/t_i$, $i = 1, 2$.

If, in addition, one of the kernels K_i is symmetrically non-degenerate, then

$$\frac{\text{osc}^{1,1}(b, R)}{|R|} \lesssim \text{Off}_{(p_1, p_2)}^{(q_1, q_2)} |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}}.$$

Proof. By Lemma 4.9 we have for each rectangle $R = I_1 \times I_2 \subset \mathbb{R}^d$ that

$$\begin{aligned} \text{osc}^{1,1}(b, R) &\lesssim |\langle [T_1, [b, T_2]]h_0, 1_{R_3} \rangle| + |\langle [T_1^*, [b, T_2]]h_1, 1_{R_2} \rangle| \\ &\quad + |\langle [T_1, [b, T_2^*]]h_2, 1_{R_1} \rangle| + |\langle [T_1^*, [b, T_2^*]]h_3, 1_{R_0} \rangle| \end{aligned}$$

for some functions h_j with $\|h_j\|_{L^\infty(R_j)} \lesssim 1$, where $R_0 = R$, $R_1 = \tilde{I}_1 \times I_2$, $R_2 = I_1 \times \tilde{I}_2$, $R_3 = \tilde{R}$. Clearly, we have

$$|\langle [T_1, [b, T_2]]h_0, 1_{R_3} \rangle| + |\langle [T_1^*, [b, T_2^*]]h_3, 1_{R_0} \rangle| \lesssim \text{Off}_{(p_1, p_2)}^{(q_1, q_2)} |R| |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}}$$

and

$$|\langle [T_1^*, [b, T_2]]h_1, 1_{R_2} \rangle| + |\langle [T_1, [b, T_2^*]]h_2, 1_{R_1} \rangle| \lesssim \text{Off}_{(s_1, s_2)}^{(t_1, t_2), 1^*} |R| |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}}.$$

The symmetry claim follows from the symmetry claims of Lemma 4.9. \square

We now define larger off-support constants that can be used to control $\text{osc}_N^{v_1, v_2}(b, R)$, uniformly on N , for general v_1, v_2 . For $s_1, s_2, t_1, t_2 \in (1, \infty)$ define

$$\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2)} = \sup \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} B(x, y) K_1(x_1, y_1) K_2(x_2, y_2) f_1(y) f_2(x) dy dx \right|,$$

where the supremum is taken over rectangles $P_1 = J_1 \times J_2$ and $P_2 = L_1 \times L_2$ with

$$\ell(J_i) = \ell(L_i) \quad \text{and} \quad \text{dist}(J_i, L_i) \sim \ell(J_i)$$

and over functions $f_j \in L^\infty(P_j)$ with

$$\|f_1\|_{L_{x_1}^{s_1} L_{x_2}^{s_2}} \leq 1 \quad \text{and} \quad \|f_2\|_{L_{x_1}^{t_1} L_{x_2}^{t_2}} \leq 1.$$

We also define the dual off-support constants $\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2), 1*}$, $\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2), 2*}$ and $\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2), *}$ in the natural way – simply replace (K_1, K_2) with (K_1^*, K_2) , (K_1, K_2^*) and (K_1^*, K_2^*) , respectively. Notice that $\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2)} \leq \widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2)}$, and if T_i is a non-degenerate CZO with kernel K_i , then $\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2)} \leq \|[T_1, [b, T_2]]\|_{L_{x_1}^{s_1} L_{x_2}^{s_2} \rightarrow L_{x_1}^{t_1} L_{x_2}^{t_2}}$ whenever the commutator exists as a bounded linear mapping. Notice also that $\widetilde{\text{Off}}_{(s_1, s_2)}^{(t_1, t_2), *}$ = $\widetilde{\text{Off}}_{(t'_1, t'_2)}^{(s'_1, s'_2)}$.

4.19. Theorem. *Let K_i be a non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$, and let $b \in L_{\text{loc}}^1$. Let $p_1, p_2, q_1, q_2 \in (1, \infty)$. Then we have for all rectangles $R = I_1 \times I_2$ that*

$$\frac{\text{osc}^{p'_1, p'_2}(b, R)}{|I_1|^{1/p'_1} |I_2|^{1/p'_2}} \lesssim \left(\widetilde{\text{Off}}_{(p_1, p_2)}^{(q_1, q_2)} + \widetilde{\text{Off}}_{(p_1, p_2)}^{(q_1, q_2), 1*} + \widetilde{\text{Off}}_{(p_1, p_2)}^{(q_1, q_2), 2*} + \widetilde{\text{Off}}_{(q'_1, q'_2)}^{(p'_1, p'_2)} \right) |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}}.$$

If one of the kernels K_i is symmetrically non-degenerate, we can drop the terms $\widetilde{\text{Off}}_{(p_1, p_2)}^{(q_1, q_2), 1}$ and $\widetilde{\text{Off}}_{(p_1, p_2)}^{(q_1, q_2), 2*}$, and if both K_1 and K_2 are symmetrically non-degenerate, then*

$$\frac{\text{osc}^{p'_1, p'_2}(b, R)}{|I_1|^{1/p'_1} |I_2|^{1/p'_2}} \lesssim \widetilde{\text{Off}}_{(p_1, p_2)}^{(q_1, q_2)} |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}}.$$

Proof. Follows directly from Lemma 4.9 similarly to Theorem 4.18. \square

Finally, we need one more off-support constant to deal with the calculations arising from considerations involving the space \dot{L}^r . Recall that characterizations of this space, such as Proposition 3.2, involve bounding a sum of oscillations. To this end, define $B(x, y) = b(x_1, x_2) - b(x_1, y_2) - b(y_1, x_2) + b(y_1, y_2)$ and for $p_1, p_2, q_1, q_2 \in (1, \infty)$ set

$$\text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)} = \sup \frac{\sum_{i=1}^N \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} B(x, y) K_1(x_1, y_1) K_2(x_2, y_2) f_{1,i}(y) f_{2,i}(x) dy dx \right|}{\left\| \sum_{i=1}^N \|f_{1,i}\|_{L^\infty 1_{P_{1,i}}} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \left\| \sum_{i=1}^N \|f_{2,i}\|_{L^\infty 1_{P_{2,i}}} \right\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}}},$$

where the supremum is taken over rectangles $P_{1,i} = J_{1,i} \times J_{2,i}$ and $P_{2,i} = L_{1,i} \times L_{2,i}$ with

$$\ell(J_{k,i}) = \ell(L_{k,i}) \quad \text{and} \quad \text{dist}(J_{k,i}, L_{k,i}) \sim \ell(J_{k,i})$$

and over functions $f_{k,i} \in L^\infty(P_{k,i})$, $k = 1, 2, i = 1, \dots, N$.

Notice that for linear operators U we have the random sign trick

$$\sum_{i=1}^N \langle U f_i, g_i \rangle = \mathbb{E} \left\langle U \left(\sum_{i=1}^N \varepsilon_i f_i \right), \sum_{j=1}^N \varepsilon_j g_j \right\rangle.$$

Thus, for functions $f_{k,i}$ as in the definition of the constant $\text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)}$ we have for certain fixed signs σ_i and random signs ε_i that

$$\begin{aligned} & \sum_{i=1}^N \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} B(x, y) K_1(x_1, y_1) K_2(x_2, y_2) f_{1,i}(y) f_{2,i}(x) \, dy \, dx \right| \\ &= \sum_{i=1}^N \sigma_i \langle [T_1, [b, T_2]] f_{1,i}, f_{2,i} \rangle \\ &= \mathbb{E} \left\langle [T_1, [b, T_2]] \left(\sum_{i=1}^N \sigma_i \varepsilon_i f_{1,i} \right), \sum_{j=1}^N \varepsilon_j f_{2,j} \right\rangle \\ &\leq \| [T_1, [b, T_2]] \|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \mathbb{E} \left\| \sum_{i=1}^N \sigma_i \varepsilon_i f_{1,i} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \left\| \sum_{i=1}^N \varepsilon_i f_{2,i} \right\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} \\ &\leq \| [T_1, [b, T_2]] \|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \left\| \sum_{i=1}^N \| f_{1,i} \|_{L^\infty} 1_{P_{1,i}} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \left\| \sum_{i=1}^N \| f_{2,i} \|_{L^\infty} 1_{P_{2,i}} \right\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} \end{aligned}$$

whenever the commutator exists as a bounded operator. Therefore, we have $\text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)} \leq \| [T_1, [b, T_2]] \|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}}$ whenever the commutator exists and so $\text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)}$ is a reasonable off-support constant.

We now state various results, where sums of oscillations are controlled by this off-support norm. Having seen how to deal with general kernels, we prefer to state these with the simplifying assumption that the underlying kernels are **symmetrically** non-degenerate. In all of these statements

$$\frac{1}{r_i} := \frac{1}{q_i} - \frac{1}{p_i}, \quad \text{if } p_i > q_i.$$

4.20. Lemma. *Let K_i be a symmetrically non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$, and let $b \in L_{\text{loc}}^1$. Let $p_1, p_2, q_1, q_2 \in (1, \infty)$ with $p_2 > q_2$. For all sparse collections \mathcal{S}_2 of cubes in \mathbb{R}^{d_2} , non-negative coefficients λ_{S_2} satisfying $\sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2} |S_2| \leq 1$ and cubes $I_1 \subset \mathbb{R}^{d_1}$, we have*

$$\sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \text{osc}^{1,1}(b, I_1 \times S_2) \lesssim \text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)} |I_1|^{1+(1/p_1-1/q_1)}.$$

Proof. Using Lemma 4.9 we estimate

$$\sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \text{osc}^{1,1}(b, I_1 \times S_2) \lesssim \sum_{i=0}^3 \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} |\langle [T_1, [b, T_2]] h_{(I_1 \times S_2)_i}, 1_{(I_1 \times S_2)_i} \rangle|$$

for some $h_{(I_1 \times S_2)_i} \in L^\infty((I_1 \times S_2)_i)$ with $\|h_{(I_1 \times S_2)_i}\|_{L^\infty} \lesssim 1$. Using the algebra

$$\frac{r'_2}{p_2} + \frac{r'_2}{q'_2} = r'_2 \left(\frac{1}{p_2} + 1 - \frac{1}{q_2} \right) = r'_2 \left(1 - \frac{1}{r_2} \right) = 1$$

we write this in the form

$$\sum_{i=0}^3 \sum_{S_2 \in \mathcal{S}_2} |\langle [T_1, [b, T_2]](\lambda_{S_2}^{r'_2/p_2} h_{(I_1 \times S_2)_i}), \lambda_{S_2}^{r'_2/q'_2} 1_{(I_1 \times S_2)_i} \rangle|.$$

This can directly be dominated with

$$\text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)} \sum_{i=0}^3 \left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2/p_2} 1_{(I_1 \times S_2)_i} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2/q'_2} 1_{(I_1 \times S_2)_i} \right\|_{L_{x_1}^{q'_1} L_{x_2}^{q'_2}}.$$

Here $1_{(I_1 \times S_2)_i} \leq 1_{I_1^* \times S_2^*}$, where $I_1^* \supset I_1 \cup \tilde{I}_1$ and $S_2^* \supset S_2 \cup \tilde{S}_2$ are concentric dilations of I_1 and S_2 by a bounded factor. Since the collection $\{S_2^* : S_2 \in \mathcal{S}_2\}$ is still sparse (with a different sparseness constant), it follows from (3.3) and $\sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2} |S_2| \leq 1$ that

$$\begin{aligned} & \left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2/p_2} 1_{(I_1 \times S_2)_i} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \leq \left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2/p_2} 1_{I_1^* \times S_2^*} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \\ & = \|1_{I_1^*}\|_{L^{p_1}} \left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2/p_2} 1_{S_2^*} \right\|_{L^{p_2}} \lesssim |I_1|^{1/p_1} \left(\sum_{S_2 \in \mathcal{S}_2} (\lambda_{S_2}^{r'_2/p_2})^{p_2} |S_2| \right)^{1/p_2} \leq |I_1|^{1/p_1}. \end{aligned}$$

Similarly, using $1_{(I_1 \times S_2)_i} \lesssim 1_{I_1^* \times S_2^*}$, we conclude that

$$\left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2}^{r'_2/q'_2} 1_{(I_1 \times S_2)_i} \right\|_{L_{x_1}^{q'_1} L_{x_2}^{q'_2}} \lesssim |I_1|^{1/q'_1},$$

and this ends the proof. \square

The symmetric version reads as follows.

4.21. Lemma. *Let K_i be a symmetrically non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$, and let $b \in L_{\text{loc}}^1$. Let $p_1, p_2, q_1, q_2 \in (1, \infty)$ with $p_1 > q_1$. For all sparse collections \mathcal{S}_1 of cubes in \mathbb{R}^{d_1} , non-negative coefficients λ_{S_1} satisfying $\sum_{S_1 \in \mathcal{S}_1} \lambda_{S_1}^{r'_1} |S_1| \leq 1$ and cubes $I_2 \subset \mathbb{R}^{d_2}$, we have*

$$\sum_{S_1 \in \mathcal{S}_1} \lambda_{S_1} \text{osc}^{1,1}(b, S_1 \times I_2) \lesssim \text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)} |I_2|^{1+(1/p_2-1/q_2)}.$$

Finally, we control the following double sum.

4.22. Lemma. *Let K_i be a symmetrically non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$, and let $b \in L_{\text{loc}}^1$. Let $p_1, p_2, q_1, q_2 \in (1, \infty)$ with $p_1 > q_1$ and $p_2 > q_2$. For all sparse collections \mathcal{S}_i of cubes in \mathbb{R}^{d_i} and non-negative coefficients λ_{i, S_i} satisfying $\sum_{S_i \in \mathcal{S}_i} \lambda_{i, S_i}^{r'_i} |S_i| \leq 1$ we have*

$$\sum_{S_1 \in \mathcal{S}_1} \sum_{S_2 \in \mathcal{S}_2} \lambda_{1, S_1} \lambda_{2, S_2} \text{osc}^{1,1}(b, S_1 \times S_2) \lesssim \text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)}.$$

Proof. Using Lemma 4.9 we estimate

$$\begin{aligned}
& \sum_{S_1 \in \mathcal{S}_1} \sum_{S_2 \in \mathcal{S}_2} \lambda_{1,S_1} \lambda_{2,S_2} \text{osc}^{1,1}(b, S_1 \times S_2) \\
& \lesssim \sum_{i=0}^3 \sum_{S_1 \in \mathcal{S}_1} \sum_{S_2 \in \mathcal{S}_2} \lambda_{1,S_1} \lambda_{2,S_2} |\langle [T_1, [b, T_2]] h_{(S_1 \times S_2)_i}, 1_{(S_1 \times S_2)_i} \rangle| \\
& = \sum_{i=0}^3 \sum_{S_1 \in \mathcal{S}_1} \sum_{S_2 \in \mathcal{S}_2} |\langle [T_1, [b, T_2]] (\lambda_{1,S_1}^{r'_1/p_1} \lambda_{2,S_2}^{r'_2/p_2} h_{(S_1 \times S_2)_i}), \lambda_{1,S_1}^{r'_1/q'_1} \lambda_{2,S_2}^{r'_2/q'_2} 1_{(S_1 \times S_2)_i} \rangle| \\
& \leq \sum_{i=0}^3 \text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)} \left\| \sum_{\substack{S_1 \in \mathcal{S}_1 \\ S_2 \in \mathcal{S}_2}} \lambda_{1,S_1}^{r'_1/p_1} \lambda_{2,S_2}^{r'_2/p_2} 1_{(S_1 \times S_2)_i} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \times \\
& \qquad \qquad \qquad \left\| \sum_{\substack{S_1 \in \mathcal{S}_1 \\ S_2 \in \mathcal{S}_2}} \lambda_{1,S_1}^{r'_1/q'_1} \lambda_{2,S_2}^{r'_2/q'_2} 1_{(S_1 \times S_2)_i} \right\|_{L_{x_1}^{q'_1} L_{x_2}^{q'_2}}.
\end{aligned}$$

Using the notation of the proof of Lemma 4.20 we have

$$\begin{aligned}
& \left\| \sum_{\substack{S_1 \in \mathcal{S}_1 \\ S_2 \in \mathcal{S}_2}} \lambda_{1,S_1}^{r'_1/p_1} \lambda_{2,S_2}^{r'_2/p_2} 1_{(S_1 \times S_2)_i} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \\
& \leq \left\| \sum_{\substack{S_1 \in \mathcal{S}_1 \\ S_2 \in \mathcal{S}_2}} \lambda_{1,S_1}^{r'_1/p_1} \lambda_{2,S_2}^{r'_2/p_2} 1_{S_1^* \times S_2^*} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \\
& = \left\| \sum_{S_1 \in \mathcal{S}_1} \lambda_{1,S_1}^{r'_1/p_1} 1_{S_1^*} \right\|_{L_{x_1}^{p_1}} \left\| \sum_{S_2 \in \mathcal{S}_2} \lambda_{2,S_2}^{r'_2/p_2} 1_{S_2^*} \right\|_{L_{x_2}^{p_2}},
\end{aligned}$$

where for both $i = 1, 2$ we have by the sparseness of the family $\{S_i^* : S_i \in \mathcal{S}_i\}$ that

$$\left\| \sum_{S_i \in \mathcal{S}_i} \lambda_{i,S_i}^{r'_i/p_i} 1_{S_i^*} \right\|_{L_{x_i}^{p_i}} \lesssim \left(\sum_{S_i \in \mathcal{S}_i} (\lambda_{i,S_i}^{r'_i/p_i})^{p_i} |S_i| \right)^{1/p_i} \leq 1,$$

and the estimate for the $L_{x_1}^{q'_1} L_{x_2}^{q'_2}$ norm is completely analogous. \square

5. NECESSARY CONDITIONS FOR COMMUTATOR BOUNDEDNESS

We have already done the main work concerning the commutator lower bounds (necessary conditions); oscillatory characterizations of function spaces from Section 3 combined with the control of oscillations by off-support constants, Theorem 4.18, Theorem 4.19, Lemma 4.20, Lemma 4.21 and Lemma 4.22 will yield the results.

Many of our necessary conditions are optimal; we will later be able to prove the sufficiency of the obtained necessary condition, see Theorem 1.3. A reason, but not the only reason, a lower bound might not be optimal is that the techniques used to obtain the lower bounds cannot distinguish between symmetric cases like

$$p_1 < q_1 \text{ and } p_2 = q_2$$

and

$$p_1 = q_1 \text{ and } p_2 < q_2,$$

and yield symmetric lower bounds: in the first case we can bound $\|b\|_{\dot{C}_{x_1}^{0,\beta_1}(\text{BMO}_{x_2})}$ and in the second case $\|b\|_{\dot{C}_{x_2}^{0,\beta_2}(\text{BMO}_{x_1})}$ by $\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}}$. Somewhat strikingly, both of these conditions happen to be optimal. However, e.g. in the upper right corner of the table of Theorem 1.3, where $p_1 > q_1$ and $p_2 < q_2$, the conditions do not match, while in the symmetric case $p_1 < q_1$ and $p_2 > q_2$ (the lower left corner) they do match.

To have clean statements and to avoid repetition, we will state all of the commutator lower bounds for *symmetrically* non-degenerate kernels K_i . From the results of Section 4 it is clear that this is not necessary – the general case just requires various dual off-support constants, and, as we saw, often it is enough that only one of the kernels is symmetrically non-degenerate to get clean statements.

Sometimes we only obtain a necessary condition in terms of an oscillatory condition that we cannot relate to a function space norm. These are recorded in this section but are not visible in the table of Theorem 1.3 (the three cases on the lower right corner).

In what follows we always have $p_1, p_2, q_1, q_2 \in (1, \infty)$, and

$$\beta_i := d_i \left(\frac{1}{p_i} - \frac{1}{q_i} \right), \quad \text{if } p_i < q_i; \quad \frac{1}{r_i} := \frac{1}{q_i} - \frac{1}{p_i}, \quad \text{if } p_i > q_i.$$

Moreover, we also let $b \in L_{\text{loc}}^1$ and K_i be a symmetrically non-degenerate kernel on \mathbb{R}^{d_i} , $i = 1, 2$. This data is not repeated in the statements of the results.

5.A. The condition $\dot{C}^{0,\beta_1}(\dot{C}^{0,\beta_2})$.

5.1. Proposition. *If $p_1 < q_1$ and $p_2 < q_2$, then*

$$\|b\|_{\dot{C}^{0,\beta_1}(\dot{C}^{0,\beta_2})} \lesssim \text{Off}_{(p_1,p_2)}^{(q_1,q_2)}.$$

Proof. By Proposition 3.5 we have

$$\|b\|_{\dot{C}^{0,\beta_1}(\dot{C}^{0,\beta_2})} \sim \sup \left\{ \frac{1}{\ell(I_1)^{\beta_1} \ell(I_2)^{\beta_2}} \frac{\text{osc}^{1,1}(b, I_1 \times I_2)}{|I_1| |I_2|} : I_i \subset \mathbb{R}^{d_i} \text{ is a cube} \right\},$$

where, by Theorem 4.18,

$$\frac{1}{\ell(I_1)^{\beta_1} \ell(I_2)^{\beta_2}} \frac{\text{osc}^{1,1}(b, I_1 \times I_2)}{|I_1| |I_2|} \lesssim \frac{1}{\ell(I_1)^{\beta_1} \ell(I_2)^{\beta_2}} \text{Off}_{(p_1,p_2)}^{(q_1,q_2)} |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} |I_2|^{\frac{1}{p_2} - \frac{1}{q_2}} = \text{Off}_{(p_1,p_2)}^{(q_1,q_2)}.$$

□

5.B. The condition $\dot{C}^{0,\beta}(\text{BMO})$.

5.2. Proposition. *If $p_1 < q_1$ and $p_2 = q_2$, then*

$$\|b\|_{\dot{C}_{x_1}^{0,\beta_1}(\text{BMO}_{x_2})} \lesssim \text{Off}_{(p_1,p_2)}^{(q_1,p_2)}.$$

If $p_1 = q_1$ and $p_2 < q_2$, then

$$\|b\|_{\dot{C}_{x_2}^{0,\beta_2}(\text{BMO}_{x_1})} \lesssim \text{Off}_{(p_1,p_2)}^{(p_1,q_2)}.$$

Proof. Let $p_1 < q_1$ and $p_2 = q_2$. By Proposition 3.6 we have

$$\|b\|_{\dot{C}_{x_1}^{0,\beta_1}(\text{BMO}_{x_2})} \sim \sup \left\{ \frac{1}{\ell(I_1)^{\beta_1}} \frac{\text{osc}^{1,1}(b, I_1 \times I_2)}{|I_1| |I_2|} : I_i \subset \mathbb{R}^{d_i} \text{ is a cube} \right\},$$

where, by Theorem 4.18,

$$\frac{1}{\ell(I_1)^{\beta_1}} \frac{\text{osc}^{1,1}(b, I_1 \times I_2)}{|I_1||I_2|} \lesssim \frac{1}{\ell(I_1)^{\beta_1}} \text{Off}_{(p_1, p_2)}^{(q_1, p_2)} |I_1|^{\frac{1}{p_1} - \frac{1}{q_1}} = \text{Off}_{(p_1, p_2)}^{(q_1, p_2)}.$$

The proof of the case $p_1 = q_1$ and $p_2 < q_2$ is symmetric. \square

5.C. The condition BMO_{rect} . We record the following lower bound concerning the diagonal of Theorem 1.3, even though our real focus is on the off-diagonal cases.

5.3. Proposition. *If $p_1 = q_1$ and $p_2 = q_2$ we have*

$$\|b\|_{\text{BMO}_{\text{rect}}(p'_1, p'_2)} \lesssim \widetilde{\text{Off}}_{(p_1, p_2)}^{(p_1, p_2)}.$$

Proof. By definition and Theorem 4.19 we have

$$\|b\|_{\text{BMO}_{\text{rect}}(p'_1, p'_2)} = \sup_{R=I_1 \times I_2} \frac{\text{osc}^{p'_1, p'_2}(b, R)}{|I_1|^{1/p'_1} |I_2|^{1/p'_2}} \lesssim \widetilde{\text{Off}}_{(p_1, p_2)}^{(p_1, p_2)}.$$

\square

5.D. Conditions related to $\dot{C}^{0, \beta}(\dot{L}^r)$ and $\text{BMO}(\dot{L}^r)$.

5.4. Proposition. *If $p_1 < q_1$ and $p_2 > q_2$ then*

$$\|b\|_{\dot{C}_{x_1}^{0, \beta_1}(\dot{L}_{x_2}^{r_2})} \lesssim \text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)}.$$

If $p_1 = q_1$ and $p_2 > q_2$ then

$$\sup \left\{ \frac{1}{|I_1|} \sum_{S_2 \in \mathcal{S}_2} \lambda_{S_2} \text{osc}^{1,1}(b, I_1 \times S_2) : I_1 \subset \mathbb{R}^{d_1}, \mathcal{S}_2 \text{ sparse}, \sum_{S_2 \in \mathcal{S}_2} |S_2| \lambda_{S_2}^{r'_2} \leq 1 \right\} \lesssim \text{Off}_{(p_1, p_2), \Sigma}^{(p_1, q_2)},$$

where the supremum on the left hand side is also dominated by $\|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})}$. The symmetric statements hold if $p_1 > q_1$ and $p_2 < q_2$ or $p_1 > q_1$ and $p_2 = q_2$.

Proof. Follows by combining Lemma 4.20 and Lemma 4.21 with Proposition 3.7 and Lemma 3.10. \square

5.E. A condition related to $\dot{L}^{r_1}(\dot{L}^{r_2})$.

5.5. Proposition. *If $p_1 > q_1$ and $p_2 > q_2$, we have*

$$\sup \sum_{S_1 \in \mathcal{S}_1} \sum_{S_2 \in \mathcal{S}_2} \lambda_{1, S_1} \lambda_{2, S_2} \text{osc}^{1,1}(b, S_1 \times S_2) \lesssim \text{Off}_{(p_1, p_2), \Sigma}^{(q_1, q_2)},$$

where the supremum is over all sparse collections \mathcal{S}_i of cubes in \mathbb{R}^{d_i} and non-negative coefficients λ_{i, S_i} satisfying $\sum_{S_i \in \mathcal{S}_i} \lambda_{i, S_i}^{r'_i} |S_i| \leq 1$. Moreover, the supremum on the left hand side is also dominated by $\|b\|_{\dot{L}_{x_1}^{r_1} \dot{L}_{x_2}^{r_2}}$.

Proof. This is simply a restatement of Lemma 4.22 combined with the connection to the norm $\|b\|_{\dot{L}_{x_1}^{r_1} \dot{L}_{x_2}^{r_2}}$ given by (3.11). \square

5.6. *Remark.* For $p_1 = p_2 =: p$ and $q_1 = q_2 =: q$, almost the same argument as in Lemma 4.22 gives the following slightly stronger conclusion: Suppose that $[T_1, [b, T_2]] : L^p \rightarrow L^q$, where $p > q$ (this can be replaced with the off-support constants as usual). Then for $1/r = 1/q - 1/p$, any sparse collections \mathcal{R} of rectangles $R = I_1 \times I_2$ in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and coefficients $\sum_{R \in \mathcal{R}} \lambda_R^{r'} |R| \leq 1$, we have

$$\sum_{R \in \mathcal{R}} \lambda_R \operatorname{osc}^{1,1}(b, R) \lesssim \|[T_1, [b, T_2]]\|_{L^p \rightarrow L^q}.$$

The difference is that the coefficients are not of the product form, while above we had $\mathcal{R} = \mathcal{S}_1 \times \mathcal{S}_2 = \{S_1 \times S_2 : S_1 \in \mathcal{S}_1, S_2 \in \mathcal{S}_2\}$. The point of the generalisation is that we can still split the coefficients $\lambda_R = \lambda_R^{r'/p} \lambda_R^{r'/q'}$ and estimate

$$\left\| \sum_{R \in \mathcal{R}} \lambda_R^{r'/p} 1_{R^*} \right\|_{L^p} \lesssim \left(\sum_{R \in \mathcal{R}} (\lambda_R^{r'/p})^p |R| \right)^{1/p} \lesssim 1;$$

in the case of mixed norms, it seems unclear how to perform this splitting of the coefficients. But even with the stronger condition, we do not know how to relate it to $b \in \dot{L}^r(\dot{L}^r)$. This problem is a variant of the question of bi-parameter sparse domination that has attracted some attention. Recently, [2] showed the *failure* of the natural sparse form bound for the strong maximal operator. However, our problem is more like finding a sparse bound for the *identity operator*!

6. SUFFICIENT CONDITIONS FOR COMMUTATOR BOUNDEDNESS

In this section we are given Calderón–Zygmund operators T_i on \mathbb{R}^{d_i} with standard kernels K_i , $i = 1, 2$. We also always assume that $b \in L_{\text{loc}}^s(\mathbb{R}^d)$ for some $s > 1$. Moreover, we let $p_1, p_2, q_1, q_2 \in (1, \infty)$ and use the familiar notation

$$\beta_i := d_i \left(\frac{1}{p_i} - \frac{1}{q_i} \right), \quad \text{if } p_i < q_i; \quad \frac{1}{r_i} := \frac{1}{q_i} - \frac{1}{p_i}, \quad \text{if } p_i > q_i.$$

This fixed data is not repeated in the statements of the results.

We aim to prove

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \lesssim \|b\|_{Y_{p_i, q_i}}$$

preferably with a function space Y_{p_i, q_i} that matches the corresponding necessary condition obtained in Section 5.

Along the way we will need several kernel representations for $\langle [T_1, [b, T_2]]f, g \rangle$ even if f and g do not have disjoint supports. To make these rigorous we recall the notion of maximal truncations of CZOs. For $\epsilon > 0$ we define for all $x_i \in \mathbb{R}^{d_i}$ and $f \in L^p(\mathbb{R}^{d_i})$, $1 \leq p < \infty$, that

$$T_{i, \epsilon} f(x_i) = \int_{|x_i - y_i| > \epsilon} K_i(x_i, y_i) f(y_i) dy_i.$$

We define the maximal truncation of T_i via the formula

$$T_{i, * } f(x_i) = \sup_{\epsilon > 0} |T_{i, \epsilon} f(x_i)|.$$

For the CZO T_i the following Cotlar's inequality is true: for $0 < r < 1$ we have for all $x_i \in \mathbb{R}^{d_i}$ that

$$T_{i, * } f(x_i) \lesssim_r M_r T_i f(x_i) + M f(x_i).$$

Here M is the Hardy–Littlewood maximal function on \mathbb{R}^{d_i} and $M_r g = (M|g|^r)^{1/r}$. The implicit constant in Cotlar’s inequality depends on $\|T_i\|_{L^2 \rightarrow L^2}$, but this is the type of CZO data that we will not track. The consequence of Cotlar’s inequality is that also T_* maps $L^p(\mathbb{R}^{d_i})$ to $L^p(\mathbb{R}^{d_i})$ for all $p \in (1, \infty)$. A standard fact also is that

$$(6.1) \quad T_i f = T_{i,0} f + a_i f,$$

where a_i is a bounded function on \mathbb{R}^{d_i} , $T_{i,0}$ is bounded on $L^p(\mathbb{R}^{d_i})$ for all $p \in (1, \infty)$, and for some sequence $\epsilon_j \rightarrow 0$ we have

$$\int T_{i,0} f \cdot g = \lim_{j \rightarrow \infty} \int T_{i,\epsilon_j} f \cdot g$$

whenever $f \in L^p(\mathbb{R}^{d_i})$ and $g \in L^{p'}(\mathbb{R}^{d_i})$ for some $p \in (1, \infty)$.

For the kernel representations it is convenient to again denote

$$B(x, y) = b(x_1, x_2) - b(x_1, y_2) - b(y_1, x_2) + b(y_1, y_2).$$

6.A. $C^{0,\alpha}(X)$. We begin with our first kernel representation.

6.2. **Proposition.** *If*

$$(x, y) \mapsto B(x, y)K_1(x_1, y_1)K_2(x_2, y_2) \in L^1_{\text{loc}}(\mathbb{R}^{2d}),$$

then for all $f, g \in L^\infty_c(\mathbb{R}^d)$ we have

$$\langle [T_1, [b, T_2]]f, g \rangle = - \iint_{\mathbb{R}^d \times \mathbb{R}^d} B(x, y)K_1(x_1, y_1)K_2(x_2, y_2)f(y)g(x) \, dy \, dx.$$

Proof. We write $T_i h = T_{i,0} h + a_i h$ as in (6.1). Notice that $[T_1, [b, T_2]] = [T_{1,0}, [b, T_{2,0}]]$ and that we have by definition

$$[T_{1,0}, [b, T_{2,0}]]f = T_{1,0}(bT_{2,0}f) + T_{2,0}(bT_{1,0}f) - bT_{1,0}T_{2,0}f - T_{1,0}T_{2,0}(bf).$$

We consider these terms separately. For all $t \in (1, s)$ we have for almost every x_2 that $x_1 \mapsto b(x_1, x_2)T_{2,0}(f(x_1, \cdot))(x_2) \in L^t(\mathbb{R}^{d_1})$. Therefore, we have

$$\langle T_{1,0}(bT_{2,0}f), g \rangle_{\mathbb{R}^{d_1}} = \lim_{j \rightarrow \infty} \langle T_{1,\epsilon_j}(bT_{2,0}f), g \rangle_{\mathbb{R}^{d_1}}.$$

Since $T_{1,*}(bT_{2,0}f)g \in L^1(\mathbb{R}^d)$ we have by dominated convergence theorem and Fubini’s theorem that

$$\langle T_{1,0}(bT_{2,0}f), g \rangle = \lim_{j \rightarrow \infty} \langle T_{1,\epsilon_j}(bT_{2,0}f), g \rangle = \lim_{j \rightarrow \infty} \int_{\mathbb{R}^{d_1}} \langle T_{2,0}f, bT_{1,\epsilon_j}^* g \rangle_{\mathbb{R}^{d_2}}.$$

Similarly as above, we further have

$$\langle T_{1,0}(bT_{2,0}f), g \rangle = \lim_{j \rightarrow \infty} \int_{\mathbb{R}^{d_1}} \lim_{k \rightarrow \infty} \langle T_{2,\rho_k} f, bT_{1,\epsilon_j}^* g \rangle_{\mathbb{R}^{d_2}} = \lim_{j \rightarrow \infty} \lim_{k \rightarrow \infty} \langle T_{1,\epsilon_j}(bT_{2,\rho_k} f), g \rangle.$$

We can similarly write out all the other terms of the commutator $\langle [T_{1,0}, [b, T_{2,0}]]f, g \rangle$. Thus, we have

$$\begin{aligned} \langle [T_{1,0}, [b, T_{2,0}]]f, g \rangle &= \lim_{j \rightarrow \infty} \lim_{k \rightarrow \infty} \langle [T_{1,\epsilon_j}, [b, T_{2,\rho_k}]]f, g \rangle \\ &= - \lim_{j \rightarrow \infty} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} \int_{\substack{y_1: |y_1 - x_1| > \epsilon_j \\ y_2: |y_2 - x_2| > \rho_k}} B(x, y)K_1(x_1, y_1)K_2(x_2, y_2)f(y)g(x) \, dy \, dx. \end{aligned}$$

The proof is finished by a final application of the dominated convergence theorem. \square

We now give the upper bound that matches Proposition 5.1.

6.3. Proposition. *If $p_1 < q_1$ and $p_2 < q_2$, we have*

$$\| [T_1, [b, T_2]] \|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \lesssim \|b\|_{\dot{C}^{0,\beta_1}(\dot{C}^{0,\beta_2})}.$$

Proof. By Proposition 6.2 we have for $b \in \dot{C}^{0,\beta_1}(\dot{C}^{0,\beta_2})$ that

$$|\langle [T_1, [b, T_2]]f, g \rangle| \leq \|b\|_{\dot{C}^{0,\beta_1}(\dot{C}^{0,\beta_2})} \langle I_{1,\beta_1} I_{2,\beta_2} |f|, |g| \rangle,$$

where

$$I_{1,\beta_1} f = \int_{\mathbb{R}^{d_1}} \frac{f(y_1)}{|x_1 - y_1|^{d_1 - \beta_1}} dy_1$$

is the fractional integral hitting on \mathbb{R}^{d_1} and I_{2,β_2} is defined similarly. The proof is completed by using the well-known boundedness of the fractional integrals:

$$I_{i,\beta_i} : L^{p_i} \rightarrow L^{q_i}.$$

Indeed, notice that by Minkowski's integral inequality and the positivity of the operator

$$\|I_{1,\beta_1} |h|\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} \leq \|I_{1,\beta_1} \|h\|_{L_{x_2}^{q_2}} \| \|_{L_{x_1}^{q_1}} \lesssim \| \|h\|_{L_{x_2}^{q_2}} \| \|_{L_{x_1}^{p_1}} = \|h\|_{L_{x_1}^{p_1} L_{x_2}^{q_2}},$$

so that no vector-valued theory is really used. \square

We move on to proving the remaining $\dot{C}^{0,\alpha}(X)$ type upper bounds. Again, a kernel representation is required.

6.4. Proposition. *If $b \in \dot{C}_{x_1}^{0,\alpha}(\text{BMO}_{x_2})$ for some $\alpha > 0$, then for all $f, g \in L_c^\infty(\mathbb{R}^d)$ we have*

$$\langle [T_1, [b, T_2]]f, g \rangle = - \iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}} K_1(x_1, y_1) \langle [b(x_1) - b(y_1), T_2] f(y_1), g(x_1) \rangle_{\mathbb{R}^{d_2}}.$$

Proof. We write

$$\begin{aligned} \langle [T_1, [b, T_2]]f, g \rangle &= \langle [T_{1,0}, [b, T_2]]f, g \rangle = \langle [T_{1,0}, [\tilde{b}, T_2]]f, g \rangle \\ &= \langle T_2[\tilde{b}, T_{1,0}]f, g \rangle - \langle [\tilde{b}, T_{1,0}]T_2 f, g \rangle, \end{aligned}$$

where $\tilde{b} = b - \langle b \rangle_{V,2}$ and $V \subset \mathbb{R}^{d_2}$ is a cube with $\text{spt}_{\mathbb{R}^{d_2}} f \cup \text{spt}_{\mathbb{R}^{d_2}} g \subset V$. We can write

$$\begin{aligned} \langle T_2[\tilde{b}, T_{1,0}]f, g \rangle &= \langle T_{1,0} f, \tilde{b} T_2^* g \rangle - \langle T_{1,0}(\tilde{b} f), T_2^* g \rangle \\ &= \int_{\mathbb{R}^{d_2}} \lim_{j \rightarrow \infty} [\langle T_{1,\varepsilon_j} f, \tilde{b} T_2^* g \rangle_{\mathbb{R}^{d_1}} - \langle T_{1,\varepsilon_j}(\tilde{b} f), T_2^* g \rangle_{\mathbb{R}^{d_1}}] \\ &= \lim_{j \rightarrow \infty} \int_{\mathbb{R}^{d_2}} [\langle T_{1,\varepsilon_j} f, \tilde{b} T_2^* g \rangle_{\mathbb{R}^{d_1}} - \langle T_{1,\varepsilon_j}(\tilde{b} f), T_2^* g \rangle_{\mathbb{R}^{d_1}}], \end{aligned}$$

where in the last step we have used the dominated convergence theorem. Notice that we can now write the above as

$$\lim_{j \rightarrow \infty} \int_{\mathbb{R}^{d_2}} \int_{\mathbb{R}^{d_1}} \int_{y_1: |x_1 - y_1| > \varepsilon_j} K_1(x_1, y_1) (\tilde{b}(x_1, y_2) - \tilde{b}(y_1, y_2)) f(y_1, y_2) T_2^* g(x_1, y_2) dy_1 dx_1 dy_2.$$

Since $b \in \dot{C}_{x_1}^{0,\alpha}(\text{BMO}_{x_2})$, by definition we know that for all $p \in (1, \infty)$ we have

$$y_2 \mapsto \frac{\tilde{b}(x_1, y_2) - \tilde{b}(y_1, y_2)}{|x_1 - y_1|^\alpha} = \frac{b(x_1, y_2) - b(y_1, y_2) - \langle b(x_1, \cdot) - b(y_1, \cdot) \rangle_V}{|x_1 - y_1|^\alpha} \in L^p(V)$$

uniformly for all x_1, y_1 with $x_1 \neq y_1$. Using the above and the boundedness of the fractional integrals we can verify that

$$(x_1, y_1, y_2) \mapsto K_1(x_1, y_1)(\tilde{b}(x_1, y_2) - \tilde{b}(y_1, y_2))f(y_1, y_2)T_2^*g(x_1, y_2) \in L^1(\mathbb{R}^{2d_1+d_2}).$$

Therefore, we have

$$\begin{aligned} & \langle T_2[\tilde{b}, T_{1,0}]f, g \rangle \\ &= \int_{\mathbb{R}^{2d_1}} \int_{\mathbb{R}^{d_2}} K_1(x_1, y_1)T_2((\tilde{b}(x_1, \cdot) - \tilde{b}(y_1, \cdot))f(y_1, \cdot))(y_2)g(x_1, y_2) dy_2 dy_1 dx_1. \end{aligned}$$

We complete the proof by noting that similarly $\langle [\tilde{b}, T_{1,0}]T_2f, g \rangle$ equals

$$\int_{\mathbb{R}^{2d_1}} \int_{\mathbb{R}^{d_2}} K_1(x_1, y_1)(\tilde{b}(x_1, y_2) - \tilde{b}(y_1, y_2))T_2f(y_1, y_2)g(x_1, y_2) dy_2 dy_1 dx_1.$$

□

We obtain the following counterpart of Proposition 5.2.

6.5. Proposition. *If $p_1 < q_1$ and $p_2 = q_2$, then*

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{p_2}} \lesssim \|b\|_{\dot{C}_{x_1}^{0, \beta_1}(\text{BMO}_{x_2})}.$$

If $p_1 = q_1$ and $p_2 < q_2$, then

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{p_1} L_{x_2}^{q_2}} \lesssim \|b\|_{\dot{C}_{x_2}^{0, \beta_2}(\text{BMO}_{x_1})}.$$

Proof. Let first $p_1 < q_1$ and $p_2 = q_2$. For $f, g \in L_c^\infty$ we have by Proposition 6.4 that

$$\begin{aligned} & |\langle [T_1, [b, T_2]]f, g \rangle| \\ &= \left| \iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}} |x_1 - y_1|^{\beta_1} K_1(x_1, y_1) \left\langle \left[\frac{b(x_1) - b(y_1)}{|x_1 - y_1|^{\beta_1}}, T_2 \right] f(y_1), g(x_1) \right\rangle_{\mathbb{R}^{d_2}} \right| \\ &\lesssim \iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}} |x_1 - y_1|^{\beta_1 - d_1} \left\| \frac{b(x_1) - b(y_1)}{|x_1 - y_1|^{\beta_1}} \right\|_{\text{BMO}_{x_2}} \|f(y_1)\|_{L_{x_2}^{p_2}} \|g(x_1)\|_{L_{x_2}^{p'_2}} \\ &\lesssim \|b\|_{\dot{C}_{x_1}^{0, \beta_1}(\text{BMO}_{x_2})} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \|g\|_{L_{x_1}^{q'_1} L_{x_2}^{p'_2}}, \end{aligned}$$

where we have used the well-known boundedness of one parameter commutators and the fractional integrals.

Next, let $p_1 = q_1$ and $p_2 < q_2$. Similarly as in Proposition 6.4, for $f, g \in L_c^\infty$ we may write

$$\langle [T_1, [b, T_2]]f, g \rangle = - \iint_{\mathbb{R}^{d_2} \times \mathbb{R}^{d_2}} K_2(x_2, y_2) \langle [b(x_2) - b(y_2), T_1] f(y_2), g(x_2) \rangle_{\mathbb{R}^{d_1}}.$$

By the sparse domination of commutators – see [34] – we have for all $r \in (1, \infty)$ that

$$\begin{aligned} & |\langle [b(x_2) - b(y_2), T_1] f(y_2), g(x_2) \rangle_{\mathbb{R}^{d_1}}| \\ &\lesssim \|b(\cdot, x_2) - b(\cdot, y_2)\|_{\text{BMO}(\mathbb{R}^{d_1})} \int_{\mathbb{R}^{d_1}} M_r^1 f(\cdot, y_2) M_r^1 g(\cdot, x_2). \end{aligned}$$

Choosing $1 < r < \min\{p_1, p_2, p'_1, q'_2\}$ and using the above we have that $|\langle [T_1, [b, T_2]]f, g \rangle|$ can be dominated by

$$\|b\|_{\dot{C}_{x_2}^{0, \beta_2}(\text{BMO}_{x_1})} \iint_{\mathbb{R}^{d_2} \times \mathbb{R}^{d_2}} |K_2(x_2, y_2)| |x_2 - y_2|^{\beta_2} \int_{\mathbb{R}^{d_1}} M_r^1 f(x_1, y_2) M_r^1 g(x_1, x_2)$$

$$\leq \|b\|_{\dot{C}_{x_2}^{0,\beta_2}(\text{BMO}_{x_1})} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \|g\|_{L_{x_1}^{p'_1} L_{x_2}^{q'_2}},$$

where we have used the boundedness of fractional integrals and mixed norm estimates of M_r^1 . This completes the proof. \square

We move on to the upper bounds that are related to the Hölder space estimates of Proposition 5.4.

6.6. Proposition. *If $p_1 < q_1$ and $p_2 > q_2$ then*

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \lesssim \|b\|_{\dot{C}_{x_1}^{0,\beta_1}(\dot{L}_{x_2}^{r_2})}.$$

If $p_1 > q_1$ and $p_2 < q_2$ then

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \lesssim \|b\|_{\dot{L}_{x_1}^{r_1}(\dot{C}_{x_2}^{0,\beta_2})}.$$

Proof. Let $p_1 < q_1$ and $p_2 > q_2$. Similarly as above for $f, g \in L_c^\infty$ we may write

$$\langle [T_1, [b, T_2]]f, g \rangle = - \iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}} K_1(x_1, y_1) \langle [b(x_1) - b(y_1), T_2]f(y_1), g(x_1) \rangle_{\mathbb{R}^{d_2}}.$$

This gives that

$$\begin{aligned} & |\langle [T_1, [b, T_2]]f, g \rangle| \\ &= \left| \iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}} |x_1 - y_1|^{\beta_1} K_1(x_1, y_1) \left\langle \left[\frac{b(x_1) - b(y_1)}{|x_1 - y_1|^{\beta_1}}, T_2 \right] f(y_1), g(x_1) \right\rangle_{\mathbb{R}^{d_2}} \right| \\ &\lesssim \iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}} |x_1 - y_1|^{\beta_1 - d_1} \left\| \frac{b(x_1) - b(y_1)}{|x_1 - y_1|^{\beta_1}} \right\|_{\dot{L}_{x_2}^{r_2}} \|f(y_1)\|_{L_{x_2}^{p_2}} \|g(x_1)\|_{L_{x_2}^{q'_2}} \\ &\lesssim \|b\|_{\dot{C}_{x_1}^{0,\beta_1}(\dot{L}_{x_2}^{r_2})} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}} \|g\|_{L_{x_1}^{q'_1} L_{x_2}^{q'_2}}, \end{aligned}$$

where we have used the elementary inequality (6.9) and the boundedness of fractional integrals.

Next, let $p_1 > q_1$ and $p_2 < q_2$. We simply estimate

$$\begin{aligned} \|[T_1, [b, T_2]]f\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} &= \|[T_1, [b - c_2, T_2]]f\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} \\ &\leq \|T_1[b - c_2, T_2]f\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} + \|[b - c_2, T_2]T_1f\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}}, \end{aligned}$$

use the mixed norm estimates of T_1 , and the estimate

$$\|[b - c_2, T_2]f\|_{L_{x_1}^{q_1} L_{x_2}^{q_2}} \lesssim \left\| \|b - c_2\|_{\dot{C}_{x_2}^{0,\beta_2}} \|f\|_{L_{x_2}^{p_2}} \right\|_{L_{x_1}^{q_1}} \leq \|b - c_2\|_{L_{x_1}^{r_1}(\dot{C}_{x_2}^{0,\beta_2})} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}}.$$

To end the proof we take the infimum over suitable $c_2(x) = c_2(x_2)$. \square

6.B. BMO and \dot{L}^r . Proposition 5.4 recorded the oscillatory lower bounds related to the space $\text{BMO}(\dot{L}^r)$. Here are the related upper bounds.

6.7. Proposition. *If $p_1 = q_1$ and $p_2 > q_2$, then*

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{p_1} L_{x_2}^{q_2}} \lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})}.$$

If $p_1 > q_1$ and $p_2 = q_2$, then

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{p_2}} \lesssim \|b\|_{\dot{L}_{x_1}^{r_1}(\text{BMO}_{x_2})}.$$

Proof. The first estimate, where $p_1 = q_1$ and $p_2 > q_2$, requires some vector-valued theory of commutators. We notice that

$$(6.8) \quad ([T_1, [b, T_2]]f)^\sharp \lesssim_\varepsilon \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})} (M_{1+\varepsilon} \|T_1 f\|_{L_{x_2}^{p_2}} + M_{1+\varepsilon} \|f\|_{L_{x_2}^{p_2}}),$$

where M is the maximal function, $M_s g = (M|g|^s)^{1/s}$ and the sharp maximal function is defined using the $L_{x_2}^{q_2}$ norm as follows

$$g^\sharp(x_1) := \sup_{I_1} \int_{I_1} \|g - \langle g \rangle_{I_1}\|_{L_{x_2}^{q_2}}.$$

Here the supremum is over the cubes $I_1 \subset \mathbb{R}^{d_1}$ centered at x_1 . The proof of this is essentially a vector-valued version of a known pointwise bound for the sharp maximal function of $[b, T_1]$, but we need to keep the operator T_2 around to exploit the elementary estimate

$$(6.9) \quad \|[b, T_2]f\|_{L_{x_2}^{q_2}} \lesssim \|b\|_{\dot{L}_{x_2}^{r_2}} \|f\|_{L_{x_2}^{p_2}}.$$

We give the full details of (6.8) for the convenience of the reader. Below we repeatedly use the well-known vector-valued boundedness of T_1 and (6.9) without explicit mention.

Fix a cube I_1 containing the implicit variable. Let $f_1 = f1_{I_1^*}$ and $f_2 = f - f_1$, where $I_1^* = 5\sqrt{d_1}I_1$. Let

$$c = \langle T_1([b - a, T_2]f_2) \rangle_{I_1, 1}, \quad a = \langle b \rangle_{I_1^*, 1}.$$

We can now write

$$[T_1, [b, T_2]]f = [T_1, [b - a, T_2]]f = -[b - a, T_2]T_1 f + T_1([b - a, T_2]f_1) + T_1([b - a, T_2]f_2).$$

Using this we split

$$\begin{aligned} \int_{I_1} \|[T_1, [b, T_2]]f - c\|_{L_{x_2}^{q_2}} &\leq \int_{I_1} \|[b - a, T_2]T_1 f\|_{L_{x_2}^{q_2}} + \int_{I_1} \|T_1([b - a, T_2]f_1)\|_{L_{x_2}^{q_2}} \\ &\quad + \int_{I_1} \|T_1([b - a, T_2]f_2) - c\|_{L_{x_2}^{q_2}} \\ &=: L_1 + L_2 + L_3. \end{aligned}$$

We have

$$L_1 \lesssim \left(\int_{I_1^*} \|b - a\|_{\dot{L}_{x_2}^{r_2}}^{(1+\varepsilon)'} \right)^{\frac{1}{(1+\varepsilon)'}} \left(\int_{I_1} \|T_1 f\|_{L_{x_2}^{p_2}}^{1+\varepsilon} \right)^{\frac{1}{1+\varepsilon}} \lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})} M_{1+\varepsilon} (\|T_1 f\|_{L_{x_2}^{p_2}}),$$

where in the last inequality we have used John-Nirenberg inequality of $\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})$, which has the same proof as the usual vector-valued John-Nirenberg inequality (the dot in $\dot{L}_{x_2}^{r_2}$ does not matter).

We move to L_2 . This time we have

$$\begin{aligned} L_2 &\lesssim |I_1|^{-\frac{1}{1+\frac{\varepsilon}{2}}} \|[b - a, T_2]f_1\|_{L_{x_1}^{1+\frac{\varepsilon}{2}}(L_{x_2}^{q_2})} \\ &\lesssim \left(\int_{I_1^*} \|b - a\|_{\dot{L}_{x_2}^{r_2}}^{1+\frac{\varepsilon}{2}} \|f\|_{L_{x_2}^{p_2}}^{1+\frac{\varepsilon}{2}} \right)^{\frac{1}{1+\frac{\varepsilon}{2}}} \lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})} M_{1+\varepsilon} \|f\|_{L_{x_2}^{p_2}}. \end{aligned}$$

We used John-Nirenberg again in the last step.

It remains to estimate L_3 . Since $I_1 \cap \text{spt } f_2 = \emptyset$, using the kernel representation of T_1 and the regularity of the kernel we have

$$\begin{aligned} L_3 &\lesssim \sum_{j=0}^{\infty} 2^{-j\alpha_1} \int_{2^{j+1}I_1^*} \|[b - a, T_2]f\|_{L_{x_2}^{q_2}} \lesssim \sum_{j=0}^{\infty} 2^{-j\alpha_1} \int_{2^{j+1}I_1^*} \|b - a\|_{\dot{L}_{x_2}^{r_2}} \|f\|_{L_{x_2}^{p_2}} \\ &\leq \sum_{j=0}^{\infty} 2^{-j\alpha_1} \int_{2^{j+1}I_1^*} \|b - \langle b \rangle_{2^{j+1}I_1^*, 1}\|_{\dot{L}_{x_2}^{r_2}} \|f\|_{L_{x_2}^{p_2}} \\ &\quad + \sum_{j=0}^{\infty} 2^{-j\alpha_1} \int_{2^{j+1}I_1^*} \|\langle b \rangle_{I_1^*, 1} - \langle b \rangle_{2^{j+1}I_1^*, 1}\|_{\dot{L}_{x_2}^{r_2}} \|f\|_{L_{x_2}^{p_2}}. \end{aligned}$$

Using John-Nirenberg again, the first sum can be bounded by $\|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})} M_{1+\varepsilon} \|f\|_{L_{x_2}^{p_2}}$. Moreover, since

$$\|\langle b \rangle_{I_1^*, 1} - \langle b \rangle_{2^{j+1}I_1^*, 1}\|_{\dot{L}_{x_2}^{r_2}} \leq \sum_{i=0}^j \|\langle b \rangle_{2^{i+1}I_1^*, 1} - \langle b \rangle_{2^i I_1^*, 1}\|_{\dot{L}_{x_2}^{r_2}} \lesssim j \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})},$$

the desired estimate follows for the second term as well. The inequality (6.8) follows.

Next, if ε is so small that $p_1/(1 + \varepsilon) > 1$, we have by the standard Fefferman-Stein inequality that

$$\begin{aligned} \|[T_1, [b, T_2]]f\|_{L_{x_1}^{p_1} L_{x_2}^{q_2}}^{p_1} &= \int_{\mathbb{R}^{d_1}} \|[T_1, [b, T_2]]f\|_{L_{x_2}^{q_2}}^{p_1} \\ &\lesssim \int_{\mathbb{R}^{d_1}} [([T_1, [b, T_2]]f)^\sharp]^{p_1} \\ &\lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})}^{p_1} \left(\|M_{1+\varepsilon} T_1 f\|_{L_{x_2}^{p_2}}^{p_1} \|L_{x_1}^{p_1}\|^{p_1} + \|M_{1+\varepsilon} f\|_{L_{x_2}^{p_2}}^{p_1} \|L_{x_1}^{p_1}\|^{p_1} \right) \\ &\lesssim \|b\|_{\text{BMO}_{x_1}(\dot{L}_{x_2}^{r_2})}^{p_1} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}}^{p_1}. \end{aligned}$$

We are done with the case $p_1 = q_1$ and $p_2 > q_2$. See Section 7 for more about vector-valued estimates of commutators.

The case $p_1 > q_1$ and $p_2 = q_2$ is the following elementary deduction. Estimate

$$\begin{aligned} \|[T_1, [b, T_2]]f\|_{L_{x_1}^{q_1} L_{x_2}^{p_2}} &= \|[T_1, [b - c_2, T_2]]f\|_{L_{x_1}^{q_1} L_{x_2}^{p_2}} \\ &\leq \|T_1[b - c_2, T_2]f\|_{L_{x_1}^{q_1} L_{x_2}^{p_2}} + \|[b - c_2, T_2]T_1 f\|_{L_{x_1}^{q_1} L_{x_2}^{p_2}}, \end{aligned}$$

use the mixed norm estimates of T_1 and

$$\|[b - c_2, T_2]f\|_{L_{x_1}^{q_1} L_{x_2}^{p_2}} \lesssim \left\| \|b - c_2\|_{\text{BMO}_{x_2}} \|f\|_{L_{x_2}^{p_2}} \right\|_{L_{x_1}^{q_1}} \leq \|b - c_2\|_{L_{x_1}^{r_1}(\text{BMO}_{x_2})} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}}.$$

To end the proof we take the infimum over suitable $c_2(x) = c_2(x_2)$. \square

The oscillatory lower bound related to $\dot{L}^{r_1} \dot{L}^{r_2}$ was recorded in Proposition 5.5. The upper bound is completely elementary and does not utilize any cancellation of the bi-commutator.

6.10. Proposition. *If $p_1 > q_1$ and $p_2 > q_2$, then*

$$\|[T_1, [b, T_2]]\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} \rightarrow L_{x_1}^{q_1} L_{x_2}^{q_2}} \lesssim \|b\|_{\dot{L}_{x_1}^{r_1}(\dot{L}_{x_2}^{r_2})}.$$

Proof. Simply use Hölder's inequality and the mixed-norm estimates of T_1 and T_2 . \square

7. VECTOR-VALUED ESTIMATES FOR COMMUTATORS

In the proof of Proposition 6.7 we used some vector-valued theory of commutators. Although the proof of Proposition 6.7 was completely self-contained, in this section we more systematically prove, and partly recall, such vector-valued estimates. This topic is of independent interest.

7.A. Vector-valued estimates for $[b, T]$. In the one-parameter setting $[b, T]$, where T is a CZO in \mathbb{R}^d , the vector-valued estimates are known and quite easy, but perhaps not so well-known in the community. They are simple as there is a certain pointwise estimate for the sharp maximal function of $[b, T]$, which has a straightforward vector-valued extension. A reference is [48]. This idea was already used in the proof of Proposition 6.7. We now recall how this can be done in a general situation.

First, a few definitions. For an extensive treatment of Banach space theory see the books [26, 27] by Hytönen, van Neerven, Veraar and Weis.

Given a Banach space X with norm $|\cdot|_X$ we let $L^p(\mathbb{R}^d; X)$, $p \in (0, \infty]$, be the Bochner space of X -valued functions with $\int_{\mathbb{R}^d} |f(x)|_X^p dx < \infty$. For $f : \mathbb{R}^d \rightarrow X$, we also define

$$f^\#(x) := \sup_{Q=Q(x,r), r>0} \frac{1}{|Q|} \int_Q |f - \langle f \rangle_Q|_X,$$

where the supremum is over the cubes $Q(x, r)$ centred at x . We have

$$f^\#(x) \sim \sup_{Q=Q(x,r), r>0} \inf_{e \in X} \frac{1}{|Q|} \int_Q |f - e|_X.$$

The UMD property is a necessary and sufficient condition for the boundedness of various one-parameter singular integrals on $L^p(\mathbb{R}^d; X) = L^p(X)$, see [26, Sec. 5.2.c and the Notes to Sec. 5.2]. We essentially only need to blackbox this fact, but below is the definition.

7.1. Definition. A Banach space X is said to be a UMD space, where UMD stands for unconditional martingale differences, if for all $p \in (1, \infty)$, all X -valued L^p -martingale difference sequences $(d_j)_{j=1}^k$ and all choices of signs $\varepsilon_j \in \{-1, 1\}$ we have

$$(7.2) \quad \left\| \sum_{j=1}^k \varepsilon_j d_j \right\|_{L^p(X)} \lesssim \left\| \sum_{j=1}^k d_j \right\|_{L^p(X)}.$$

The $L^p(X)$ -norm is with respect to the measure space where the martingale differences are defined.

A standard property of UMD spaces is that if (7.2) holds for one $p_0 \in (1, \infty)$ it holds for all $p \in (1, \infty)$ [26, Theorem 4.2.7]. Moreover, if X is UMD then so is the dual space X^* [26, Prop. 4.2.17].

Suppose now that X_1 is a Banach space and X_2, X_3 are UMD spaces such that there exists a bilinear map $X_1 \times X_2 \rightarrow X_3$, which we denote by

$$(x_1, x_2) \rightarrow x_1 x_2,$$

so that

$$|x_1 x_2|_{X_3} \leq |x_1|_{X_1} |x_2|_{X_2}.$$

Let f take values in X_2 and b take values in X_1 . For a CZO T on \mathbb{R}^d a standard estimate as in the proof of Proposition 6.7 gives that for all $\varepsilon > 0$ we have

$$(7.3) \quad ([b, T]f)^\sharp \lesssim_\varepsilon \|b\|_{\text{BMO}(X_1)} (M_{1+\varepsilon}(|Tf|_{X_2}) + M_{1+\varepsilon}|f|_{X_2}),$$

where M is the maximal function and $M_s g = (M|g|^s)^{1/s}$. The sharp maximal function is defined using the norm of X_3 . To get this estimate, already the boundedness $T: L^q(X_3) \rightarrow L^q(X_3)$, $q \in (1, \infty)$, is required – this is true as X_3 is UMD. The vector-valued estimate

$$(7.4) \quad \| [b, T]f \|_{L^p(X_3)} \lesssim \|b\|_{\text{BMO}(X_1)} \|f\|_{L^p(X_2)}, \quad p \in (1, \infty),$$

follows from (7.3) using the Fefferman-Stein inequality as in the proof of Proposition 6.7. Here the UMD property of X_2 is needed via the boundedness $T: L^p(X_2) \rightarrow L^p(X_2)$.

Notice that the proof of Proposition 6.7 was almost just this vector-valued estimate – we just needed to work a bit more to get the norm $\|b\|_{\text{BMO}_{x_1}(L_{x_2}^{r_2})}$ instead of $\|b\|_{\text{BMO}_{x_1}(L_{x_2}^{r_2})}$.

Alternatively, it is also possible to directly prove a vector-valued version of the sparse domination of commutators [34]:

$$\begin{aligned} |\langle [b, T]f, g \rangle| &\lesssim \sum_{Q \in \mathcal{S}} \langle |b - \langle b \rangle_Q|_{X_1} |f|_{X_2} \rangle_Q \langle |g|_{X_3^*} \rangle_Q |Q| \\ &\quad + \sum_{Q \in \mathcal{S}} \langle |f|_{X_2} \rangle_Q \langle |b - \langle b \rangle_Q|_{X_1} |g|_{X_3^*} \rangle_Q |Q|. \end{aligned}$$

7.B. Vector-valued estimates for bi-commutators. We will prove the bi-parameter analogue of the vector-valued estimate (7.4). There is no equally cheap way for this as in the above one-parameter case. However, we want to show that it can still be done with different methods.

We need more definitions and tools.

Dyadic notation. Let \mathcal{D} be some fixed dyadic lattice in \mathbb{R}^d . For a fixed $Q \in \mathcal{D}$ and $f \in L_{\text{loc}}^1(X)$ we define as follows.

- If $k \in \mathbb{Z}$, $k \geq 0$, then $Q^{(k)}$ denotes the unique cube $S \in \mathcal{D}$ for which $Q \subset S$ and $\ell(Q) = 2^{-k} \ell(S)$.
- The dyadic children of Q are denoted by $\text{ch}(Q) = \{Q' \in \mathcal{D}: (Q')^{(1)} = Q\}$.
- The average operator is $E_Q f = \langle f \rangle_Q \mathbf{1}_Q$.
- The martingale difference $\Delta_Q f$ is defined by $\Delta_Q f = \sum_{Q' \in \text{ch}(Q)} E_{Q'} f - E_Q f$.
- For $k \in \mathbb{Z}$, $k \geq 0$, we define the martingale difference block

$$\Delta_{Q,k} f = \sum_{\substack{S \in \mathcal{D} \\ S^{(k)} = Q}} \Delta_S f.$$

Haar functions are used to decompose martingale differences $\Delta_Q f$ in terms of rank-one operators. For an interval $I \subset \mathbb{R}$, we denote by I_l and I_r the left and right halves of the interval I , respectively. We define $h_I^0 = |I|^{-1/2} \mathbf{1}_I$ and $h_I^1 = |I|^{-1/2} (\mathbf{1}_{I_l} - \mathbf{1}_{I_r})$. Let now $Q = I_1 \times \cdots \times I_d \in \mathcal{D}$, and define the Haar function h_Q^η , $\eta = (\eta_1, \dots, \eta_d) \in \{0, 1\}^d$, via

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

If $\eta \neq 0$, the Haar function is cancellative: $\int h_Q^\eta = 0$. We may write $\Delta_Q f = \sum_{\eta \neq 0} \langle f, h_Q^\eta \rangle h_Q^\eta$, where $\langle f, h_Q^\eta \rangle = \int f h_Q^\eta$. We exploit notation by suppressing the presence of η , and simply write h_Q for some $h_Q^\eta, \eta \neq 0$. Similarly, we write $\Delta_Q f = \langle f, h_Q \rangle h_Q$.

When performing bi-parameter analysis and thinking of \mathbb{R}^d as a product space $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, we denote a dyadic grid in \mathbb{R}^{d_i} by \mathcal{D}^i and denote the related dyadic rectangles in \mathbb{R}^d by $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$. If $R = I_1 \times I_2 \in \mathcal{D}$ we set $h_R = h_{I_1} \otimes h_{I_2}$. We also define $\Delta_{I_1}^1 f(x) := \Delta_{I_1}(f(\cdot, x_2))(x_1)$ and define $\Delta_{I_2}^2 f$ analogously. Then we set $\Delta_R f = \Delta_{I_1}^1 \Delta_{I_2}^2 f = \Delta_{I_2}^2 \Delta_{I_1}^1 f$. We record that $\Delta_{I_1}^1 f = h_{I_1} \otimes \langle f, h_{I_1} \rangle_1$, $\Delta_{I_2}^2 f = \langle f, h_{I_2} \rangle_2 \otimes h_{I_2}$ and $\Delta_R f = \langle f, h_R \rangle h_R$. Bi-parameter martingale blocks may also be defined in the natural way

$$\Delta_R^{j_1, j_2} f = \sum_{J_1: J_1^{(j_1)} = I_1} \sum_{J_2: J_2^{(j_2)} = I_2} \Delta_{J_1 \times J_2} f.$$

More about Banach spaces. In addition to UMD we will need the ‘‘property (α) ’’. Bi-parameter extensions require this – see [27, Sec. 8.3.e]. Suppose that for all N , all scalars $a_{i,j}$ and all $e_{i,j} \in X, 1 \leq i, j \leq N$, there holds

$$\left(\mathbb{E} \mathbb{E}' \left| \sum_{1 \leq i, j \leq N} \varepsilon_i \varepsilon'_j a_{i,j} e_{i,j} \right|_X^2 \right)^{1/2} \lesssim \max_{i,j} |a_{i,j}| \left(\mathbb{E} \mathbb{E}' \left| \sum_{1 \leq i, j \leq N} \varepsilon_i \varepsilon'_j e_{i,j} \right|_X^2 \right)^{1/2}.$$

If this holds, the Banach space X is said to satisfy the property (α) of Pisier. Here $\{\varepsilon_k\}_k$ is a *collection of independent random signs* – that is, the following holds. We have $\varepsilon_k: \mathcal{M} \rightarrow \{-1, 1\}$, where (\mathcal{M}, ρ) is a probability space, the collection $\{\varepsilon_k\}_k$ is independent and $\rho(\{\varepsilon_k = 1\}) = \rho(\{\varepsilon_k = -1\}) = 1/2$.

Instead of dealing with UMD spaces with property (α) we will be working with the formally stronger (although for concrete examples essentially the same) assumption that X is a UMD function lattice. For proofs and background on function lattices see [1, 25, 39, 40]. A normed space X is a Banach function lattice if the following four conditions hold. Let $(\Omega, \mathcal{A}, \mu)$ be a σ -finite measure space.

- (1) Every $e \in X$ is a measurable function $e: \Omega \rightarrow \mathbb{R}$ (an equivalence class).
- (2) If $e: \Omega \rightarrow \mathbb{R}$ is measurable, $u \in X$ and $|e(\omega)| \leq |u(\omega)|$ for μ -a.e. $\omega \in \Omega$, then $e \in X$ and $|e|_X \leq |u|_X$.
- (3) There is an element $e \in X$ so that $e > 0$ (i.e. $e(\omega) > 0$ for μ -a.e. $\omega \in \Omega$).
- (4) If e_i, e are non-negative, $e_i \in X, e_i \leq e_{i+1}, e_i(\omega) \rightarrow e(\omega)$ for μ -a.e. $\omega \in \Omega$ and $\sup_i |e_i|_X < \infty$, we have $e \in X$ and $|e|_X \rightarrow |e|_X = \sup_i |e_i|_X$.

Such a space X is automatically a Banach space. For a measurable $u: \Omega \rightarrow \mathbb{R}$ with $eu \in L^1(\mu)$ for all $e \in X$, we define

$$\Lambda_u: X \rightarrow \mathbb{R}, e_u(e) = \int_{\Omega} e(\omega) u(\omega) d\mu(\omega).$$

In this case $\Lambda_u \in X^*$ (the dual space of X). We set $X' \subset X^*$ to consist of those elements of X^* that have the form Λ_u for some u like above and make the obvious identification. The Banach function lattice X' is called the Köthe dual of X .

We mostly work with UMD Banach function lattices, where $X' = X^*$ is automatic. Indeed, UMD spaces are always reflexive, and in reflexive Banach lattices $X' = X^*$ holds.

When $X' = X^*$ is not automatic, we prefer to tacitly assume it for clarity. A UMD function lattice X also automatically satisfies the property (α) of Pisier – this explains why this assumption is not explicitly needed.

Next, we explain some maximal function estimates by Bourgain [3] and Rubio de Francia [47] that are valid in UMD function lattices (see also [17]). We also explain some standard square function estimates, see e.g. the recent paper [25].

Let X be a UMD function lattice, $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, where \mathcal{D}^i is a dyadic grid in \mathbb{R}^{d_i} . Given $f: \mathbb{R}^d \rightarrow X$ and $g: \mathbb{R}^{d_1} \rightarrow X$ we define the dyadic (lattice) maximal functions

$$M_{\mathcal{D}^1} g(x_1, \omega) := \sup_{I \in \mathcal{D}^1} \frac{1_I(x_1)}{|I|} \int_I |g(x_1, \omega)| dx_1 \text{ and } M_{\mathcal{D}} f(x, \omega) := \sup_{R \in \mathcal{D}} \frac{1_R(x)}{|R|} \int_R |f(x, \omega)| dx.$$

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

Similarly, define the square functions

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

Notice that $|x| \in X$ and $|x|^2 \in X$ are defined in the natural pointwise sense, as X is a function lattice. Define also

$$S_{\mathcal{D}^1, M_{\mathcal{D}^2}} f = \left(\sum_{I \in \mathcal{D}^1} \frac{1_I}{|I|} \otimes [M_{\mathcal{D}^2} \langle f, h_I \rangle_1]^2 \right)^{1/2}$$

and define $S_{\mathcal{D}^2, M_{\mathcal{D}^1}} f$ analogously. We are ready to state the following standard estimates that are at least recorded in [25].

7.5. Lemma. *Let $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, where \mathcal{D}^i is a dyadic grid in \mathbb{R}^{d_i} , and let X be a UMD function lattice. Then we have for all $p_1, p_2 \in (1, \infty)$ that*

$$\|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)} \sim \|S_{\mathcal{D}} f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)} \sim \|S_{\mathcal{D}^1}^1 f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)} \sim \|S_{\mathcal{D}^2}^2 f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)}.$$

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

$$\left\| \left(\sum_j |M f_j|^s \right)^{1/s} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)} \lesssim \left\| \left(\sum_j |f_j|^s \right)^{1/s} \right\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)}.$$

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

$$\|S_{\mathcal{D}^1, M_{\mathcal{D}^2}} f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)} + \|S_{\mathcal{D}^2, M_{\mathcal{D}^1}} f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)} \lesssim \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X)}.$$

Product BMO and vector-valued H^1 – BMO estimates. Let again X be a Banach function lattice, $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, where \mathcal{D}^i is a dyadic grid in \mathbb{R}^{d_i} . A locally integrable function $b: \mathbb{R}^d \rightarrow X$ belongs to the dyadic X -valued product BMO space $\text{BMO}_{\mathcal{D}, \text{prod}}(p, X)$, $0 < p < \infty$, if

$$\sup_{\Omega} \frac{1}{|\Omega|^{1/p}} \left\| \left(\sum_{\substack{R \in \mathcal{D} \\ R \subset \Omega}} |\langle b, h_R \rangle|^2 \frac{1_R}{|R|} \right)^{1/2} \right\|_{L^p(X)} < \infty,$$

where the supremum is taken over all open subsets $\Omega \subset \mathbb{R}^d$ of finite measure. We define $\text{BMO}_{\mathcal{D}, \text{prod}}(X) := \text{BMO}_{\mathcal{D}, \text{prod}}(2, X)$. With this definition we will not require a Banach-valued bi-parameter John–Nirenberg inequality for which we do not have a convenient

reference. In the scalar-valued case it is well-known that all these norms are equivalent. The product BMO is the supremum over all dyadic grids of the dyadic product BMO norms.

7.6. *Remark.* The following explanation gives one way to arrive at the definition of product BMO in our generality. If X is UMD and $p \in (1, \infty)$ then

$$\left(\int_{J_i} |b - \langle b \rangle_{J_i}|_X^p \right)^{1/p} = \left\| \sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} \langle b, h_{I_i} \rangle h_{I_i} \right\|_{L^p(X)} \sim \mathbb{E} \left\| \sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} \varepsilon_{I_i} \langle b, h_{I_i} \rangle \frac{1_{I_i}}{|I_i|^{1/2}} \right\|_{L^p(X)}.$$

If X is a UMD function lattice we further have

$$\mathbb{E} \left\| \sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} \varepsilon_{I_i} \langle b, h_{I_i} \rangle \frac{1_{I_i}}{|I_i|^{1/2}} \right\|_{L^p(X)} \sim \left\| \left(\sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} |\langle b, h_{I_i} \rangle|^2 \frac{1_{I_i}}{|I_i|} \right)^{1/2} \right\|_{L^p(X)}.$$

The last similarity can be seen as follows. The Kahane–Khintchine inequality (see [26, Theorem 6.2.4]) says if X is a Banach space, we have for all $x_1, \dots, x_M \in X$ and $p, q \in (0, \infty)$ that

$$\left(\mathbb{E} \left| \sum_{m=1}^M \varepsilon_m x_m \right|_X^p \right)^{1/p} \sim \left(\mathbb{E} \left| \sum_{m=1}^M \varepsilon_m x_m \right|_X^q \right)^{1/q}.$$

Applying this twice we see that

$$\mathbb{E} \left\| \sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} \varepsilon_{I_i} \langle b, h_{I_i} \rangle \frac{1_{I_i}}{|I_i|^{1/2}} \right\|_{L^p(X)} \sim \left\| \left(\mathbb{E} \left| \sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} \varepsilon_{I_i} \langle b, h_{I_i} \rangle \frac{1_{I_i}}{|I_i|^{1/2}} \right|_X^2 \right)^{1/2} \right\|_{L^p}.$$

It remains to use the equivalence

$$\left(\mathbb{E} \left| \sum_j \varepsilon_j e_j \right|_X^2 \right)^{1/2} \sim \left| \left(\sum_j |e_j|^2 \right)^{1/2} \right|_X$$

for all $e_j \in X$. This theorem of Khintchine–Maurey holds in all Banach lattices with finite cotype – in particular, in UMD function lattices. For a proof see Theorem 7.2.13 in [27].

This explains that in one-parameter the dyadic BMO(X) norm (if X is UMD) also has the equivalent form

$$\sup_{\substack{J_i \in \mathcal{D}^i \\ I_i \subset J_i}} \left\| \left(\sum_{\substack{I_i \in \mathcal{D}^i \\ I_i \subset J_i}} |\langle b, h_{I_i} \rangle|^2 \frac{1_{I_i}}{|I_i|} \right)^{1/2} \right\|_{L^p(X)}.$$

The product BMO is simply the bi-parameter analogue of this form, where the supremum is taken over all open sets. The rectangular BMO is the analogue of this, where the supremum is over dyadic rectangles. This is because

$$\sum_{\substack{P \in \mathcal{D} \\ P \subset R = I_1 \times I_2}} \langle b, h_P \rangle h_P = (b - \langle b \rangle_{I_{1,1}} - \langle b \rangle_{I_{2,2}} + \langle b \rangle_R) 1_R.$$

Let $b: \mathbb{R}^d \rightarrow X$ and $f: \mathbb{R}^d \rightarrow X'$, where X is a Banach function lattice and X' is the Köthe dual of X . Denote $\langle b, f \rangle = \int_{\mathbb{R}^d} \{b(x), f(x)\}_X dx$, where $\{\cdot, \cdot\}_X$ denotes the duality pairing between X and X' , in other words, we have $\{e, e'\}_X = \int_{\Omega} e(\omega) e'(\omega) d\mu(\omega)$ for $e \in X$ and $e' \in X'$.

The following H^1 -BMO duality in the lattice case is convenient. The proof is similar to the scalar-valued case.

7.7. Lemma. *Let X be a Banach function lattice. Then we have*

$$|\langle b, f \rangle| \lesssim \|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X)} \|S_{\mathcal{D}}f\|_{L^1(X')}.$$

Proof. We may assume $\|S_{\mathcal{D}}f\|_{L^1(X')} < \infty$. Estimate

$$|\langle b, f \rangle| \leq \sum_{R \in \mathcal{D}} \{|\langle b, h_R \rangle|, |\langle f, h_R \rangle|\}_X.$$

Given $k \in \mathbb{Z}$ we define

$$U_k = \{x : |S_{\mathcal{D}}f(x)|_{X'} > 2^{-k}\},$$

and $\widehat{\mathcal{R}}_k = \{R \in \mathcal{D} : |R \cap U_k| > |R|/2\}$. If $R \in \widehat{\mathcal{R}}_k$, then $R \subset \tilde{U}_k := \{x : M_{\mathcal{D}}1_{U_k} > 1/2\}$, and, of course, $|\tilde{U}_k| \lesssim |U_k|$.

For all $R_0 \in \mathcal{D}$ and $x \in R_0$ we have

$$|S_{\mathcal{D}}f(x)|_{X'} \geq \frac{|\langle f, h_{R_0} \rangle|_{X'}}{|R_0|^{1/2}}.$$

If $|\langle f, h_{R_0} \rangle|_{X'} > 0$, then this implies $R_0 \in \widehat{\mathcal{R}}_k$ for all large enough k . As $|U_k| \rightarrow 0$ when $k \rightarrow -\infty$ we also have that $R_0 \notin \widehat{\mathcal{R}}_k$ for all small enough k . Let $\mathcal{R}_k = \widehat{\mathcal{R}}_k \setminus \widehat{\mathcal{R}}_{k-1}$, $k \in \mathbb{Z}$, and notice that by above

$$\begin{aligned} & \sum_{R \in \mathcal{D}} \{|\langle b, h_R \rangle|, |\langle f, h_R \rangle|\}_X \\ & \leq 2 \sum_{k \in \mathbb{Z}} \sum_{R \in \mathcal{R}_k} \left\langle |\langle b, h_R \rangle| \frac{1_R}{|R|^{1/2}}, |\langle f, h_R \rangle| \frac{1_R}{|R|^{1/2}} 1_{\tilde{U}_k} 1_{U_{k-1}^c} \right\rangle \\ & \lesssim \sum_{k \in \mathbb{Z}} \left\| \left(\sum_{\substack{R \in \mathcal{D} \\ R \subset \tilde{U}_k}} |\langle b, h_R \rangle|^2 \frac{1_R}{|R|} \right)^{1/2} \right\|_{L^2(X)} \left\| \left(\sum_{R \in \mathcal{D}} |\langle f, h_R \rangle|^2 \frac{1_R}{|R|} \right)^{1/2} 1_{\tilde{U}_k} 1_{U_{k-1}^c} \right\|_{L^2(X')} \\ & \leq \sum_{k \in \mathbb{Z}} \|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X)} |\tilde{U}_k|^{1/2} \times \|2^{-(k-1)} 1_{\tilde{U}_k}\|_{L^2} \\ & \lesssim \sum_{k \in \mathbb{Z}} \|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X)} 2^{-k} |U_k| \sim \|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X)} \|S_{\mathcal{D}}f\|_{L^1(X')}. \quad \square \end{aligned}$$

Paraproducts. When we represent a singular integral using dyadic model operators [22, 23], one type of model operator that appears is a paraproduct. However, this will not be the only source of paraproducts in our arguments. In fact, the main source comes from the desire to expand products of functions bf – that appear naturally in commutators – using paraproducts. In this section we explain how this is done.

For $i \in \{1, 2\}$ we define the one-parameter paraproducts

$$A_1^i(b, f) = \sum_{I_i \in \mathcal{D}^i} \Delta_{I_i}^i b \Delta_{I_i}^i f, \quad A_2^i(b, f) = \sum_{I_i \in \mathcal{D}^i} \Delta_{I_i}^i b E_{I_i}^i f \quad \text{and} \quad A_3^i(b, f) = \sum_{I_i \in \mathcal{D}^i} E_{I_i}^i b \Delta_{I_i}^i f.$$

By writing $b = \sum_{I_i} \Delta_{I_i}^i b$ and $f = \sum_{J_i} \Delta_{J_i}^i f$, and collapsing sums such as $\sum_{J_i: I_i \subsetneq J_i} \Delta_{J_i}^i f = E_{I_i}^i f$, we formally have

$$bf = \sum_{I_i} \Delta_{I_i}^i b \Delta_{I_i}^i f + \sum_{I_i \subsetneq J_i} \Delta_{I_i}^i b \Delta_{J_i}^i f + \sum_{J_i \subsetneq I_i} \Delta_{I_i}^i b \Delta_{J_i}^i f = \sum_{j=1}^3 A_j^i(b, f).$$

Sometimes, when we pair bf with a tensor product of Haar functions, only one of the appearing Haar functions is cancellative. In this type of a situation we expand using these one-parameter paraproducts, like above, in the parameter $i \in \{1, 2\}$ where the cancellative Haar function appears.

A feature of such decompositions is that the paraproduct $A_3^i(b, f)$ appears (e.g. such paraproducts do not appear in the dyadic representation of singular integrals). Indeed, there is an average on b , and so no BMO type philosophy on that parameter can characterize its boundedness. However, these will appear in commutator decompositions and the cancellation present in the commutator is key to handling such terms. In fact, none of the one-parameter paraproducts are bounded by themselves with the assumption $b \in \text{BMO}_{\text{prod}}$.

We continue to define bi-parameter paraproducts, which can be obtained by chaining two one-parameter paraproducts. For $j_1, j_2 \in \{1, 2, 3\}$ define formally

$$A_{j_1, j_2}(b, f) = A_{j_1}^1 A_{j_2}^2(b, f)$$

so that e.g.

$$A_{1,2}(b, f) = \sum_{I_2 \in \mathcal{D}^2} A_1^1(\Delta_{I_2}^2 b, E_{I_2}^2 f) = \sum_{R=I_1 \times I_2 \in \mathcal{D}} \Delta_R b \Delta_{I_1}^1 E_{I_2}^2 f.$$

We can again expand the product bf as follows

$$bf = \sum_{j_1, j_2 \in \{1, 2, 3\}} A_{j_1, j_2}(b, f).$$

Now, those bi-parameter paraproducts where $j_1, j_2 \in \{1, 2\}$, will be automatically bounded if b is in product BMO. The rest will have to be handled by exploiting the cancellation of the commutator.

We now formulate the vector-valued setting, where we can prove the boundedness of our paraproducts.

7.8. Definition. We say that (X_1, X_2, X_3) is a compatible triple of function lattices, if each X_i is a Banach lattice defined using the same underlying measure space (Ω, μ) , the spaces X_2 and X_3 are UMD, and we have

$$|x_1 x_2|_{X_3} \leq |x_1|_{X_1} |x_2|_{X_2}.$$

7.9. Proposition. Let $p_1, p_2 \in (1, \infty)$ and (X_1, X_2, X_3) be a compatible triple of function lattices in the sense of Definition 7.8. For $j_1, j_2 \in \{1, 2\}$ we have the paraproduct estimate

$$\|A_{j_1, j_2}(b, f)\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X_3)} \lesssim \|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X_1)} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X_2)}.$$

Proof. We will show the estimate for $A_{1,2}(b, f)$ – the other cases are very similar and follow by dominating the pairings below by different combinations of square functions and maximal functions.

Let $g \in L_{x_1}^{p'_1} L_{x_2}^{p'_2}(X'_3)$ and estimate

$$\begin{aligned} |\langle A_{1,2}(b, f), g \rangle| &\leq \sum_{R=I_1 \times I_2} \left\{ |\langle b, h_R \rangle|, \left| \left\langle f, h_{I_1} \otimes \frac{1_{I_2}}{|I_2|} \right\rangle \right| |\langle g, h_{I_1} h_{I_1} \otimes h_{I_2} \rangle| \right\} \\ &\lesssim \|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X_1)} \left\| \left(\sum_{R=I_1 \times I_2} \left| \left\langle f, h_{I_1} \otimes \frac{1_{I_2}}{|I_2|} \right\rangle \right|^2 |\langle g, h_{I_1} h_{I_1} \otimes h_{I_2} \rangle|^2 \frac{1_R}{|R|} \right)^{1/2} \right\|_{L^1(X'_1)}. \end{aligned}$$

Here we used Lemma 7.7. We can now further dominate this by

$$\|S_{\mathcal{D}^1, M_{\mathcal{D}^2}} f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X_2)} \|S_{\mathcal{D}^2, M_{\mathcal{D}^1}} g\|_{L_{x_1}^{p'_1} L_{x_2}^{p'_2}(X'_3)},$$

and the proof is finished using Lemma 7.5. \square

7.10. *Remark.* In the one-parameter case, paraproduct estimates are known to hold in all UMD spaces (i.e., without imposing any additional function lattice structure), but the proofs become slightly more demanding under such minimal assumptions. For background and a modern proof of such estimates see e.g. [28].

We are now ready to give the proof of the bi-parameter analogue of (7.4). The proof follows the modern paradigm of dyadic representation theorems of singular integrals combined with the paraproduct decompositions of commutators.

7.11. *Remark.* A vector-valued estimate for $[b, T]$, where T is a bi-parameter CZO [41], could extremely likely also be proved with an adaptation of the strategy below.

7.12. **Theorem.** *Let $p_1, p_2 \in (1, \infty)$ and (X_1, X_2, X_3) be a compatible triple of function lattices in the sense of Definition 7.8. Let T_i be a CZO in \mathbb{R}^{d_i} , $i = 1, 2$. Then we have*

$$\|[T_1, [b, T_2]]f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2}(X_3)} \lesssim \|b\|_{\text{BMO}_{\text{prod}}(X_1)} \|f\|_{L_{x_1}^{p'_1} L_{x_2}^{p'_2}(X_3)}.$$

Proof. The proof is a vector-valued adaptation of the recent commutator decompositions [36, 37, 38]. In particular, see [37] for a two-weight version of this bound in the scalar-valued setting. To begin the proof, we recall that it is enough to fix lattices \mathcal{D}^i and to estimate $[U_1, [b, U_2]]f$, where U_1 and U_2 are DMOs – dyadic model operators – appearing in the representation theorem [22, 23]. That is, we have that $U_i \in \{S_i, \pi_i\}$, where S_i is a so-called dyadic shift in \mathbb{R}^{d_i} and π_i is a dyadic paraproduct in \mathbb{R}^{d_i} . We will recall the definitions during the proof. It is enough to obtain a polynomial dependence on the complexity of the shift.

We will only explicitly show the case $[S_1, [b, \pi_2]]f$ (see [37] for the decompositions used in the other cases). The dyadic shift S_1 has the form

$$S_1 f = \sum_{\substack{K_1 \in \mathcal{D}^1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} a_{K_1, I_1, J_1} h_{J_1} \otimes \langle f, h_{I_1} \rangle_1,$$

where $i_1, j_1 \geq 0$ and $|a_{K_1, I_1, J_1}| \leq |I_1|^{1/2} |J_1|^{1/2} |K_1|^{-1}$. The paraproduct π_2 has the form

$$\pi_2 f = \sum_{K_2 \in \mathcal{D}^2} a_{K_2} \langle f \rangle_{K_2, 2} \otimes h_{K_2},$$

where $\|(a_{K_2})\|_{\text{BMO}(\mathcal{D}^2)} = \sup_{P_2 \in \mathcal{D}^2} \left(\frac{1}{|P_2|} \sum_{K_2 \subset P_2} |a_{K_2}|^2 \right)^{1/2} \leq 1$. We now write out

$$[S_1, [b, \pi_2]]f = S_1(b\pi_2 f) - b\pi_2 S_1 f - S_1 \pi_2(bf) + \pi_2(bS_1 f).$$

We identify the products

$$b\pi_2 f, b\pi_2 S_1 f, bf, bS_1 f$$

that we want to expand using paraproducts. Of course, it is possible to expand these in the one-parameter sense (in either parameter) or in the bi-parameter sense. What is actually useful is determined by the form of the operators S_1 and π_2 , and looking at where the cancellative Haar functions appear. This leads us to e.g. expand

$$\begin{aligned} b\pi_2 f &= \sum_{k_1, k_2 \in \{1, 2, 3\}} A_{k_1, k_2}(b, \pi_2 f) \\ &= \sum_{k_1, k_2 \in \{1, 2\}} A_{k_1, k_2}(b, \pi_2 f) + \sum_{\substack{(k_1, k_2) \neq (3, 3) \\ 3 \in \{k_1, k_2\}}} A_{k_1, k_2}(b, \pi_2 f) + A_{3, 3}(b, \pi_2 f), \end{aligned}$$

where the idea of the latter grouping is that in the case $k_1, k_2 \in \{1, 2\}$ we can directly use Proposition 7.9, while the remaining terms cannot be handled alone. They need to be grouped with terms coming from the other product expansions, which are as follows:

$$b\pi_2 S_1 f = \sum_{k_1, k_2 \in \{1, 2\}} A_{k_1, k_2}(b, \pi_2 S_1 f) + \sum_{\substack{(k_1, k_2) \neq (3, 3) \\ 3 \in \{k_1, k_2\}}} A_{k_1, k_2}(b, \pi_2 S_1 f) + A_{3, 3}(b, \pi_2 S_1 f),$$

$$bf = \sum_{k_1=1}^2 A_{k_1}^1(b, f) + A_3^1(b, f)$$

and

$$bS_1 f = \sum_{k_1=1}^2 A_{k_1}^1(b, S_1 f) + A_3^1(b, S_1 f).$$

None of the one-parameter paraproducts are bounded as they are. However, when $k_1 \neq 3$, they will be paired with **one** other term coming from the other product expansions. In this sense, they are similar to the terms of the form $A_{k_1, k_2}(b, F)$, $(k_1, k_2) \neq (3, 3)$, $3 \in \{k_1, k_2\}$. Finally, all the four terms $A_{3, 3}(b, \pi_2 f)$, $A_{3, 3}(b, \pi_2 S_1 f)$, $A_3^1(b, f)$ and $A_3^1(b, S_1 f)$ will be grouped together.

The above explanation leads us to the long expansion

$$\begin{aligned} [S_1, [b, \pi_2]]f &= \sum_{k_1, k_2 \in \{1, 2\}} S_1(A_{k_1, k_2}(b, \pi_2 f)) - \sum_{k_1, k_2 \in \{1, 2\}} A_{k_1, k_2}(b, \pi_2 S_1 f) \\ &+ \left\{ \sum_{\substack{(k_1, k_2) \neq (3, 3) \\ 3 \in \{k_1, k_2\}}} S_1(A_{k_1, k_2}(b, \pi_2 f)) + \sum_{k_1=1}^2 \pi_2(A_{k_1}^1(b, S_1 f)) \right. \\ &\quad \left. - \sum_{\substack{(k_1, k_2) \neq (3, 3) \\ 3 \in \{k_1, k_2\}}} A_{k_1, k_2}(b, \pi_2 S_1 f) - \sum_{k_1=1}^2 S_1 \pi_2(A_{k_1}^1(b, f)) \right\} \\ &+ \{[S_1(A_{3, 3}(b, \pi_2 f)) - S_1 \pi_2(A_3^1(b, f))] + [\pi_2(A_3^1(b, S_1 f)) - A_{3, 3}(b, \pi_2 S_1 f)]\}. \end{aligned}$$

Using Proposition 7.9 and the fact (see [28]) that S_1 and π_2 are bounded in $L_{x_1}^{s_1} L_{x_2}^{s_2}(X)$ for all $s_1, s_2 \in (1, \infty)$ and for all UMD function lattices X , we have that the first two terms in the expansion above automatically satisfy the desired estimate. The big group in the

middle consists of the various terms, where exactly one of the paraproduct indices is 3 or we have a one-parameter paraproduct with an indice $k_1 \neq 3$. They will not all be handled together, but rather by always pairing two suitable ones together (notice that there are 6 terms with a plus sign and 6 terms with a minus sign). It is not straightforward to explain which terms should be paired together – it can be seen by analysing their forms. We get to them later.

We start our work with the last term, which is the combination of the four terms, where all the paraproduct indices are 3. First, we have by the definition of $A_{3,3}$ and π_2 that

$$\begin{aligned} A_{3,3}(b, \pi_2 f) &= \sum_{L_1, L_2} \langle b \rangle_{L_1 \times L_2} \langle \pi_2 f, h_{L_1} \otimes h_{L_2} \rangle h_{L_1} \otimes h_{L_2} \\ &= \sum_{L_1, L_2} \langle b \rangle_{L_1 \times L_2} \left\langle \sum_{K_2} a_{K_2} \langle f \rangle_{K_2, 2} \otimes h_{K_2}, h_{L_1} \otimes h_{L_2} \right\rangle h_{L_1} \otimes h_{L_2} \\ &= \sum_{L_1, K_2} a_{K_2} \langle b \rangle_{L_1 \times K_2} \left\langle f, h_{L_1} \otimes \frac{1_{K_2}}{|K_2|} \right\rangle h_{L_1} \otimes h_{K_2}. \end{aligned}$$

Now, using the definition of S_1 we calculate

$$\begin{aligned} S_1(A_{3,3}(b, \pi_2 f)) &= \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} a_{K_1, I_1, J_1} h_{J_1} \otimes \langle A_{3,3}(b, \pi_2 f), h_{I_1} \rangle_1 \\ &= \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} a_{K_1, I_1, J_1} h_{J_1} \otimes \left\langle \sum_{L_1, K_2} a_{K_2} \langle b \rangle_{L_1 \times K_2} \left\langle f, h_{L_1} \otimes \frac{1_{K_2}}{|K_2|} \right\rangle h_{L_1} \otimes h_{K_2}, h_{I_1} \right\rangle_1 \\ &= \sum_{K_2} a_{K_2} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} a_{K_1, I_1, J_1} \langle b \rangle_{I_1 \times K_2} \left\langle f, h_{I_1} \otimes \frac{1_{K_2}}{|K_2|} \right\rangle h_{J_1} \otimes h_{K_2}. \end{aligned}$$

The next task is to perform a completely analogous calculation to get that

$$S_1 \pi_2(A_3^1(b, f)) = \sum_{K_2} a_{K_2} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} a_{K_1, I_1, J_1} \langle \langle b \rangle_{I_1, 1} \langle f, h_{I_1} \rangle_1 \rangle_{K_2} h_{J_1} \otimes h_{K_2}.$$

We can now write

$$(7.13) \quad \begin{aligned} &S_1(A_{3,3}(b, \pi_2 f)) - S_1 \pi_2(A_3^1(b, f)) \\ &= \sum_{K_2} a_{K_2} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} a_{K_1, I_1, J_1} \left\langle [\langle b \rangle_{I_1 \times K_2} - \langle b \rangle_{I_1, 1}] \langle f, h_{I_1} \rangle_1 \right\rangle_{K_2} h_{J_1} \otimes h_{K_2}. \end{aligned}$$

Here the term $\langle [\langle b \rangle_{I_1 \times K_2} - \langle b \rangle_{I_1, 1}] \langle f, h_{I_1} \rangle_1 \rangle_{K_2}$ appears, which we further expand by applying the martingale difference expansion $g 1_{K_2} = \sum_{I_2 \subset K_2} \langle g, h_{I_2} \rangle h_{I_2} + \langle g \rangle_{K_2} 1_{K_2}$ to the function $\langle b \rangle_{I_1, 1}$. This gives that

$$\left\langle [\langle b \rangle_{I_1 \times K_2} - \langle b \rangle_{I_1, 1}] \langle f, h_{I_1} \rangle_1 \right\rangle_{K_2} = - \left\langle \left[\sum_{I_2 \subset K_2} \langle \langle b, h_{I_2} \rangle_2 \rangle_{I_1} h_{I_2} \right] \langle f, h_{I_1} \rangle_1 \right\rangle_{K_2}$$

$$= -\frac{1}{|K_2|} \sum_{I_2 \subset K_2} \langle \langle b, h_{I_2} \rangle_2 \rangle_{I_1} \langle f, h_{I_1} \otimes h_{I_2} \rangle.$$

Plugging this into (7.13) gives the final formula

$$\begin{aligned} & S_1(A_{3,3}(b, \pi_2 f)) - S_1 \pi_2(A_3^1(b, f)) \\ &= -\sum_{K_2} \frac{a_{K_2}}{|K_2|} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{I_2 \subset K_2} a_{K_1, I_1, J_1} \langle \langle b, h_{I_2} \rangle_2 \rangle_{I_1} \langle f, h_{I_1} \otimes h_{I_2} \rangle_{h_{J_1} \otimes h_{K_2}}. \end{aligned}$$

This is clear progress as there is now one cancellative Haar function paired with b . On the other hand, this is simply not enough, since we need two cancellative Haar functions to exploit the product BMO assumption. However, with very similar calculation as above we see that

$$\begin{aligned} & \pi_2(A_3^1(b, S_1 f)) - A_{3,3}(b, \pi_2 S_1 f) \\ &= \sum_{K_2} \frac{a_{K_2}}{|K_2|} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{I_2 \subset K_2} a_{K_1, I_1, J_1} \langle \langle b, h_{I_2} \rangle_2 \rangle_{J_1} \langle f, h_{I_1} \otimes h_{I_2} \rangle_{h_{J_1} \otimes h_{K_2}}. \end{aligned}$$

The only difference is that we have a plus sign in front and that $\langle \langle b, h_{I_2} \rangle_2 \rangle_{I_1}$ is replaced by $\langle \langle b, h_{I_2} \rangle_2 \rangle_{J_1}$. We have arrived at the combined formula

$$\begin{aligned} & [S_1(A_{3,3}(b, \pi_2 f)) - S_1 \pi_2(A_3^1(b, f))] + [\pi_2(A_3^1(b, S_1 f)) - A_{3,3}(b, \pi_2 S_1 f)] \\ &= \sum_{K_2} \frac{a_{K_2}}{|K_2|} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{I_2 \subset K_2} a_{K_1, I_1, J_1} [\langle \langle b, h_{I_2} \rangle_2 \rangle_{J_1} - \langle \langle b, h_{I_2} \rangle_2 \rangle_{I_1}] \\ & \quad \times \langle f, h_{I_1} \otimes h_{I_2} \rangle_{h_{J_1} \otimes h_{K_2}}. \end{aligned}$$

Next, we add and subtract $\langle \langle b, h_{I_2} \rangle_2 \rangle_{K_1}$ and e.g. write

$$\langle \langle b, h_{I_2} \rangle_2 \rangle_{J_1} - \langle \langle b, h_{I_2} \rangle_2 \rangle_{K_1} = \sum_{l_1=1}^{j_1} \langle b, h_{J_1^{(l_1)}} \otimes h_{I_2} \rangle_{h_{J_1^{(l_1)}} \otimes h_{K_2}}.$$

It is now enough to fix $l_1 \in \{1, \dots, j_1\}$ and consider, for $g \in L_{x_1}^{p'_1} L_{x_2}^{p'_2}(X'_3)$, the following dualized term

$$\begin{aligned} & \sum_{K_1} \sum_{L_1^{(j_1-l_1)} = K_1} \sum_{I_2} \left\{ |\langle b, h_{L_1} \otimes h_{I_2} \rangle|, \right. \\ & \quad \left. \sum_{\substack{I_1^{(i_1)} = K_1 \\ J_1^{(l_1)} = L_1}} |L_1|^{-1/2} |a_{K_1, I_1, J_1}| |\langle f, h_{I_1} \otimes h_{I_2} \rangle| \sum_{K_2 \supset I_2} \frac{|a_{K_2}|}{|K_2|} |\langle g, h_{J_1} \otimes h_{K_2} \rangle| \right\}. \end{aligned}$$

Notice that we can estimate

$$\begin{aligned} \sum_{K_2 \supset I_2} \frac{|a_{K_2}|}{|K_2|} |\langle g, h_{J_1} \otimes h_{K_2} \rangle| &= \left\langle \sum_{K_2 \supset I_2} |a_{K_2}| |\langle g, h_{J_1} \otimes h_{K_2} \rangle| \frac{1_{K_2}}{|K_2|} \right\rangle_{I_2} \\ &\leq |J_1|^{1/2} \left\langle \sum_{K_2} |a_{K_2}| |\langle g, h_{K_2} \rangle_2| \frac{1_{K_2}}{|K_2|} \right\rangle_{I_2}, \end{aligned}$$

and use the normalisation of $|a_{K_1, I_1, J_1}|$ to estimate

$$\begin{aligned} |L_1|^{-1/2} |a_{K_1, I_1, J_1}| |\langle f, h_{I_1} \otimes h_{I_2} \rangle| &\leq |L_1|^{-1/2} \frac{|I_1|^{1/2} |J_1|^{1/2}}{|K_1|} |\langle \langle \Delta_{K_1, i_1} f, h_{I_2} \rangle_2, h_{I_1} \rangle| \\ &\leq |L_1|^{-1/2} \frac{|J_1|^{1/2}}{|K_1|} \int_{I_1} |\langle \Delta_{K_1, i_1} f, h_{I_2} \rangle_2|. \end{aligned}$$

Now, by summing over I_1, J_1 we have reduced to bounding

$$\begin{aligned} \sum_{K_1} \sum_{L_1^{(j_1-i_1)}=K_1} \sum_{I_2} \left\{ |\langle b, h_{L_1} \otimes h_{I_2} \rangle|, \right. \\ \left. |L_1|^{1/2} \langle |\langle \Delta_{K_1, i_1} f, h_{I_2} \rangle_2 \rangle_{K_1} \left\langle \sum_{K_2} |a_{K_2}| \langle |\langle g, h_{K_2} \rangle_2 \rangle_{L_1} \frac{1_{K_2}}{|K_2|} \right\rangle_{I_2} \right\}. \end{aligned}$$

We use Lemma 7.7 to dominate this by $\|b\|_{\text{BMO}_{\mathcal{D}, \text{prod}}(X_1)} \leq \|b\|_{\text{BMO}_{\text{prod}}(X_1)}$ multiplied with

$$\left\| \left(\sum_{K_1} \sum_{I_2} [M_{\mathcal{D}^1} \langle \Delta_{K_1, i_1} f, h_{I_2} \rangle_2]^2 \otimes \frac{1_{I_2}}{|I_2|} \right)^{1/2} M_{\mathcal{D}} \left(\sum_{K_2} |a_{K_2}| \langle |\langle g, h_{K_2} \rangle_2 \rangle_{L_1} \otimes \frac{1_{K_2}}{|K_2|} \right) \right\|_{L^1(X'_1)}.$$

After using the estimate $|e_2 e'_3|_{X'_1} \leq |e_2|_{X_2} |e'_3|_{X'_3}$ this term decouples naturally by Hölder's inequality. After this the proof can be ended using Lemma 7.5 and the $L_{x_1}^{p'_1} L_{x_2}^{p'_2}(X'_3)$ boundedness of the paraproduct

$$g \mapsto \sum_{K_2} |a_{K_2}| \langle |\langle g, h_{K_2} \rangle_2 \rangle_{L_1} \otimes \frac{1_{K_2}}{|K_2|}.$$

It only remains to estimate the terms in the big block, i.e., the terms

$$\begin{aligned} \sum_{\substack{(k_1, k_2) \neq (3, 3) \\ 3 \in \{k_1, k_2\}}} S_1(A_{k_1, k_2}(b, \pi_2 f)) + \sum_{k_1=1}^2 \pi_2(A_{k_1}^1(b, S_1 f)) \\ - \sum_{\substack{(k_1, k_2) \neq (3, 3) \\ 3 \in \{k_1, k_2\}}} A_{k_1, k_2}(b, \pi_2 S_1 f) - \sum_{k_1=1}^2 S_1 \pi_2(A_{k_1}^1(b, f)). \end{aligned}$$

Here we always have to pair two terms together, one with a plus and one with a minus sign, to induce cancellation to b . For example, we will consider

$$S_1(A_{3,1}(b, \pi_2 f)) - A_{3,1}(b, \pi_2 S_1 f).$$

Using that $\pi_2 S_1 = S_1 \pi_2$ and the boundedness of π_2 , it is enough to consider

$$\begin{aligned} S_1(A_{3,1}(b, f)) - A_{3,1}(b, S_1 f) \\ = \sum_{\substack{K_1 \\ I_1^{(i_1)}=J_1^{(j_1)}=K_1}} \sum_{K_2} a_{K_1, I_1, J_1} (\langle \langle b, h_{K_2} \rangle_2 \rangle_{I_1} - \langle \langle b, h_{K_2} \rangle_2 \rangle_{J_1}) \langle f, h_{I_1} \otimes h_{K_2} \rangle_{h_{J_1} \otimes h_{K_2}}. \end{aligned}$$

This can be bounded like the previous term, except it is simpler. This ends our proof. \square

We note that the upper bound related to the lower bound of Proposition 5.3 – that is, the upper bound related to the diagonal $p_1 = q_1$ and $p_2 = q_2$ of Theorem 1.3 – is just the scalar-valued version of the above result. However, this scalar-valued result is very well-known. For general CZOs like here, the non-mixed norm version appears in [9]. The scalar-valued mixed-norm version could also directly be proved e.g. by extrapolating the estimate in [37], which is a two-weight version of [9].

Lastly, even if the bi-commutators $[T_1, [b, T_2]]$ are the main topic of this paper, we record how the vector-valued estimate for the bi-commutator implies an estimate for a commutator with three CZOs. Establishing a dotted version of the space $L_{x_3}^{r_3}$ would require some modifications as in the bi-commutator case.

7.14. Proposition. *Let $p_1, p_2, p_3, q_3 \in (1, \infty)$, $p_3 > q_3$ and define $r_3 \in (1, \infty)$ via*

$$\frac{1}{r_3} = \frac{1}{p_3} - \frac{1}{p_3}.$$

Let T_i be a CZO in \mathbb{R}^{d_i} , $i = 1, 2, 3$. Then we have

$$\|[T_1, [T_2, [b, T_3]]]f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} L_{x_3}^{q_3}} \lesssim \|b\|_{\text{BMO}_{\text{prod}, x_1, x_2}(L_{x_3}^{r_3})} \|f\|_{L_{x_1}^{p_1} L_{x_2}^{p_2} L_{x_3}^{p_3}}.$$

It is an interesting, but very complicated, problem to extend all of our results to tri-commutators, or even higher order commutators.

APPENDIX A. A COUNTEREXAMPLE

The Ferguson–Lacey theorem [15] claims that

$$\|[H_1, [b, H_2]]\|_{L_{x_1}^2 L_{x_2}^2} \sim \|b\|_{\text{BMO}(\mathbb{R} \times \mathbb{R})},$$

where the right-hand side is the product-BMO norm of Chang–Fefferman, and H_i is the Hilbert transform in the i th coordinate. This result has been extended to higher order iterated commutators and more general operators in place of H_i in several subsequent papers.

However, all these results suffer from a gap that was first observed by A. Volberg and recently presented in [21, Section 10]. Namely, they all rely on a claim of the form

$$(A.1) \quad \|P\phi\|_2 \gtrsim \|\phi\|_2,$$

where P is a projection given by a Fourier multiplier whose symbol is the indicator of a product of half-spaces (in [15], simply $\mathbb{R}_+ \times \mathbb{R}_+ \subset \mathbb{R}^2$) and the Fourier transform of ϕ is “symmetric”. In practice, ϕ takes the form $\phi = |\gamma|^2$ for some function γ , so that ϕ is non-negative and in particular real-valued. On the other hand, in each case, the function γ is relatively arbitrary, so that little more than the non-negativity of ϕ can be asserted without further justification. This, on the other hand, is not enough to guarantee the estimate (A.1), as a simple example given in [21, Section 10] shows. We will give another example here with some additional properties.

We define the Fourier transform of $f \in L^1(\mathbb{R}^d)$ by

$$\hat{f}(\xi) := \int_{\mathbb{R}^d} f(x) e^{-i2\pi x \cdot \xi} dx.$$

With this normalization, the Gaussian $\varphi(x) = e^{-\pi|x|^2}$ satisfies $\hat{\varphi} = \varphi$.

If $Q \in \mathcal{L}(\mathbb{R}^d)$ is an invertible matrix, a standard change-of-variables shows that

$$[f(Q\cdot)]^\wedge(\xi) = \hat{f}(Q^{-\top}\xi) |\det Q|^{-1}.$$

In particular, we have

$$[f(Q\cdot)]^\wedge(\xi) = \hat{f}(Q^{-1}\xi), \quad \text{if } Q^\top = Q \text{ and } \det Q = 1.$$

Let then $d = 2$ and

$$Q_N := \frac{1}{2\sqrt{N}} \begin{pmatrix} 1+N & 1-N \\ 1-N & 1+N \end{pmatrix}.$$

Then $Q_N^\top = Q_N$, $\det Q_N = 1$, and

$$Q_N^{-1} := \frac{1}{2\sqrt{N}} \begin{pmatrix} N+1 & N-1 \\ N-1 & N+1 \end{pmatrix}.$$

We consider the functions $f_N(x) := \varphi(Q_N x)$. These satisfy

$$\hat{f}_N(\xi) = \varphi(Q_N^{-1}\xi) = \exp(-\pi|Q_N^{-1}\xi|^2),$$

where

$$|Q_N^{-1}\xi|^2 = \left| \frac{1}{2\sqrt{N}} \begin{pmatrix} (N+1)\xi_1 + (N-1)\xi_2 \\ (N-1)\xi_1 + (N+1)\xi_2 \end{pmatrix} \right|^2 = \frac{1}{2N} \left(N^2(\xi_1 + \xi_2)^2 + (\xi_1 - \xi_2)^2 \right).$$

Now consider $\xi_1, \xi_2 \geq 0$. Then $(\xi_1 + \xi_2)^2 \geq \xi_1^2 + \xi_2^2 = |\xi|^2$, and thus

$$|Q_N^{-1}\xi|^2 \geq \frac{N}{2} |\xi|^2 = \left| \sqrt{\frac{N}{2}} \xi \right|^2, \quad \forall \xi_1, \xi_2 \geq 0.$$

This in turn implies that

$$\begin{aligned} \iint_{\xi_1, \xi_2 \geq 0} |\hat{f}_N(\xi)|^2 d\xi &= \iint_{\xi_1, \xi_2 \geq 0} |\varphi(Q_N^{-1}\xi)|^2 d\xi \leq \iint_{\xi_1, \xi_2 \geq 0} \left| \varphi\left(\sqrt{\frac{N}{2}} \xi\right) \right|^2 d\xi \\ &= \frac{2}{N} \iint_{\xi_1, \xi_2 \geq 0} |\varphi(\xi)|^2 d\xi = \frac{1}{2N} \iint_{\mathbb{R}^2} |\varphi(\xi)|^2 d\xi, \end{aligned}$$

whereas

$$\iint_{\mathbb{R}^2} |\hat{f}_N(\xi)|^2 d\xi = \iint_{\mathbb{R}^2} |\varphi(Q_N^{-1}\xi)|^2 d\xi = \iint_{\mathbb{R}^2} |\varphi(\xi)|^2 d\xi.$$

If P is the Fourier multiplier whose symbol is the indicator of $\mathbb{R}_+ \times \mathbb{R}_+$, then the non-negative function f_N satisfies

$$\|P f_N\|_2 = \|\widehat{P f_N}\|_2 = \|\hat{f}_N\|_{L^2(\mathbb{R}_+ \times \mathbb{R}_+)} \leq \frac{1}{\sqrt{2N}} \|\varphi\|_2 = \frac{1}{\sqrt{2N}} \|\hat{f}_N\|_2 = \frac{1}{\sqrt{2N}} \|f_N\|_2$$

showing the impossibility of (A.1).

Let us also observe that

$$\begin{aligned} \|f_N\|_p^p &= \int_{\mathbb{R}^2} |\varphi(Q_N x)|^p dx \\ &= \int_{\mathbb{R}^2} |\varphi(x)|^p dx \quad (\det Q_N = 1) \\ &= \int_{\mathbb{R}^2} e^{-\pi|x|^2 p} dx = \int_{\mathbb{R}^2} e^{-\pi|x|^2} \frac{dx}{p} = \frac{1}{p} \end{aligned}$$

for all $p \in (0, \infty)$, independent of N . That is, we have

$$\|f_N\|_p \sim_p 1 \quad \forall p \in (0, \infty),$$

another feature of the functions for which (A.1) is applied in [15], yet this additional property is still not enough to guarantee (A.1).

REFERENCES

- [1] A. Amenta, E. Lorist, M. Veraar, Rescaled extrapolation for vector-valued functions, *Publ. Mat.* 63 (2019) 155–182. [35](#)
- [2] A. Barron, J. M. Conde-Alonso, Y. Ou, G. Rey, Sparse domination and the strong maximal function. *Adv. Math.* 345 (2019), 1–26. [26](#)
- [3] J. Bourgain, Extension of a result of Benedek, Calderón and Panzone, *Ark. Mat.* 22 (1984) 91–95. [36](#)
- [4] C. Cabrelli, M. T. Lacey, U. Molter, J. Pipher, Variations on the theme of Journé’s lemma, *Houston J. Math.* 32 (2006), no. 3, 833–861. [3](#)
- [5] 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. [3](#)
- [6] 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. [3](#)
- [7] R. Coifman, R. Rochberg, G. Weiss, Factorization theorems for Hardy spaces in several variables, *Ann. of Math.* (2) 103 (1976) 611–635. [1](#), [2](#), [4](#)
- [8] R. Coifman, P.-L. Lions, Y. Meyer, S. Semmes. Compensated compactness and Hardy spaces, *J. Math. Pures Appl.* 72 (3) (1993) 247–286. [2](#)
- [9] L. Dalenc, Y. Ou, Upper bound for multi-parameter iterated commutators *Publ. Mat.* 60 (2016) 191–220. [45](#)
- [10] G. David, J.-L. Journé, A boundedness criterion for generalized Calderón-Zygmund operators. *Ann. of Math.* (2) 120 (1984) 371–397. [6](#)
- [11] X.T. Duong, J. Li, Y. Ou, J. Pipher, B. Wick, Commutators of multi-parameter flag singular integrals and applications, *Anal. PDE* 12 (2019) 1325–1355. [3](#), [4](#)
- [12] X.T. Duong, J. Li, Y. Ou, J. Pipher, B. Wick, Weighted Estimates of Singular Integrals and Commutators in the Zygmund Dilation Setting, preprint, arXiv:1905.00999, 2019 [3](#), [4](#)
- [13] X. T. Duong, J. Li, Y. Ou, B. Wick, D. Yang, Product BMO, Little BMO, and Riesz Commutators in the Bessel Setting, *J. Geom. Anal.* 28 (2018) 2558–2601. [3](#), [4](#)
- [14] R. Fefferman, E. Stein, Singular integrals on product spaces, *Adv. Math.* 45 (1982) 117–143. [3](#)
- [15] S. H. Ferguson, M. T. Lacey, A characterization of product BMO by commutators, *Acta Math.* 189 (2002) 143–160. [3](#), [4](#), [45](#), [47](#)
- [16] S. Ferguson, C. Sadosky, Characterizations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures, *J. Anal. Math.* 81 (2000) 239–267.
- [17] J. García-Cuerva, R. Macías, J. L. Torrea, The Hardy–Littlewood property of Banach lattices, *Israel J. Math.* 83 (1993) 177–201. [36](#)
- [18] I. Holmes, M. Lacey, B. Wick, Commutators in the two-weight setting, *Math. Ann.* 367 (2017) 51–80. [4](#)
- [19] I. Holmes, M. Lacey, B. Wick, Bloom’s inequality: commutators in a two-weight setting, *Arch. Math. (Basel)* 106 (2016) 53–63. [4](#)
- [20] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators, *Anal. PDE* 11 (2018) 1693–1740. [4](#)
- [21] I. Holmes, S. Treil, A. Volberg, Dyadic bi-parameter repeated commutator and dyadic product BMO, preprint, arXiv:2101.00763, 2021. [3](#), [45](#)
- [22] T. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, *Ann. of Math.* 175 (2012) 1473–1506. [6](#), [38](#), [40](#)
- [23] T. Hytönen, Representation of singular integrals by dyadic operators, and the A_2 theorem, *Expo. Math.* 35 (2017) 166–205. [6](#), [38](#), [40](#)
- [24] T. Hytönen, The L^p -to- L^q boundedness of commutators with applications to the Jacobian operator, preprint, arXiv:1804.11167, 2018. [2](#), [4](#), [6](#), [7](#), [8](#), [14](#), [16](#)
- [25] T. Hytönen, H. Martikainen, E. Vuorinen, Multi-parameter estimates via operator-valued shifts, *Proc. Lond. Math. Soc.* 119 (6) (2019) 1560–1597. [35](#), [36](#)

- [26] T. Hytönen, J. van Neerven, M. Veraar, L. Weis, *Analysis in Banach Spaces, Volume I: Martingales and Littlewood-Paley Theory*, Springer-Verlag, 2016. [12](#), [33](#), [37](#)
- [27] T. Hytönen, J. van Neerven, M. Veraar, L. Weis, *Analysis in Banach Spaces, Volume II: Probabilistic Methods and Operator Theory*, Springer-Verlag, 2017. [33](#), [35](#), [37](#)
- [28] T. Hänninen, T. Hytönen, Operator-valued dyadic shifts and the $T(1)$ theorem, *Monatsh. Math.* 180 (2016) 213–253. [40](#), [41](#)
- [29] T. Iwaniec, Nonlinear commutators and Jacobians. Proceedings of the conference dedicated to Professor Miguel de Guzmán (El Escorial, 1996). *J. Fourier Anal. Appl.* 3 (1997), Special Issue, 775–796. [2](#)
- [30] S. Janson, Mean oscillation and commutators of singular integral operators, *Ark. Mat.* 16 (1978), no. 2, 263–270. [2](#)
- [31] J.-L. Journé, A covering lemma for product spaces, *Proc. Amer. Math. Soc.* 96 (1986) 593–598. [3](#)
- [32] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Multiparameter Riesz commutators, *Amer. J. Math.* 131 (2009) 731–769. [3](#), [4](#)
- [33] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Iterated Riesz Commutators: A Simple Proof of Boundedness, *Contemp. Math.*, 505, Amer. Math. Soc., Providence, RI, 2010, 171–178. [3](#)
- [34] 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](#), [29](#), [34](#)
- [35] A. Lerner, S. Ombrosi, I. Rivera-Ríos, Commutators of singular integrals revisited, *Bull. London Math. Soc.* 51 (2019) 107–119. [4](#)
- [36] K. Li, H. Martikainen, E. Vuorinen, Bloom type inequality for bi-parameter singular integrals: efficient proof and iterated commutators, *Int. Math. Res. Not. IMRN* 11 (2021) 8153–8187. [4](#), [40](#)
- [37] K. Li, H. Martikainen, E. Vuorinen, Bloom type upper bounds in the product BMO setting, *J. Geom. Anal.* 30 (2020) 3181–3203. [4](#), [40](#), [45](#)
- [38] K. Li, H. Martikainen, E. Vuorinen, Bilinear Calderón-Zygmund theory on product spaces, *J. Math. Pures Appl.* 138 (2020) 356–412. [40](#)
- [39] J. Lindenstrauss, L. Tzafriri, *Classical Banach Spaces II: Function Spaces*, Springer-Verlag, 1979. [35](#)
- [40] E. Lorist, Maximal functions, factorization, and the \mathcal{R} -boundedness of integral operators, Master’s thesis, Delft University of Technology, Delft, the Netherlands, 2016. [35](#)
- [41] H. Martikainen, Representation of bi-parameter singular integrals by dyadic operators, *Adv. Math.* 229 (2012) 1734–1761. [40](#)
- [42] Z. Nehari, On bounded bilinear forms, *Ann. of Math.* 65 (1957) 153–162. [1](#), [4](#)
- [43] Y. Ou, Multi-parameter singular integral operators and representation theorem, *Rev. Mat. Iberoam.* 33 (2017) 325–350.
- [44] Y. Ou, S. Petermichl, E. Strouse, Higher order Journé commutators and characterizations of multi-parameter BMO, *Adv. Math.* 291 (2016) 24–58. [3](#)
- [45] S. Petermichl, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol, *C. R. Acad. Sci. Paris Sér. I Math.* 330 (2000) 455–460.
- [46] J. Pipher, Journé’s covering lemma and its extension to higher dimensions, *Duke Math. J.* 53 (1986), no. 3, 683–690. [3](#)
- [47] J. L. Rubio de Francia, Martingale and integral transforms of Banach space valued functions, *Lectures Notes in Mathematics* 1221, Springer-Verlag, 1985, pp. 195–222. [36](#)
- [48] C. Segovia, J. L. Torrea, Vector-valued commutators and applications, *Indiana Univ. Math. J.* 38 (1989) 959–971. [33](#)

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

Email address: `emil.airta@helsinki.fi`

Email address: `tuomas.hytonen@helsinki.fi`

Email address: `henri.martikainen@helsinki.fi`

Email address: `tuomas.v.oikari@helsinki.fi`

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

Email address: `kli@tju.edu.cn`