Annales Fennici Mathematici
Volumen 48, 2023, 81–112

Off-diagonal estimates for bi-parameter commutators

Tuomas Oikari

Abstract. We study the boundedness of commutators of bi-parameter singular integrals between mixed spaces

$$[b,T]\colon L^{p_1}L^{p_2}\to L^{q_1}L^{q_2}$$

in the off-diagonal situation $q_i,p_i\in(1,\infty)$ where we allow $q_i\neq p_i$. Boundedness is fully characterized in terms of oscillatory testing conditions on the function b for a total seven out of the nine possible arrangements of the integrability exponents.

Fraktionaalisia estimaatteja kahden parametrin kommutaattoreille

Tiivistelmä. Tutkimme kahden parametrin singulaari-integraalien kommutaattoreiden rajoittuneisuutta iteroitujen avaruuksien

$$[b,T]\colon L^{p_1}L^{p_2}\to L^{q_1}L^{q_2}$$

välillä kun $q_i,p_i\in(1,\infty)$ ja erityisesti fraktionaalisessa tilanteessa $q_i\neq p_i$. Yhteensä seitsemässä tapauksessa yhdeksästä saavutamme täyden karaterisoinnin rajoittuneisuudella symbolia b koskevilla oskillatorisilla testiehdoilla.

1. Introduction and preliminaries

The first commutator results concern the commutator of the Hilbert transform

$$[b,H]f=bHf-H(bf)$$

whose boundedness was characterized in the classical theorem of Nehari in [15] through Hankel operators. Later, Coifman, Rochberg and Weiss [3] generalized Nehari's result and showed that

$$\|b\|_{\mathrm{BMO}}\lesssim\sum_{i=j}^{d}\|[b,R_j]\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)}\lesssim\|b\|_{\mathrm{BMO}}:=\sup_I\fint_I|b-\langle b\rangle_I|,\quad p\in(1,\infty),\tag{1.1}$$

where the supremum is taken over all cubes $I\subset\mathbb{R}^d$ and $\langle b\rangle_I=\frac{1}{|I|}\int_I b$. The upper bound in (1.1) was proved for a wide class of bounded singular integrals, while the lower bound especially involved the Riesz transforms. Later, the lower bound in (1.1) was improved separately by both Janson [10] and Uchiyama [17] by finding non-degeneracy assumptions on the kernel that cover any single Riesz transform (in contrast to (1.1) involving all the d Riesz transforms). Janson [10] also covers the off-diagonal situation when $1<p<q<\infty$ and provides the characterization

$$\|[b,T]\|_{L^p\to L^q}\sim\sup_Q\ell(Q)^{-\alpha}\fint_Q|b-\langle b\rangle_Q|,\quad\alpha:=d\left(\frac{1}{p}-\frac{1}{q}\right).$$

https://doi.org/10.54330/afm.125812
2020 Mathematics Subject Classification: Primary 42B20.
Key words: Calderón–Zygmund operators, singular integrals, multi-parameter analysis, commutators.
T. Oikari was supported by the Academy of Finland project No. 306901, by the Finnish Centre of Excellence in Analysis and Dynamics Research project No. 307333, and by the three-year research grant of the University of Helsinki No. 75160010.

The remaining range with $1 < q < p < \infty$ was characterised through an approximate weak factorization (awf) argument by Hytönen in [7],

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

In this article we will also use the awf argument, but now in the bi-parameter setting, we recall the basic idea below in Section 2.

The $p = q$ characterization yields factorizations of the Hardy space H^1, see [3], and implies div-curl lemmas relevant for compensated compactness, see [2]. In Lindberg [13] and Hytönen [7] the characterization of the case $q < p$ is connected with a conjecture of Iwaniec [9] on the prescribed Jacobian problem. It is crucial in all of these applications that we have both commutator upper and lower bounds.

In this article, we work in the product ambient space $\mathbb{R}^d = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$ and study the boundedness of the bi-parameter commutators $[b,T]$, where T is now a bi-parameter singular integral operator. Due to the product space nature of the problem, it is natural to allow different integrability exponents in the first and the second parameter slots, thereby leading to the question of $L^{p_1}L^{p_2}$-to-$L^{q_1}L^{q_2}$ boundedness. With $L^{p_1}L^{p_2}$ or $L^{p_1}_{x_1}L^{p_2}_{x_2}$ we denote the mixed-norm space $L^{p_1}(\mathbb{R}^{d_1};L^{p_2}(\mathbb{R}^{d_2}))$. More precisely, we identify $f\colon \mathbb{R}^d\to\mathbb{C}$ satisfying

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

with the function $\phi_f \in L^{p_1}(\mathbb{R}^{d_1};L^{p_2}(\mathbb{R}^{d_2}))$, $\phi_f(x_1) = f(x_1,\cdot)$, and similarly for the other iterated function spaces appearing in this article, see e.g. the table below in Theorem 1.2.

In accordance with the three qualitatively different regimes $p < q$, $p = q$ and $p > q$ in the one-parameter setup, there will now be nine cases depending on the relative size of both of the pairs p_1, q_1 and p_2, q_2. The exact statements of our results are spread throughout the text; the following Theorem 1.2 is a condensed version of the obtained results.

1.2. Theorem. *Let T be a non-degenerate bi-parameter Calderón–Zygmund operator on $\mathbb{R}^d = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$, fix the exponents $p_1,p_2,q_1,q_2\in(1,\infty)$ and set*

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

Let $b\colon\mathbb{R}^d\to\mathbb{C}$ be a function with some local integrability depending on p_1,p_2,q_1,q_2 (L^∞_{loc} works in all cases, for example). Then, denoting $\|[b,T]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{q_1}_{x_1}L^{q_2}_{x_2}} = N_{p,q}$ we have the upper- and lower bounds

	$p_1<q_1$	$p_1=q_1$	$p_1>q_1$
$p_2<q_2$	$b=$ *constant*, $N_{p,q}=0$	$b(\cdot,x_2)=$ *constant*, $N_{p,q}\sim\|b(x_1,\cdot)\|_{\dot C^{0,\alpha_2}_{x_2}}$	$b=$ *constant*, $N_{p,q}=0$
$p_2=q_2$	$b(x_1,\cdot)=$ *constant*, $N_{p,q}\sim\|b(\cdot,x_2)\|_{\dot C^{0,\alpha_1}_{x_1}}$	$N_{p,q}\sim\|b\|_{\mathrm{bmo}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}$	$\inf_{c\in\mathbb{C}}\|b-c\|_{L^\infty_{x_2}L^{r_1}_{x_1}}\lesssim N_{p,q}\lesssim\inf_{c\in\mathbb{C}}\|b-c\|_{L^{r_1}_{x_1}L^\infty_{x_2}}$
$p_2>q_2$	$b=$ *constant* $N_{p,q}=0$	$N_{p,q}\sim\inf_{c\in\mathbb{C}}\|b-c\|_{L^\infty_{x_1}L^{r_2}_{x_2}}$	$N_{p,q}\lesssim\inf_{c\in\mathbb{C}}\|b-c\|_{L^{r_1}_{x_1}L^{r_2}_{x_2}}$

Our main focus is on the off-diagonal cases $(p_1, p_2) \neq (q_1, q_2)$ with the diagonal being well-known and lately studied e.g. by Holmes, Petermichl and Wick [6], and by Li, Martikainen and Vuorinen [12].

While some of the upper bounds in the off-diagonal situations in the table of Theorem 1.2 are quick by few applications of Hölder's inequality, or completely trivial in the constant cases, the rest are not effortless and require for example the use of the representation theorem and other tools inherent to the bi-parameter setting. Especially, the results do not follow as corollaries of the one-parameter theory. This being said, the most work is found with the lower bounds. We prove the lower bounds through the awf argument but now in the bi-parameter setting. In the two cases where we fail to achieve a full characterization, the problems are mainly due to the fact that the awf argument is symmetric with respect to both of the parameters, while the norm $\|\cdot\|_{L^t_{x_1} L^s_{x_2}}$ has a built-in order to it. This limitation is not new and was expected, as we already saw it in Airta, Hytönen, Li, Martikainen and Oikari [1], where we provided a similar table as in Theorem 1.2 above, but for the bi-parameter commutator $[T_2, [b, T_1]]$, where each T_i is a singular integral on $\mathbb{R}^{d_i}$. In [1] we achieved a fully satisfactory characterization of the boundedness of the commutator in only four cases, this is in line with $[T_2, [b, T_1]]$ being generally considered a harder operator to work with than $[b, T]$. Perhaps this difference is best reflected through the fact that the diagonal characterization ($p_1 = p_2 = q_1 = q_2$) in terms of the proposed product BMO is open in the first case, see for example the discussion in [1], whereas the boundedness of $[b, T]$ on the diagonal is fully understood and captured by the simpler little bmo.

Acknowledgements. We thank Henri Martikainen, Emil Vuorinen and Tuomas Hytönen for discussions and comments that improved the article.

In the remaining part of this section we provide the relevant definitions. The reader who is familiar with this material may immediately skip them and move to the next Section 2.

1.1. Singular integrals. We denote the diagonal with $\Delta = \Delta^{(d_i,d_i)} = \{(x_i, y_i) \in \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} : x_i = y_i\}$ and call

$$K_i \colon \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \setminus \Delta \to \mathbb{C}$$

a standard Calderón–Zygmund kernel on $\mathbb{R}^{d_i}$ if the size estimate

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

and, for some $\delta > 0$, the regularity estimates

$$|K_i(x_i, y_i) - K_i(x_i', y_i)| + |K_i(y_i, x_i) - K_i(y_i, x_i')| \leq C \frac{|x_i - x_i'|^\delta}{|x_i - y_i|^{d_i + \delta}}$$

whenever $|x_i - x_i'| \leq |x_i - y_i|/2$, are satisfied. The best constant in these estimates is denoted by $\|K\|_{\mathrm{CZ}(d_i,\delta)}$ and the collection of all such kernels is denoted as $\mathrm{CZ}(d_i, \delta)$.

1.3. Definition. Let $\Sigma_i = \Sigma(\mathbb{R}^{d_i})$ be the linear span of the indicator functions of cubes. A singular integral operator (SIO) is then a linear mapping $T_i \colon \Sigma_i \to L^1_{\mathrm{loc}}(\mathbb{R}^{d_i})$ such that

$$\langle T_i f, g \rangle = \int_{\mathbb{R}^{d_i}} \int_{\mathbb{R}^{d_i}} K(x, y) f(y) g(x) \, \mathrm{d}y \, \mathrm{d}x, \quad \mathrm{spt}(f) \cap \mathrm{spt}(g) = \emptyset, \quad f, g \in \Sigma_i,$$

where $K \in \mathrm{CZ}(d_i, \delta)$.

1.4. Definition. A Calderón–Zygmund operator (CZO) is simply an SIO T_i that is bounded from $L^p(\mathbb{R}^{d_i}) \to L^p(\mathbb{R}^{d_i})$ for all (equivalently, for some) $p \in (1,\infty)$. Given a CZO T_i with a kernel $K_i \in \mathrm{CZ}(d_i,\delta)$, let us denote $\|T\|_{\mathrm{CZO}(d_i,\delta)} = \|T\|_{L^2(\mathbb{R}^{d_i})\to L^2(\mathbb{R}^{d_i})} + \|K_i\|_{\mathrm{CZ}(d_i,\delta)}$.

1.2. Bi-parameter singular integrals. We give the definition of Martikainen [14] of bi-parameter SIOs, see also the last Section 5 for the original definition by Journé. Now we start working in the ambient space $\mathbb{R}^d = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$. Again, we let $\Sigma_i = \Sigma(\mathbb{R}^{d_i})$ be the linear span of the indicator functions of the cubes of $\mathbb{R}^{d_i}$ and then let $\Sigma = \Sigma(\mathbb{R}^d)$ be the linear span of $\Sigma_1\otimes\Sigma_2 = \{f_1\otimes f_2 \colon f_i\in\Sigma_i\}$. We assume that we are given a linear operator T along with a full adjoint T^* and partial adjoints T_1^*, T_2^*, i.e., four operators $T, T^*, T_1^*, T_2^* \colon \Sigma \to L^1_{\mathrm{loc}}(\mathbb{R}^d)$ that satisfy

$$\begin{aligned}\langle T(f_1\otimes f_2), g_1\otimes g_2\rangle &= \langle T_1^*(g_1\otimes f_2), g_1\otimes g_2\rangle = \langle T_2^*(f_1\otimes g_2), g_1\otimes f_2\rangle\\ &= \langle T^*(g_1\otimes g_2), f_1\otimes f_2\rangle.\end{aligned}$$

These operators will be assumed to have bi-parameter kernels, recalled next.

1.2.1. Bi-parameter kernels. Let $\delta>0$. We assume to have a kernel

$$K\colon \mathbb{R}^d\times\mathbb{R}^d\setminus\Delta\to\mathbb{C},$$

where $\Delta = \{(x,y)\in(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})^2 \colon x_1=y_1 \text{ or } x_2=y_2\}$, that satisfies the size estimate

$$|K(x,y)| \le C|x_1-y_1|^{-d_1}|x_2-y_2|^{-d_2}, \tag{1.5}$$

the regularity estimate

$$\begin{aligned}&|K(x,y)-K((x_1,x_2'),y)-K((x_1',x_2),y)+K(x',y)|\\ &\le C\frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{d_1+\delta}}\frac{|x_2-x_2'|^\delta}{|x_2-y_2|^{d_2+\delta}},\end{aligned}$$

whenever $|x_i-x_i'|\le\frac12|x_i-y_i|$ for $i=1,2$, and the mixed size-regularity estimate

$$|K((x_1,x_2),y)-K((x_1',x_2),y)| \le C\frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{d_1+\delta}}|x_2-y_2|^{-d_2},$$

whenever $|x_1-x_1'|\le\frac12|x_1-y_1|$. We also assume the symmetric estimates to the stated regularity and size-regularity estimates to hold in the other parameter slots. The collection of all such kernels is denoted $\mathrm{CZ}((d_1,d_2),\delta)$ and the best constant C in these estimates is denoted with $\|K\|_{\mathrm{CZ}((d_1,d_2),\delta)}$.

1.2.2. Full kernel representation. Let $f=f_1\otimes f_2, g=g_1\otimes g_2\in\Sigma$ be such that for both indices $i\in\{1,2\}$ we have $\mathrm{spt}(f_i)\cap\mathrm{spt}(g_i)=\emptyset$. Then we assume the representation

$$\langle Tf,g\rangle = \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}}\int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} K(x,y)(f_1\otimes f_2)(y)(g_1\otimes g_2)(x)\,\mathrm{d}y\,\mathrm{d}x,$$

where $K\in\mathrm{CZ}((d_1,d_2),\delta)$. Note that this implies the analogous kernel representations for T^{1*}, T^{2*}, T^*.

1.2.3. Partial kernel representations. Now, let $f=f_1\otimes f_2, g=g_1\otimes g_2\in\Sigma$ be such that for one index $j\in\{1,2\}$ we have $\mathrm{spt}(f_j)\cap\mathrm{spt}(g_j)=\emptyset$. Then, we assume the representation

$$\langle T(f_1\otimes f_2), g_1\otimes g_2\rangle = \int_{\mathbb{R}^{d_j}}\int_{\mathbb{R}^{d_j}} K_{f_i,g_i}(x_j,y_j)f_j(y_j)g_j(x_j)\,\mathrm{d}y_j\,\mathrm{d}x_j,$$

where $K_{f_i,g_i} \in \mathrm{CZ}(\delta, d_j)$ is such that $\|K_{f_i,g_i}\|_{\mathrm{CZ}(\delta,d_j)} \leq C(f_i, g_i)$ for some positive constant that depends on the functions f_i, g_i. We also assume these constants to have the following bounds

$$C(1_P, 1_P) + C(1_P, a_P) + C(a_P, 1_P) \leq C|P|$$

for all functions $a_P \in \Sigma_i$ such that $a_P = 1_P a_P$, $|a_P| \leq 1$, and $\int a_P = 0$, where P is a cube on $\mathbb{R}^{d_i}$.

1.6. Definition. A linear operator T with the full and partial kernel representations as described in this section, is called a bi-parameter singular integral operator.

1.7. Definition. A bi-parameter singular integral operator T such that

$$\|T\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} + \|T^{1*}\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} < \infty$$

for some $p \in (1, \infty)$ (equivalently, for all p) is called a bi-parameter Calderón–Zygmund operator.

1.3. Basic notation. We write all identities almost everywhere. For example, if a function can be made to satisfy a property (e.g. to be a constant, or continuous, etc.) by redefining it in a set of Lebesgue measure zero, we say that the function satisfies that property.

We denote cubes in $\mathbb{R}^{d_1}$ by I, and cubes in $\mathbb{R}^{d_2}$ by J—that is, the dimension of the cube can be read from which symbol we are using. Various rectangles then take the form $I \times J$. The side-length and the diameter of a cube I are denoted respectively by $\ell(I)$ and $\operatorname{diam}(I)$. Centre-points of cubes and rectangles are denoted as c_Q, c_R.

Often integral pairings need to be taken with respect to one of the variables only. For example, if $f\colon \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{C}$ and $h_I\colon \mathbb{R}^{d_1} \to \mathbb{C}$, then $\langle f, h_I\rangle\colon \mathbb{R}^{d_2} \to \mathbb{C}$ is defined by

$$\langle f, h_I\rangle(x_2) = \int_{\mathbb{R}^{d_1}} f(y_1, x_2) h(y_1)\,\mathrm{d}y_1.$$

On several occasions we use operators that only act on one of the variables, e.g. the maximal function $\mathsf{M}\colon L^p_{x_2} \to L^p_{x_2}$ and we denote it acting on a function of two parameters as $\mathsf{M}f(x_1, x_2) = \mathsf{M}(f(x_1, \cdot))(x_2)$. If it is unclear on what parameter slots these auxiliary operators are acting, we denote $\mathsf{M}^i, \mathsf{M}^{\alpha_i}$, etc.

Throughout the exponents p_1, p_2, q_1, q_2 will always be in the range $(1, \infty)$ but this will not always be mentioned. We will sometimes write $p = (p_1, p_2)$ and $q = (q_1, q_2)$ to shorten notation and this will be clear from the context.

We denote $A \lesssim B$, if $A \leq CB$ for some constant $C > 0$ depending only on the dimension of the underlying space, on the integrability exponents and on other unimportant absolute constants appearing in the assumptions. Then $A \sim B$, if $A \lesssim B$ and $B \lesssim A$. Subscripts on constants ($C_{a,b,c,\ldots}$) and quantifiers ($\lesssim_{a,b,c,\ldots}$) signify their dependence on those subscripts.

2. Approximate weak factorization in the bi-parameter setting

We will next go through the awf argument for proving commutator lower bounds in the bi-parameter setting. We refer the reader to consult [7] for a lengthier discussion in the standard one-parameter setting. Still, let us recall some important points.

When a commutator lower bound is proved, the full norm $\|[b, T]]\|_{L^{p_1}_{x_1} L^{p_2}_{x_2} \to L^{q_1}_{x_1} L^{q_2}_{x_2}}$ is not actually needed but so-called off-support versions of the norm we denote as

$\mathcal{O}_{p,q}(b;K)$ and $\mathcal{O}^{\Sigma}_{p,q}(b;K)$ are used and these can be defined even if we only have $b \in L^1_{\mathrm{loc}}$. Indeed, in defining these off-support norms what we use is the assumption

$$Tf(x) = \int_{\mathbb{R}^d} K(x,y)f(y)\,\mathrm{d}y, \quad x \notin \operatorname{spt}(f),$$

and this only involves the kernel. It is actually true in all cases that we are estimating the testing conditions on b by the off-support norms rather than with the full operator norm. Consequently, where we achieve a full characterization we also obtain as immediate corollaries the information

$$\mathcal{O}_{(p_1,p_2),(q_1,q_2)}(b;K) \sim \|[b,T]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{q_1}_{x_1}L^{q_2}_{x_2}}. \tag{2.1}$$

At the heart of the business lies the notion of non-degeneracy.

2.2. Definition. A bi-parameter kernel K is called non-degenerate, if for each $x=(x_1,x_2)\in\mathbb{R}^d$ and two radii $r_1,r_2>0$, there exists $y=(y_1,y_2)$ such that

$$|K((x_1,x_2),(y_1,y_2))| \gtrsim r_1^{-d_1}r_2^{-d_2}, \quad |x_1-y_1|>r_1, \quad |x_2-y_2|>r_2.$$

To obtain commutator lower bounds, we will also assume that the kernel K satisfies the size estimate (1.5) and the mixed size-regularity conditions

$$|K((x_1,x_2),y)-K((x_1',x_2),y)| \le C\frac{1}{|x_1-y_1|^{d_1}}\omega_1\left(\frac{|x_1-x_1'|}{|x_1-y_1|}\right)\frac{1}{|x_2-y_2|^{d_2}}, \tag{2.3}$$

whenever $|x_1-x_1'|\le 1/2|x_1-y_1|$, of which we also have the three other variants.

Notice that given the points x,y as in Definition 2.2, it follows from the size estimate that

$$r_1^{-d_1}r_2^{-d_2} \lesssim |K(x,y)| \lesssim |x_1-y_1|^{-d_1}|x_2-y_2|^{-d_2} \lesssim |x_1-y_1|^{-d_1}r_2^{-d_2}, \tag{2.4}$$

hence $|x_1-y_1|\lesssim r_1$, and similarly we see that $|x_2-y_2|\lesssim r_2$, and consequently that

$$|x_i-y_i|\sim r_i, \quad i=1,2. \tag{2.5}$$

Of the functions ω_i appearing in the mixed- and full regularity estimates we ask that they are increasing, subadditive and satisfy $\omega_i(\alpha)\to 0$ as $\alpha\to 0$. We will use a single function ω to deal with all the parameter slots, as we have $\omega_i \le \max_{i\in\{1,2,3,4\}}\omega_i =: \omega$, and ω is a function that satisfies the same assumptions as each single ω_i.

Obviously the class of standard bi-parameter CZ-kernels is encompassed here, however, it is a larger class in another sense also: we do not require any kind of full regularity conditions, see Section 1.2.

2.6. Proposition. *Let K be a non-degenerate bi-parameter kernel as in Definition 2.2 that satisfies the size estimate* (1.5) *and the mixed size-regularity estimates* (2.3). *Fix a constant $A\ge 3$ and let $R=I\times J$ be a rectangle. Then, there exists a rectangle $\widetilde{R}=\widetilde{I}\times\widetilde{J}$ such that $\ell(\widetilde{I})=\ell(I)$ and $\ell(\widetilde{J})=\ell(J)$, and that is localized as*

$$\operatorname{dist}(I,\widetilde{I})\sim A\operatorname{diam}(I), \quad \operatorname{dist}(J,\widetilde{J})\sim A\operatorname{diam}(J) \tag{2.7}$$

and which satisfies the following: for all $x\in R$ and $y\in\widetilde{R}$ we have that

$$|K\big(x,y\big)-K\big(c_R,c_{\widetilde{R}}\big)| \lesssim A^{-(d_1+d_2)}|R|^{-1}\omega\big(1/A\big), \tag{2.8}$$

and if we choose A large enough, we also have,

$$\left|\int_R K(x,y)\,\mathrm{d}x\right| \sim \left|\int_{\widetilde{R}} K(x,y)\,\mathrm{d}y\right| \sim \int_R |K(x,y)|\,\mathrm{d}x \sim \int_{\widetilde{R}}|K(x,y)|\,\mathrm{d}y \sim A^{-(d_1+d_2)}.$$

Proof. Let $c_R = (c_I, c_J) \in \mathbb{R}^{d_1+d_2}$ be the centre of a rectangle R. By the non-degeneracy of K, we find a point $c_{\widetilde{R}} = (c_{\widetilde{I}}, c_{\widetilde{J}})$ such that

$$|c_I - c_{\widetilde{I}}| \geq A\ell(I), \quad |c_J - c_{\widetilde{J}}| \geq A\ell(J) \tag{2.9}$$

that is the centre of a rectangle $\widetilde{R} = \widetilde{I} \times \widetilde{J}$ and satisfies

$$|K(c_R, c_{\widetilde{R}})| \gtrsim A^{-(d_1+d_2)}\ell(I)^{-d_1}\ell(J)^{-d_2} = A^{-(d_1+d_2)}|R|^{-1}.$$

The claims on the line (2.7) follow immediately from the remarks following the definition of non-degeneracy, see the lines (2.4) and (2.5). Moreover, by the size estimate and (2.9) we have that $|K(c_R, c_{\widetilde{R}})| \lesssim A^{-(d_1+d_2)}|R|^{-1}$ and consequently that

$$|K(c_R, c_{\widetilde{R}})| \sim A^{-(d_1+d_2)}|R|^{-1}. \tag{2.10}$$

Now let $x \in R$ and $y \in \widetilde{R}$ be arbitrary. To see why (2.8) holds, we use the mixed size-regularity conditions (2.3). We have

$$\begin{aligned}
|K\big(x,y\big) - K\big(c_R, c_{\widetilde{R}}\big)| &\leq |K\big([x_1,x_2],[y_1,y_2]\big) - K\big([c_I,x_2],[y_1,y_2]\big)| \\
&\quad + |K\big([c_I,x_2],[y_1,y_2]\big) - K\big([c_I,c_J],[y_1,y_2]\big)| \\
&\quad + |K\big([c_I,c_J],[y_1,y_2]\big) - K\big([c_I,c_J],[c_{\widetilde{I}},y_2]\big)| \\
&\quad + |K\big([c_I,c_J],[c_{\widetilde{I}},y_2]\big) - K\big([c_I,c_J],[c_{\widetilde{I}},c_{\widetilde{J}}]\big)| \\
&\lesssim A^{-(d_1+d_2)}|R|^{-1}\omega\big(1/A\big),
\end{aligned}$$

where for example the estimate for the first of the four intermediate terms derives as

$$\begin{aligned}
|K([x_1,x_2],[y_1,y_2]) - K([c_I,x_2],[y_1,y_2])| &\lesssim \frac{1}{|c_I - y_1|^{d_1}}\omega\left(\frac{|x_1 - c_I|}{|c_I - y_1|}\right)\frac{1}{|x_2 - y_2|^{d_2}} \\
&\lesssim A^{-d_1}\ell(I)^{-d_1}\omega\big(C/A\big)A^{-d_2}\ell(J)^{-d_2} \\
&\lesssim A^{-(d_1+d_2)}|R|^{-1}\omega\big(1/A\big),
\end{aligned}$$

where used the fact that $A \geq 3$ to apply the mixed size-regularity estimates, and also the sub-additivity of ω.

Now, the last four claims involving the integrals follow by choosing A sufficiently large, by subtracting and adding $K(c_R, c_{\widetilde{R}})$, and using the estimates (2.8) and (2.10). □

2.11. Proposition. *Let K be a non-degenarate bi-parameter kernel as in Proposition 2.6. Let $R = I \times J$ be a fixed rectangle and let f be a locally integrable function such that $\operatorname{spt}(f) \subset R$, $\int f = 0$. Then, for a choice of the constant A large enough, the function f can be written as*

$$f = [h_1 T g_1 - g_1 T^* h_1] + [h_2 T^* g_2 - g_2 T h_2] + \tilde{f}, \tag{2.12}$$

where the appearing auxiliary functions satisfy

$$g_1 = 1_{\widetilde{R}}, \quad g_2 = 1_R, \quad \operatorname{spt}(h_1) \subset R, \quad \operatorname{spt}(h_2) \subset \widetilde{R}, \quad \operatorname{spt}(\tilde{f}) \subset R \tag{2.13}$$

and

$$|h_1(x)| \lesssim A^d|f(x)|, \quad |h_2(x)| \lesssim A^d\langle|f|\rangle_R 1_{\widetilde{R}}(x), \quad |\tilde{f}(x)| \lesssim \omega(\frac{1}{A})\langle|f|\rangle_R 1_R(x), \tag{2.14}$$

and we have $\int \tilde{f} = 0$.

Proof. Let $\widetilde{R} = \widetilde{I} \times \widetilde{J}$ be the rectangle as obtained by Proposition 2.6 and let $g_1 := 1_{\widetilde{R}}$. We decompose the function f as

$$f = h_1 T g_1 - g_1 T^* h_1 + \tilde{f}, \quad h_1 = \frac{f}{T g_1}, \quad \tilde{w} = g_1 T^* h_1.$$

The only problem with the above factorization is that h_1 might a priori involve a division by zero, the following estimates show that this is not the case. Let $x \in R$, then

$$T g_1(x) = \int_{\widetilde{R}} K(x,y)\,\mathrm{d}y = \int_{\widetilde{R}} (K(x,y) - K(c_R, c_{\widetilde{R}}))\,\mathrm{d}y + \int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}y = I + II.$$

It follows by Proposition 2.6 that

$$|I| \lesssim A^{-(d_1+d_2)}\omega(1/A), \quad |II| \sim A^{-(d_1+d_2)}$$

and hence for A sufficiently large that $|T g_1(x)| \sim A^{-(d_1+d_2)}$, making h_1 well-defined. Also, by the above we have

$$|h_1(x)| \lesssim A^{d_1+d_2}|f(x)|,$$

which establishes the left-most estimate on the line (2.14). Then, to estimate the first iteration error term $\tilde{w}$, let $y \in R$ and write

$$\frac{f}{T g_1}(y) = \left(\frac{f}{T g_1} - \frac{f}{\int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}z}\right)(y) + \frac{f(y)}{\int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}z} = III + IV.$$

By Proposition 2.6 it follows that

$$\begin{aligned} |III| &= \left| f(y) \int_{\widetilde{R}} (K(y,z) - K(c_R, c_{\widetilde{R}}))\,\mathrm{d}z \right| \times \left| \int_{\widetilde{R}} K(y,z)\,\mathrm{d}z \int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}z \right|^{-1} \\ &\lesssim \frac{|f(y)| A^{-d}\omega(\frac{1}{A})}{A^{-d}A^{-d}} = A^d \omega\left(\frac{1}{A}\right)|f(y)|, \end{aligned}$$

and hence, we have with $x \in \widetilde{R}$ that

$$\begin{aligned} &\left| T^* \left(\frac{f}{T g_1} - \frac{f}{\int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}y} \right) \right| (x) \lesssim A^d \omega\left(\frac{1}{A}\right) \cdot \int_R |K(y,x)||f(y)|\,\mathrm{d}y \\ &\lesssim A^d \omega\left(\frac{1}{A}\right) \cdot A^{-d} \langle |f| \rangle_R 1_{\widetilde{R}}(x) = \omega\left(\frac{1}{A}\right) \cdot \langle |f| \rangle_R 1_{\widetilde{R}}(x), \end{aligned}$$

where we simply used the size estimate.

By the zero mean of f on R we have

$$\begin{aligned} \left| T^*(IV)(y) \right| &= \left| T^* \left(\frac{f}{\int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}y} \right)(x) \right| = \frac{\left| \int_R (K(y,x) - K(c_{\widetilde{R}}, c_R) f(y)\,\mathrm{d}y \right|}{\left| \int_{\widetilde{R}} K(c_R, c_{\widetilde{R}})\,\mathrm{d}y \right|} \\ &\lesssim \omega\Big(\frac{1}{A}\Big) A^{-d} \langle |f| \rangle_R 1_{\widetilde{R}}(x) \cdot A^d = \omega\Big(\frac{1}{A}\Big) \cdot \langle |f| \rangle_R 1_{\widetilde{R}}(x), \end{aligned}$$

where we used the mixed size-regularity estimates and Proposition 2.6. Hence, combining the above parts, we obtain

$$|\tilde{w}(x)| \lesssim \omega\Big(\frac{1}{A}\Big) \langle |f| \rangle_R 1_{\widetilde{R}}(x).$$

It is also immediate from the definitions that

$$\int_{\widetilde{R}} \tilde{w} = \int g_1 T^* \left(\frac{f}{T g_1} \right) = \int T g_1 \frac{f}{T g_1} = \int f = 0. \tag{2.15}$$

Now, let $g_2 = 1_R$. By repeating the above argument, but now starting with the function $\tilde{f}$ supported on the rectangle $\widetilde{R}$ we write

$$\tilde{w} = h_2 T^* g_2 - g_2 T h_2 + \tilde{\tilde{f}}, \quad h_2 = \frac{\tilde{w}}{T g_2}, \quad \tilde{f} = g_2 T h_2.$$

With the same arguments and proofs as above, the function h_2 is well-defined and for $x \in \widetilde{R}$ we have that

$$|h_2(x)| \lesssim A^d |\tilde{f}(x)| \lesssim A^d \omega\Big(\frac{1}{A}\Big) \langle |f| \rangle_R 1_{\widetilde{R}}(x) \lesssim A^d \langle |f| \rangle_R 1_{\widetilde{R}}(x)$$

and for $x \in R$, with A large enough, that

$$|\tilde{f}(x)| \lesssim \omega\Big(\frac{1}{A}\Big) \cdot \langle |\tilde{f}| \rangle_{\widetilde{R}} 1_R(x) \lesssim \omega\Big(\frac{1}{A}\Big)^2 \langle |f| \rangle_R 1_R(x) \lesssim \omega\Big(\frac{1}{A}\Big) \langle |f| \rangle_R 1_R(x).$$

Moreover, as in (2.15), the secont iteration error term $\widetilde{f}$ inherits the zero mean from $\widetilde{w}$. □

Let us notate the oscillation of a function $b \in L^1_{\mathrm{loc}}$ over a rectangle $R = I \times J$ with

$$\operatorname{osc}(b; R) = \fint_R |b - \langle b \rangle_R|.$$

2.16. Proposition. *Let K be a symmetrically non-degenerate bi-parameter kernel and $b \in L^1_{\mathrm{loc}}$. Then, for all rectangles $R = I \times J$ we have*

$$|R| \operatorname{osc}(b; R) \lesssim |\langle [b, T] g_1, h_1 \rangle| + |\langle [b, T] h_2, g_2 \rangle|,$$

where the appearing functions are as in 2.11,

$$g_1 = 1_{\widetilde{R}}, \quad g_2 = 1_R, \quad |h_1(x)| \lesssim_A 1_R(x), \quad |h_2(x)| \lesssim_A 1_{\widetilde{R}}(x).$$

Proof. As $b \in L^1_{\mathrm{loc}}$, we find a function f of zero mean supported on R such that $\|f\|_\infty \le 1$ and

$$|R| \operatorname{osc}(b; R) \sim \int bf.$$

By Proposition 2.11 we write and estimate the right-hand side as

$$\begin{aligned} \int bf &= \int b[h_1 T g_1 - g_1 T^* h_1] + \int b[h_2 T^* g_2 - g_2 T h_2] + \int b\tilde{f} \\ &\le |\langle [b, T] g_1, h_1 \rangle| + |\langle [b, T] h_2, g_2 \rangle| + \left| \int b\tilde{f} \right| \end{aligned}$$

and the error term further to

$$\left| \int b\tilde{f} \right| \le \|\tilde{f}\|_\infty \int_R \big| b - \langle b \rangle_R \big| \lesssim \omega(1/A) |R| \operatorname{osc}(b; R).$$

By having the above estimates together we obtain

$$|R| \operatorname{osc}(b; R) \lesssim |\langle [b, T] g_1, h_1 \rangle| + |\langle [b, T] h_2, g_2 \rangle| + \omega(1/A) |R| \operatorname{osc}(b; R). \tag{2.17}$$

As $b \in L^1_{\mathrm{loc}}$, by choosing A large enough, we absorb the common term in (2.17) to the left-hand side. □

The first off-support norm we use is

2.18. Definition. Let $b \in L^1_{\mathrm{loc}}$ and define

$$\mathcal{O}^A_{(t_1,t_2),(s_1,s_2)}(b;K) = \sup_{\substack{R=I\times J\\ \widetilde{R}=\widetilde{I}\times\widetilde{J}}} \left| \frac{\int_{\mathbb{R}^d\times\mathbb{R}^d}(b(x)-b(y))K(x,y)f(y)g(x)\,\mathrm{d}y\,\mathrm{d}x}{|I|^{1/t_1+1/s_1'}|J|^{1/t_2+1/s_2'}} \right|,$$

where the supremum is taken over rectangles $R = I \times J$ and $\widetilde{R} = \widetilde{I} \times \widetilde{J}$ with

$$\operatorname{dist}(I,\widetilde{I}) \sim A\ell(I) \quad \text{and} \quad \operatorname{dist}(J,\widetilde{J}) \sim A\ell(J)$$

and over functions $f \in L^\infty(R)$ and $g \in L^\infty(\widetilde{R})$ with

$$\|f\|_{L^\infty} \le 1 \quad \text{and} \quad \|g\|_{L^\infty} \le 1.$$

2.19. Remark. When $p = (p_1,p_2), q = (q_1,q_2)$ we may write $\mathcal{O}^A_{p,q}(b;K) = \mathcal{O}^A_{(p_1,p_2),(q_1,q_2)}(b;K)$. From this point onwards we will fix the constant A large enough so that we may always use the conclusions of all the above stated propositions where the constant A appears and we will drop the superscript A and simply write $\mathcal{O}_{p,q}(b;K)$.

Relating the oscillation to the off-support norm, we have the following

2.20. Proposition. *Let K be a non-degenerate bi-parameter kernel, $b \in L^1_{\mathrm{loc}}$ and $s_i, t_i \in (1,\infty)$. Then, for all rectangles $R = I \times J$ we have*

$$\operatorname{osc}(b;R) \lesssim \mathcal{O}_{(t_1,t_2),(s_1,s_2)}(b;K)|I|^{1/t_1-1/s_1}|J|^{1/t_2-1/s_2}.$$

Proof. By Proposition 2.16 we write

$$|R|\operatorname{osc}(b;R) \lesssim |\langle [b,T]g_1,h_1\rangle| + |\langle [b,T]h_2,g_2\rangle|,$$

for functions h_i, g_i as in Proposition 2.16. By the definition of the off-support norm we estimate

$$\begin{aligned}|\langle [b,T]g_1,h_1\rangle| &\le \mathcal{O}_{(t_1,t_2)(s_1,s_2)}|I|^{1/t_1+1/s_1'}|J|^{1/t_2+1/s_2'}\\ &= |R|\mathcal{O}_{(t_1,t_2),(s_1,s_2)}(b;K)|I|^{1/t_1-1/s_1}|J|^{1/t_2-1/s_2}\end{aligned}$$

and similarly for the other term. Dividing with $|R|$, the claimed estimate follows. □

3. Upper and lower bounds

In this section we prove all the stated lower bounds and those upper bounds that admit a short proof, with the remaining upper bounds postponed to Sections 4 and 5.

3.1. The case $p_i = q_i > 1$, $i = 1,2$. This case is not new, other proofs are contained e.g. in [6] and [12] both that treat the problem in the Bloom setup. Given the awf argument prepared in the previous section, the arguments are shortly stated and we gather them here in the unweighted off-diagonal setting in Proposition 3.1. Also, when in addition all the exponents are the same, we record how to derive the bi-parameter Bloom type lower bound directly from Proposition 2.20. Proposition 3.2 is not new either, the special case with the Riesz transforms is contained in [6] and the result with the same non-degeneracy assumptions as we use is in [12]. In [12] the slightly simpler median method is used which limits their considerations to real-valued functions b, on the other hand the median method works also for iterated commutator.

3.1. Proposition. *Let $1 < p_i = q_i < \infty$, $i = 1, 2$ and assume that $b \in L^1_{\mathrm{loc}}$. Then,*

$$\|b\|_{\mathrm{bmo}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})} \lesssim \mathcal{O}_{(p_1,p_2),(p_1,p_2)}(b;K) \le \|[b,T]]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{p_1}_{x_1}L^{p_2}_{x_2}} \lesssim \|b\|_{\mathrm{bmo}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}$$

Proof. The first estimate is immediate from 2.20, while the second follows by a simple application of Hölders' inequality. Hence, the only claim left to show is the upper bound

$$\|[b,T]]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{p_1}_{x_1}L^{p_2}_{x_2}} \lesssim \|b\|_{\mathrm{bmo}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}.$$

This, is proved with exactly the same argument as the commutator upper bounds are proved in [12], the fact that we have mixed norms appear, contrary to the non-mixed cases, plays no significant role in the proof at all. □

Let us then turn to the Bloom type lower bound. Recall that a positive function μ is in the bi-parameter A_p if

$$[\mu]_{A_p(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})} := \sup_R \langle\mu\rangle_R \langle\mu^{-\frac{p'}{p}}\rangle_R^{\frac{p}{p'}} < \infty,$$

and for a positive locally integral function ν we write $b \in \mathrm{bmo}_\nu$, if

$$\|b\|_{\mathrm{bmo}_\nu} := \sup_R \frac{1}{\nu(R)} \int_R |b - \langle b\rangle_R| < \infty, \quad \nu(R) = \int_R \nu.$$

Notice that if we have two weights $\lambda, \mu \in A_p$, then by a simple application of Hölder's inequality we have that $\nu = (\mu/\lambda)^{1/p} \in A_2$.

Also in the Bloom case, we use an off-support norm. The only difference compared to Definition 2.18 is that now the normalization is modified and we consider the quantity

$$\mathcal{O}_{(p,p)}(b;K;\mu;\lambda) = \sup \left| \frac{\int_{\mathbb{R}^d\times\mathbb{R}^d} (b(x)-b(y))K(x,y)f(y)g(x)\,\mathrm{d}y\,\mathrm{d}x}{\mu(R)^{1/p}[\lambda^{-\frac{p'}{p}}(\widetilde{R})]^{1/p'}} \right|$$

where the supremum is taken over all such functions f and g as in the Definition 2.18.

3.2. Proposition. *Let $b \in L^1_{\mathrm{loc}}$ and let μ, λ be bi-parameter A_p weights. Then, we have that*

$$\|b\|_{\mathrm{bmo}(\nu)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \mathcal{O}^A_{(p,p)}(b;K;\mu;\lambda) \le \|[b,T]]\|_{L^p(\mu)\to L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b\|_{\mathrm{bmo}(\nu)},$$

where $\nu := (\frac{\mu}{\lambda})^{1/p}$.

Proof. By Proposition 2.16 we estimate

$$\begin{aligned}\int_R |b - \langle b\rangle_R| = |R|\operatorname{osc}(b;R) &\lesssim \sum_{i=1,2} |\langle [b,T]g_i, h_i\rangle| \\ &\le \mathcal{O}_{(p,p)}(b;K;\mu;\lambda)\big(\mu(R)^{1/p}[\lambda^{-\frac{p'}{p}}(\widetilde{R})]^{1/p'} + \mu(\widetilde{R})^{1/p}[\lambda^{-\frac{p'}{p}}(R)]^{1/p'}\big).\end{aligned}$$

Since A_p weights are doubling and

$$\operatorname{dist}(I,\widetilde{I}) \sim \operatorname{diam}(I), \quad \operatorname{dist}(J,\widetilde{J}) \sim \operatorname{diam}(J),$$

it follows that

$$\mu(\widetilde{R}) \sim_{[\mu]_{A_p}} \mu(R), \quad \lambda(\widetilde{R}) \sim_{[\lambda]_{A_p}} \lambda(R), \quad \nu(\widetilde{R}) \sim_{[\mu]_{A_p},[\lambda]_{A_p}} \nu(R). \tag{3.3}$$

Hence, we estimate the left-term of the previous estimate with the index $i = 1$ as

$$\begin{aligned}\mu(\widetilde{R})^{1/p}[\lambda^{-\frac{p'}{p}}(R)]^{1/p'} &= \mu(\widetilde{R})^{1/p}\left(\int_R \lambda^{-\frac{p'}{p}}\right)^{1/p'} \sim_{A,[\lambda]_{A_p}} \mu(R)^{1/p}\left(\int_R \lambda^{-\frac{p'}{p}}\right)^{1/p'} \\ &\overset{*}{\leq} [\mu]_{A_p}^{1/p}\langle\nu\rangle_R\lambda(R)^{1/p}\left(\int_R \lambda^{-\frac{p'}{p}}\right)^{1/p'} \leq [\mu]_{A_p}^{1/p}[\lambda]_{A_p}^{1/p}\nu(R) \\ &\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \nu(R)\end{aligned}$$

where in the estimate marked with $*$ we used that

$$\mu(R)^{1/p}\lambda(R)^{-1/p} \leq [\mu]_{A_p}^{1/p}\langle\nu\rangle_R,$$

which follows by a few applications of Hölder's inequality and a rearranging of the estimate

$$1 \leq \langle\nu\rangle_R\langle\nu^{-1}\rangle_R \leq \langle\nu\rangle_R\langle\lambda\rangle_R^{1/p}\langle\mu^{-\frac{p'}{p}}\rangle_R^{1/p'} \leq \langle\nu\rangle_R\langle\lambda\rangle_R^{1/p}\langle\mu\rangle_R^{-1/p}[\mu]_{A_p}^{1/p}.$$

Using the other estimates from the line (3.3) it follows that $\mu(R)^{1/p}[\lambda^{-\frac{p'}{p}}(\widetilde{R})]^{1/p'}$ satisfies the same estimate, and hence, we have shown the first estimate,

$$\int_R |b - \langle b\rangle_R| \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \mathcal{O}_{(p,p)}(b;K;\mu;\lambda)\nu(R).$$

For the middle estimate, by Hölder's inequality we immediately have that

$$\mathcal{O}^A_{(p,p)}(b;K;\mu;\lambda) \leq \|[b,T]\|_{L^p(\mu)\to L^p(\lambda)},$$

and the right-most estimate $\|[b,T]\|_{L^p(\mu)\to L^p(\lambda)} \lesssim \|b\|_{\mathrm{bmo}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}$ is proved in exactly the stated form in [12]. □

3.2. The three cases $p_i < q_i$, $i = 1, 2$ and $p_1 < q_1$, $p_2 > q_2$, and $p_2 < q_2$, $p_1 > q_1$. In these three cases we find that the commutator is bounded if and only if b is a constant function almost everywhere. By redefining b in a set of measure zero we may assume that b is a constant.

3.4. Proposition. *Let $b \in L^1_{\mathrm{loc}}$, $p_i < q_i$, $i = 1, 2$, and assume that $\mathcal{O}_{p,q}(b;K) < \infty$. Then, b is a constant. Conversely, if b is a constant, then $[b,T] = 0$.*

Proof. Only one direction is non-trivial. Fix a point $x_2 \in \mathbb{R}^d$ and consider a sequence of cubes $\mathbb{R}^{d_2} \supset J_k \to \{x_2\}$. The Lebesgue differentiation theorem shows that

$$⨍_I |b(x_1,x_2) - \langle b(\cdot,x_2)\rangle_I|\,\mathrm{d}x_1 = \lim_{k\to\infty}\operatorname{osc}(b;I\times J_k)$$

for almost every $x_1 \in \mathbb{R}^{d_1}$. By Proposition 2.20 we dominate the right-hand side with

$$\mathcal{O}_{p,q}(b;K)|I|^{1/p_1-1/q_1}\lim_{k\to\infty}|J_k|^{1/p_2-1/q_2} = 0,$$

where in the last step we used that $1/p_2 - 1/q_2 > 0$. This shows that $b(\cdot,x_2)$ is a constant on all cubes $I \subset \mathbb{R}^{d_1}$, hence on $\mathbb{R}^{d_1}$. Similarly we see that $b(x_1,\cdot)$ is a constant almost everywhere on $\mathbb{R}^{d_2}$. It follows that b is a constant almost everywhere. □

3.5. Proposition. *Let $p_1 < q_1$, $p_2 > q_2$ and assume that $\mathcal{O}_{p,q}(b;K) < \infty$. Then, b is a constant. Conversely, if b is a constant, then $[b,T] = 0$.*

Proof. By the same argument as above we see that $b(x_1,\cdot)$ is a constant and hence for any choice of a cube $J\subset\mathbb{R}^{d_2}$ we have that

$$\fint_I |b(x_1,x_2)-\langle b(\cdot,x_2)\rangle_I|\,\mathrm{d}x_1 = \fint_I\fint_J |b-\langle b\rangle_{I\times J}| \lesssim \mathcal{O}_{p,q}(b;K)|I|^{1/p_1-1/q_1}|J|^{1/p_2-1/q_2}.$$

As $1/p_2-1/q_2<0$, letting $|J|\to\infty$ shows that

$$\fint_I |b(x_1,x_2)-\langle b(\cdot,x_2)\rangle_I|\,\mathrm{d}x_1 = 0.$$

Hence, also $b(\cdot,x_2)$ is a constant and consequently b is a constant. □

The symmetric case with a symmetric proof is

3.6. Proposition. *Let $p_1>q_1$, $p_2<q_2$ and assume that $\mathcal{O}_{p,q}(b;K)<\infty$. Then, b is a constant. Conversely, if b is a constant, then $[b,T]=0$.*

3.3. The cases $p_1<q_1$, $p_2=q_2$, and $p_2<q_2$, $p_1=q_1$. In these cases the function b is constant in one variable slot and Hölder continuous in the other.

3.7. Proposition. *Let $b\in L^1_{\mathrm{loc}}$, $p_1<q_1$ and $p_2=q_2$ and $\mathcal{O}_{p,q}(b;K)<\infty$. Then, $b(x_1,\cdot)$ is a constant and there holds that*

$$\|b(\cdot,x_2)\|_{\dot{C}^{0,\alpha_1}_{x_1}} \lesssim \mathcal{O}_{p,q}(b;K). \tag{3.8}$$

Conversely, if b satisfies the above properties, then

$$\|[b,T]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{q_1}_{x_1}L^{p_2}_{x_2}} \lesssim \|b(\cdot,x_2)\|_{\dot{C}^{0,\alpha_1}_{x_1}}.$$

Proof. We see by the same argument as above that $b(x_1,\cdot)$ is a constant for almost every $x_1\in\mathbb{R}^{d_1}$, and by redefining in a set of measure zero, constant everywhere. Thus, for every $x_2\in\mathbb{R}^{d_2}$ there holds that

$$\fint_I |b(x_1,x_2)-\big\langle b(\cdot,x_2)\big\rangle_I|\,\mathrm{d}x_1 = \fint_I\fint_J |b-\langle b\rangle_{I\times J}| \lesssim \mathcal{O}_{p,q}(b;K)|I|^{1/p_1-1/q_1}|J|^{1/p_2-1/p_2} = \mathcal{O}_{p,q}(b;K)|I|^{1/p_1-1/q_1}$$

and this implies (3.8). The converse direction is proved in Proposition 5.5 of Section 4. □

The symmetric case with a symmetric proof is

3.9. Proposition. *Let $p_2<q_2$ and $p_1=q_1$ and assume that $\mathcal{O}_{p,q}(b;K)<\infty$. Then, $b(\cdot,x_2)$ is a constant and*

$$\|b(x_1,\cdot)\|_{\dot{C}^{0,\alpha_2}_{x_2}} \lesssim \mathcal{O}_{p,q}(b;K).$$

Conversely, if the above conclusions hold, then

$$\|[b,T]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{p_1}_{x_1}L^{q_2}_{x_2}} \lesssim \operatorname*{ess\,sup}_{x_1\in\mathbb{R}^{d_1}} \|b(x_1,\cdot)\|_{\dot{C}^{0,\alpha_2}_{x_2}}.$$

Proof. The proof for the first part of the claim is completely symmetric with proof in the previous case. The converse direction is proved in Section 5, see Proposition 4.3. □

Recapping, in all the above cases where we concluded the function b to be a constant, we have the corresponding upper bounds as stated in Theorem 1.2 (i.e. $b=$ constant implies that $[b,T]=0$ which implies that $N_{p,q}=0$) and hence have constituted a full characterization of the boundedness of $[b,T]$, in these cases. Both

upper bounds for the cases where we concluded the function b to be constant in one and have the Hölder continuity criterion in the other variable are lengthier and will be presented later in sections 4 and 5.

3.4. The case $p_1 = q_1$ and $p_2 > q_2$. We first recall some basic background. A dyadic grid on $\mathbb{R}^d$ is a collection $\mathcal{D} = \mathcal{D}(\mathbb{R}^d)$ of cubes with side-lengths in the powers of two such that:

(1) for each $k \in \mathbb{Z}$ the collection $\{Q \in \mathcal{D} \colon \ell(Q) = 2^k\}$ is a disjoint cover of $\mathbb{R}^d$,
(2) for $Q, P \in \mathcal{D}$ there holds that $Q \cap P \in \{Q, P, \emptyset\}$.

Given a cube Q, we let $\mathcal{D}(Q)$ denote the system of dyadic cubes inside Q that is attained from iteratively bisecting the sides of Q; we use sparse collections made up of elements of $\mathcal{D}(Q)$. A collection of sets $\mathscr{S}$ is said to be γ-sparse, if each $Q \in \mathscr{S}$ has a major subset E_Q such that $|E_Q| > \gamma|Q|$ and these sets E_Q are pairwise disjoint.

The stopping time family inside a fixed cube Q_0 is given by the following algorithm. For a given cube $Q \in \mathcal{D}$ we denote

$$S(f;Q) = \{P \in \mathcal{D},\ P \subset Q \text{ is maximal with } \langle |f| \rangle_P > 2\langle |f| \rangle_Q\}$$

and let

$$\mathscr{S} = \bigcup_k \mathscr{S}_k, \quad \mathscr{S}_{k+1} = \bigcup_{P \in \mathscr{S}_k} S(f;P), \quad \mathscr{S}_0 = \{Q_0\}.$$

For a given collection $\mathscr{S} \subset \mathcal{D}$ of dyadic cubes and for each $Q \in \mathscr{S}$ we let $\mathrm{ch}_{\mathscr{S}}(Q)$ consist of the maximal cubes $P \in \mathscr{S}$ such that $P \subsetneq Q$. For a given cube $P \in \mathscr{S}$ we denote $E_P = P \setminus \bigcup_{Q \in \mathrm{ch}_{\mathscr{S}} P} Q$, and for each $P \in \mathcal{D}$ we let $\Pi P := \Pi_{\mathscr{S}} P$ denote the minimal cube Q in $\mathscr{S}$ such that $P \subset Q$ (on the condition that it exists). With this notation then,

$$\mathrm{ch}_{\mathscr{S}}(P) = \{Q \in \mathscr{S} \colon Q \subsetneq P,\ \Pi Q = P\}.$$

A variant of the following lemma is contained in [7].

3.10. Lemma. *Fix a cube Q and let f be a bounded function of zero mean supported on Q. Then, there exists a sparse collection $\mathscr{S} = \bigcup_{k=1}^N \mathscr{S}_k \subset \mathcal{D}_Q$ such that*

$$f = \sum_{k=0}^N \sum_{P \in \mathscr{S}_k} f_P, \quad f_P = \sum_{\Pi_{\mathscr{S}} Q = P} \Delta_Q f,$$

where the number N is finite and depends only on $\|f\|_{L^\infty(Q)}$, and moreover, there holds that

(1) $\int f_P = 0$, *for all* $P \in \mathscr{S}$,
(2) $\sum_{k=0}^N \sum_{P \in \mathscr{S}_k} \|f_P\|_\infty^s 1_P \lesssim_s (\mathsf{M}f)^s$ *for all* $s > 0$.

In the remaining lower bounds we use the off-support norm given in the following

3.11. Definition. We let

$$\mathcal{O}_{p,q}^{\Sigma,A}(b;K) = \sup \frac{\sum_{i=1}^N \left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} (b(x) - b(y)) K(x,y) f_i(y) g_i(x) \,\mathrm{d}y \,\mathrm{d}x \right|}{\| \sum_{i=1}^N \|f_i\|_\infty 1_{R_i} \|_{L^{p_1}_{x_1} L^{p_2}_{x_2}} \| \sum_{i=1}^N \|g_i\|_\infty 1_{\widetilde{R}_i} \|_{L^{q_1'}_{x_1} L^{q_2'}_{x_2}}},$$

where the supremum is taken over rectangles $R_i = I_i \times J_i$ and $\widetilde{R}_i = \widetilde{I}_i \times \widetilde{J}_i$ with

$$\ell(L_i) = \ell(\widetilde{L}_i) \quad \text{and} \quad \operatorname{dist}(L_i, \widetilde{L}_i) \sim A\ell(L_i) \quad \text{for} \quad L = I, J$$

and over functions $f_i \in L^\infty(R_i)$, $g_i \in L^\infty(\widetilde{R}_i)$, $i = 1, \dots, N$.

3.12. Remark. Again, we will suppress the superscript A from $\mathcal{O}_{p,q}^{\Sigma,A}$ and just write $\mathcal{O}_{p,q}^{\Sigma}$. Using that for linear operators U there holds

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

for Rademacher random signs ε_i, it follows by Hölder's inequality that

$$\mathcal{O}_{p,q}^{\Sigma}(b;K) \leq \|[b,T]\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}\to L_{x_1}^{q_1}L_{x_2}^{q_2}}.$$

Consequently, this is a reasonable off-support constant.

3.13. Proposition. *Let $p_1 = q_1$, $q_2 < p_2$ and set $1/q_2 = 1/r_2 + 1/p_2$, and assume that $b \in L^1_{\mathrm{loc},x_1}L^{r_2}_{\mathrm{loc},x_2}$. Then, there holds that*

$$\inf_{c\in\mathbb{C}}\|b-c\|_{L_{x_1}^{\infty}L_{x_2}^{r_2}} \lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K) \leq \|[b,T]\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}\to L_{x_1}^{p_1}L_{x_2}^{q_2}} \lesssim \inf_{c\in\mathbb{C}}\|b-c\|_{L_{x_1}^{\infty}L_{x_2}^{r_2}}. \tag{3.14}$$

Proof. Let $c \in \mathbb{C}$ be a constant and denote $\widetilde{b}(x_1,x_2) = b(x_1,x_2) - c$. Then, let $f\colon \mathbb{R}^{d_2} \to \mathbb{C}$ be such that

$$1_J f = f, \quad \int f = 0, \quad \|f\|_{L^{r_2'}(\mathbb{R}^{d_2})} \leq 1. \tag{3.15}$$

Then, according to Lemma 3.10, we let $\mathscr{S}^2$ be the sparse collection of cubes inside J with respect to the function f and with $R = I \times J$ write

$$\int_R \widetilde{b}f = \int \widetilde{b}\cdot 1_I \otimes \left(\sum_{k=0}^{N}\sum_{P\in\mathscr{S}_k^2} f_P\right) = \sum_{k=0}^{N}\sum_{P\in\mathscr{S}_k^2}\int \widetilde{b}\cdot 1_I\otimes f_P.$$

The last step follows from that the left-hand side is integrable and that $\sum_{P\in\mathscr{S}_k^2}\widetilde{b}\cdot 1_I\otimes f_P$ are disjointly supported for each fixed k. Then, as the functions $1_I\otimes f_P$ satisfy the assumptions of Proposition 2.11 on the cubes $I\times P$ we write

$$\int \widetilde{b}\cdot 1_I\otimes f_P = \big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle + \big\langle [b,T]h_{\widetilde{P}}, g_P\big\rangle + \int \widetilde{b}\widetilde{f}_P,$$

where in line with Proposition 2.11 we notate $g_1 = g_{\widetilde{P}}$, $g_2 = g_P$, $h_1 = h_P$, $h_2 = h_{\widetilde{P}}$, and where we use that the commutator annihilates constants to change $\widetilde{b}$ to b inside the commutator. Consequently,

$$\begin{aligned}\left|\int_R \widetilde{b}f\right| &= \left|\sum_{k=0}^{N}\sum_{P\in\mathscr{S}_k^2}\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle + \big\langle [b,T]h_{\widetilde{P}}, g_P\big\rangle + \int \widetilde{b}\widetilde{f}_P\right| \\ &\leq \sum_{k=0}^{N}\sum_{P\in\mathscr{S}_k^2}\big|\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle\big| + \sum_{k=0}^{N}\sum_{P\in\mathscr{S}_k^2}\big|\big\langle [b,T]h_{\widetilde{P}}, g_P\big\rangle\big| + \left|\int \widetilde{b}\widetilde{f}_{\Sigma}\right|,\end{aligned} \tag{3.16}$$

where we denote $\widetilde{f}_{\Sigma} = \sum_{k=0}^{N}\sum_{P\in\mathscr{S}_k^2}\widetilde{f}_P$.

Let us then focus on the first sum on the right-hand side. The collection $\mathscr{S}_k^2$ is not necessarily finite and the off-support norm 3.11 only controls finite sums. Hence we write

$$\mathscr{S}_{k,j}^2 = \mathscr{S}_k^2 \cap \mathscr{S}_j^2, \quad \mathscr{S}^2 = \bigcup_{j=1}^{\infty}\mathscr{S}_j^2, \quad \mathscr{S}_j^2 \subset \mathscr{S}_{j+1}^2$$

for some finite collections $\mathscr{S}_j^2 \subset \mathscr{S}^2$ and

$$\sum_{k=0}^{N} \sum_{P\in\mathscr{S}_k^2} \big|\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle\big| = \lim_{j\to\infty} \sum_{k=0}^{N} \sum_{P\in\mathscr{S}_{k,j}^2} \big|\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle\big|$$

Notice that the term $\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle$ is bilinear and hence that we may replace $g_{\widetilde{P}}$ with $\alpha_P g_{\widetilde{P}}$ and h_P with $\alpha_P^{-1}h_P$, for any $\alpha_P \neq 0$. We choose $\alpha_P = \|f_P\|_\infty^{\frac{r_2'}{p_2}}$ and estimate

$$\begin{aligned}
&\sum_{k=0}^{N} \sum_{P\in\mathscr{S}_{k,j}^2} \big|\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle\big| = \sum_{k=0}^{N} \sum_{P\in\mathscr{S}_{k,j}^2} \Big|\Big\langle [b,T]\,\|f_P\|_\infty^{\frac{r_2'}{p_2}} 1_{\widetilde{I\times P}}, \|f_P\|_\infty^{-\frac{r_2'}{p_2}} h_P\Big\rangle\Big| \\
&\lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K) \Big\| \sum_{k=0}^{N} \sum_{P\in\mathscr{S}_{k,j}^2} \|f_P\|_\infty^{\frac{r_2'}{p_2}} 1_{\widetilde{I\times P}} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}} \Big\| \sum_{k=0}^{N} \sum_{P\in\mathscr{S}_{k,j}^2} \|f_P\|_\infty^{\frac{r_2'}{q_2'}} 1_{I\times P} \Big\|_{L_{x_1}^{p_1'}L_{x_2}^{q_2'}},
\end{aligned} \tag{3.17}$$

in the last step we used the Definition 3.11 of $\mathcal{O}_{p,q}^{\Sigma}$, the estimate $\|h_P^1\|_\infty \lesssim \|f_P\|_\infty$ and the identity $1-\frac{r_2'}{p_2} = \frac{r_2'}{q_2'}$. By Lemma 3.10 we have

$$\sum_{k=0}^{N} \sum_{P\in\mathscr{S}_{k,j}^2} \big\|f_P\big\|_\infty^{\frac{r_2'}{q_2'}} 1_{I\times P} \lesssim 1_I \otimes (\mathsf{M}f)^{r_2'/q_2'}$$

and this is enough to control the right-most term of the display (3.17). To obtain the similar estimate for the other term we argue as follows. With the rectangle $I\times P$ fixed, write the reflected rectangle as $\widetilde{I\times P} = \widetilde{I}_{I\times P} \times \widetilde{P}_{I\times P}$. Then, by Proposition 2.6 we have

$$\operatorname{dist}(\widetilde{I}_{I\times P}, I) \sim \operatorname{diam}(I), \quad \operatorname{dist}(\widetilde{P}_{I\times P}, P) \sim \operatorname{diam}(P),$$

and hence, there exists some absolute bounded positive constant C such that $C\widetilde{I\times P} \supset I\times P \supset I\times E_P$. This shows that the collection $\big\{C\widetilde{I\times P}\colon P\in\mathscr{S}\big\}$ of rectangles is sparse with the major subsets $I\times E_P$. Hence, we have

$$\begin{aligned}
&\Big\| \sum_{k=0}^{N} \sum_{P\in\mathscr{S}^2} \big\|f_P\big\|_\infty^{\frac{r_2'}{p_2}} 1_{\widetilde{I\times P}} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}} \leq \Big\| \sum_{k=0}^{N} \sum_{P\in\mathscr{S}^2} \big\|f_P\big\|_\infty^{\frac{r_2'}{p_2}} 1_{C\widetilde{I\times P}} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}} \\
&\overset{*}{\lesssim} \Big\| \sum_{k=0}^{N} \sum_{P\in\mathscr{S}^2} \big\|f_P\big\|_\infty^{\frac{r_2'}{p_2}} 1_{I\times E_P} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}} \leq \Big\| \sum_{k=0}^{N} \sum_{P\in\mathscr{S}^2} \big\|f_P\big\|_\infty^{\frac{r_2'}{p_2}} 1_{I\times P} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}},
\end{aligned}$$

where the estimate marked with $*$ can be seen by dualizing and using sparseness, indeed, we have with any function such that $\|g\|_{L_{x_1}^{p_1'}L_{x_2}^{p_2'}} \leq 1$, with any constants a_j, and with any sparse collection $\{R_j, E_{R_j}\}_j$ of rectangles, that

$$\begin{aligned}
\int \sum_j a_j 1_{R_j} g &= \sum_j a_j \langle g\rangle_{R_j} |R_j| \lesssim \sum_j |a_j| \langle |g|\rangle_{R_j} |E_{R_j}| \leq \int \mathsf{M}_{\mathsf{S}} g \sum_j |a_j| 1_{E_{R_j}} \\
&\leq \Big\| \sum_j |a_j| 1_{E_{R_j}} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}} \|\mathsf{M}_{\mathsf{S}} g\|_{L_{x_1}^{p_1'}L_{x_2}^{p_2'}} \lesssim \Big\| \sum_j |a_j| 1_{E_{R_j}} \Big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}},
\end{aligned}$$

where $\mathsf{M_S}$ is the bi-parameter strong maximal function. Hence, we have again reduced to the pointwise estimate $\sum_{k=0}^N \sum_{P\in\mathscr{S}^2} \big\|f_P\big\|_\infty^{\frac{r_2'}{p_2}} 1_{I\times P} \lesssim 1_I \otimes (\mathsf{M}f)^{\frac{r_2'}{p_2}}$ true by Lemma 3.10. The same estimates also holds for the other term on the line (3.16). Putting the above together, we have now shown that

$$\begin{aligned}
&\sum_{k=0}^N \sum_{P\in\mathscr{S}_k^2} \big|\big\langle [b,T]g_{\widetilde{P}}, h_P\big\rangle\big| + \sum_{k=0}^N \sum_{P\in\mathscr{S}_k^2} \big|\big\langle [b,T]h_{\widetilde{P}}, g_P\big\rangle\big| \\
&\lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K)\big\|1_I\otimes(\mathsf{M}f)^{\frac{r_2'}{p_2}}\big\|_{L_{x_1}^{p_1}L_{x_2}^{p_2}}\big\|1_I\otimes(\mathsf{M}f)^{\frac{r_2'}{q_2'}}\big\|_{L_{x_1}^{p_1'}L_{x_2}^{q_2'}} \\
&\lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K)|I|^{1/p_1+1/q_1'} = \mathcal{O}_{p,q}^{\Sigma}(b;K)|I|,
\end{aligned} \tag{3.18}$$

where we used the boundedness of the maximal function and that $\|f\|_{L^{r_2'}(\mathbb{R}^{d_2})} \le 1$. The estimate (3.18) is uniform in j and hence from (3.16) we find that

$$\left|\int_R \widetilde{b}f\right| \lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K)|I| + \left|\int_R \widetilde{b}\widetilde{f}_\Sigma\right|. \tag{3.19}$$

To have control over the error term, we use Proposition 2.11 and Lemma 3.10 to find

$$|\widetilde{f}_\Sigma| \le \sum_{k=0}^N \sum_{P\in\mathscr{S}_k^2} |\widetilde{f}_P| \lesssim \omega\left(\frac{1}{A}\right) 1_I \otimes \sum_{k=0}^N \sum_{P\in\mathscr{S}_k^2} \|f_P\|_\infty 1_P \lesssim \omega\left(\frac{1}{A}\right) 1_I\otimes \mathsf{M}f. \tag{3.20}$$

By (3.20) and Hölder's inequality we have

$$\begin{aligned}
\left|\int_R \widetilde{b}\widetilde{f}_\Sigma\right| &\le \int_I \|\tilde{b}\|_{L_{x_2}^{r_2}(J)} \|\widetilde{f}_\Sigma\|_{L_{x_2}^{r_2'}(J)} \lesssim \int_I \|\tilde{b}\|_{L_{x_2}^{r_2}(J)} \omega\left(\frac{1}{A}\right) \|\mathsf{M}f\|_{L_{x_2}^{r_2'}(J)} \\
&\lesssim \omega\left(\frac{1}{A}\right) \int_I \|\tilde{b}\|_{L_{x_2}^{r_2}(J)}
\end{aligned}$$

and hence continuing from the line (3.19) after dividing by $|I|$ that

$$\left|\fint_I \int_J \widetilde{b}f\right| \lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K) + \omega\left(\frac{1}{A}\right) \fint_I \|\tilde{b}\|_{L_{x_2}^{r_2}(J)}.$$

Hence, by having $I \to \{x_1\}$, the Lebesgue differentiation theorem shows that

$$\left|\int_J \widetilde{b}(x_1,x_2) f(x_2)\,\mathrm{d}x_2\right| \lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K) + \omega(\frac{1}{A})\|\tilde{b}(x_1,x_2)\|_{L_{x_2}^{r_2}(J)}.$$

Since $\widetilde{b}(x_1,\cdot) \in L^{r_2}(J)$ we have

$$\sup_{(3.15)} \left|\int_J \widetilde{b}(x_1,x_2) f(x_2)\,\mathrm{d}x_2\right| = \|\tilde{b}(x_1,x_2)\|_{L_{x_2}^{r_2}(J)},$$

where the supremum is taken over all such f as were considered on the line (3.15). Consequently, we have shown that

$$\|\tilde{b}(x_1,x_2)\|_{L_{x_2}^{r_2}(J)} \lesssim \mathcal{O}_{p,q}^{\Sigma}(b;K) + \omega(\frac{1}{A})\|\tilde{b}(x_1,x_2)\|_{L_{x_2}^{r_2}(J)}.$$

The term shared on both sides of the estimate is finite almost everywhere and hence by absorbing the common term to the left-hand side we conclude with the left-most estimate of (3.14).

The estimate on the middle was already discussed earlier in section 3.11 and it remains to show the right-most estimate. As the commutator is unchanged modulo constants, we find that

$$\|[b,T]f\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}} = \|[b-c,T]f\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}} \leq \|(b-c)Tf\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}} + \|T(b-c)f\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}}.$$

From here, by the mixed norm estimates of T, it is enough to estimate

$$\|(b-c)f\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}} \leq \big\|\|b-c\|_{L^{r_2}_{x_2}}\|f\|_{L^{p_2}_{x_2}}\big\|_{L^{p_1}_{x_1}} \leq \|b-c\|_{L^{\infty}_{x_1}L^{r_2}_{x_2}}\|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}},$$

where we used that $1/q_2 = 1/r_2 + 1/p_2$. Taking the infimum over all $c \in \mathbb{C}$ shows the claim. □

3.21. Proposition. *Let $p_2 = q_2$ and $q_1 < p_1$ and assume that $b \in L^1_{\mathrm{loc},x_2}L^{r_1}_{\mathrm{loc},x_1}$. Then, there holds that*

$$\inf_{c\in\mathbb{C}} \|b-c\|_{L^{\infty}_{x_2}L^{r_1}_{x_1}} \lesssim \mathcal{O}^{\Sigma}_{p,q}(b;K) \leq \|[b,T]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{q_1}_{x_1}L^{p_2}_{x_2}} \lesssim \inf_{c\in\mathbb{C}} \|b-c\|_{L^{r_1}_{x_1}L^{\infty}_{x_2}}.$$

Proof. The left-most estimate is completely symmetric with the proof of Proposition 3.13 and the estimate on the middle is immediate by Hölder's inequality. The right-most estimate follows by the invariance of the commutator modulo additive constants, the mixed norm estimates of T, and Hölders inequality. □

3.5. The case $p_1 > q_1$ and $p_2 > q_2$. In this case, again, it follows immediately by Hölder's inequality, the invariance of the commutator modulo additive constants, and the mixed norm estimates of T, that

$$\mathcal{O}^{\Sigma}_{p,q}(b;K) \leq \|[b,T]\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}\to L^{q_1}_{x_1}L^{q_2}_{x_2}} \leq \inf_{c\in\mathbb{C}} \|b-c\|_{L^{r_1}_{x_1}L^{r_2}_{x_2}}.$$

Then, we would like to prove a lower bound for $\mathcal{O}^{\Sigma}_{p,q}(b;K)$ that gets as close to $\inf_{c\in\mathbb{C}}\|b-c\|_{L^{r_1}_{x_1}L^{r_2}_{x_2}}$ as possible. Let us first discuss the non-mixed case, where we have a full characterization.

3.22. Proposition. *Let $p_1 = p_2 > q_1 = q_2$, define $1/r = 1/q_1 - 1/p_1$, and assume that $b \in L^1_{\mathrm{loc}}$. Then, there holds that*

$$\inf_{c\in\mathbb{C}} \|b-c\|_{L^r(\mathbb{R}^d)} \sim \mathcal{O}^{\Sigma}_{p,q}(b;K) \sim \|[b,T]\|_{L^p(\mathbb{R}^d)\to L^q(\mathbb{R}^d)}.$$

Proof. The following oscillatory characterization is recorded e.g. as Proposition 3.2. in [1]. Let $r \in (1,\infty)$, then there holds that

$$\begin{aligned} &\inf_{c\in\mathbb{C}} \|b-c\|_{L^r(\mathbb{R}^d)} \\ &\sim \sup_{\mathscr{S}} \left\{ \sum_{Q\in\mathscr{S}} \lambda_Q |Q| \operatorname{osc}(b,Q) \colon \mathscr{S} \text{ is } 1/2\text{-sparse}, \sum_{Q\in\mathscr{S}} |Q|\lambda_Q^{r'} \leq 1 \right\}, \end{aligned} \tag{3.23}$$

where the sparse collections $\mathscr{S}$ consist of cubes of $\mathbb{R}^d$. Now, fix any sparse collection $\mathscr{S}$ as in the supremum. Then, identically as in the proof of Proposition 3.24, we can bound

$$\sum_{Q\in\mathscr{S}} \lambda_Q |Q| \operatorname{osc}(b,Q) \lesssim \mathcal{O}^{\Sigma}_{p,q}(b;K).$$

The remaining bounds $\mathcal{O}^{\Sigma}_{p,q}(b;K) \lesssim \|[b,T]\|_{L^p(\mathbb{R}^d)\to L^q(\mathbb{R}^d)} \lesssim \inf_{c\in\mathbb{C}}\|b-c\|_{L^r(\mathbb{R}^d)}$ were already discussed above in the mixed case. □

In the mixed case we are unable to prove the desired lower bound and what we have is the following

3.24. Proposition. *Let $p_1 > q_1, p_2 > q_2$ and let $b \in L^1_{\mathrm{loc}}$. Let $\mathscr{S}^i$ denote a 1/2-sparse collection on $\mathbb{R}^{d_i}$ with associated coefficients $\{\lambda_{I_i}\}$ such that $\sum_{I_i \in \mathscr{S}^i} \lambda_{I_i}^{r_i'} |I_i| \leq 1$. Then,there holds that*

$$\sup_{\mathscr{S}^1, \mathscr{S}^2} \left[\sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1} \lambda_{I_2} |I_1||I_2| \operatorname{osc}(b, I_1 \times I_2) \right] \lesssim \mathcal{O}^{\Sigma}_{p,q}(b; K).$$

Technically this limitation is due to the failure of finding any useful rectangular sparse oscillatory characterization of the mixed space $L^s_{x_1} L^t_{x_2}$, when $s \neq t$, that would correspond with that of the one on the line (3.23).

Proof of Proposition 3.24. Without loss of generality we may assume that the collections $\mathscr{S}^i$ are finite. First, by Proposition 2.16 we have

$$|I_1||I_2| \operatorname{osc}(b, I_1 \times I_2) \lesssim |\langle [b,T] g^1_{I_1 \times I_2}, h^1_{I_1 \times I_2}, \rangle| + |\langle [b,T] h^2_{I_1 \times I_2}, g^2_{I_1 \times I_2} \rangle| \tag{3.25}$$

where we write $h^i_{I_1 \times I_2}, g^i_{I_1 \times I_2}$ for the functions g_i, h_i. Also, let $R(h^i_{I_1 \times I_2})$ and $R(g^i_{I_1 \times I_2})$ stand respectively for the rectangles on which $h^i_{I_1 \times I_2}$ and $g^i_{I_1 \times I_2}$ are supported. Then, by (3.25), the relation $1/r_i = 1/q_i - 1/p_i$, and the Definition 3.11 of the off-support norm, we estimate

$$\begin{aligned}
&\sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1} \lambda_{I_2} |I_1||I_2| \operatorname{osc}(b, I_1 \times I_2) \\
&\lesssim \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1} \lambda_{I_2} |\langle [b,T] g^1_{I_1 \times I_2}, h^1_{I_1 \times I_2}, \rangle| + \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1} \lambda_{I_2} |\langle [b,T] h^2_{I_1 \times I_2}, g^2_{I_1 \times I_2} \rangle|
\end{aligned}$$

and let us estimate the sums as

$$\begin{aligned}
&\lesssim \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1} \lambda_{I_2} |\langle [b,T] h^i_{I_1 \times I_2}, g^i_{I_1 \times I_2} \rangle| \\
&= \sum_{i=1,2} \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} |\langle [b,T] (\lambda_{I_1}^{r_1'/p_1} \lambda_{I_2}^{r_2'/p_2} h^i_{I_1 \times I_2}), \lambda_{I_1}^{r_1'/q_1'} \lambda_{I_2}^{r_2'/q_2'} g^i_{I_1 \times I_2} \rangle| \\
&\leq \mathcal{O}^{\Sigma}_{p,q}(b; K) \sum_{i=1,2} \left\| \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1}^{r_1'/p_1} \lambda_{I_2}^{r_2'/p_2} 1_{R(h^i_{I_1 \times I_2})} \right\|_{L^{p_1}_{x_1} L^{p_2}_{x_2}} \\
&\qquad \times \left\| \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1}^{r_1'/q_1'} \lambda_{I_2}^{r_2'/q_2'} 1_{R(g^i_{I_1 \times I_2})} \right\|_{L^{q_1'}_{x_1} L^{q_2'}_{x_2}}.
\end{aligned}$$

Using that the coefficients are of product form, we can then, for example, estimate one of the terms as

$$\begin{aligned}
&\left\| \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1}^{r_1'/q_1'} \lambda_{I_2}^{r_2'/q_2'} 1_{R(g^2_{I_1 \times I_2})} \right\|_{L^{q_1'}_{x_1} L^{q_2'}_{x_2}} = \left\| \sum_{I_1 \in \mathscr{S}^1} \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_1}^{r_1'/q_1'} \lambda_{I_2}^{r_2'/q_2'} 1_{I_1 \times I_2} \right\|_{L^{q_1'}_{x_1} L^{q_2'}_{x_2}} \\
&= \left\| \sum_{I_1 \in \mathscr{S}^1} \lambda_{I_1}^{r_1'/q_1'} 1_{I_1} \right\|_{L^{q_1'}_{x_1}} \left\| \sum_{I_2 \in \mathscr{S}^2} \lambda_{I_2}^{r_2'/q_2'} 1_{I_2} \right\|_{L^{q_2'}_{x_2}} \lesssim 1,
\end{aligned}$$

where in the last step we used the sparseness of the collections $\mathscr{S}^i$ and the assumed bounds $\sum_{I_i \in \mathscr{S}^i} \lambda_{I_i}^{r_i'} |I_i| \leq 1$. The remaining three terms are estimated in the same

fashion, basically repeating the arguments that we already went through in the proof of Proposition 3.13. □

4. Upper bound for the case $p_1 = q_1$, $p_2 < q_2$

We are now left with two cases and we first deal with this one. We will use the representation of bi-parameter CZO's as dyadic model operators; this is maybe surprising as the corresponding lower bound obtained in Proposition 3.9 seems simple and should perhaps yield an easier proof. We will prove

4.1. Proposition. *Let $p_1 = q_1$ and $p_2 < q_2$, let $b(x_1, \cdot) \in \dot{C}^{0,\alpha_2}(\mathbb{R}^{d_2})$ and $b(\cdot, x_2) =$ constant. Then, we have*

$$\|[b,T]f\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}} \lesssim \|b(x_1,\cdot)\|_{\dot{C}^{0,\alpha}_{x_2}(\mathbb{R}^{d_2})}\|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}.$$

The dyadic representation theorem of bi-parameter CZO's of Martikainen [14] is the following

4.2. Theorem. *Given a bi-parameter CZO, it can be written as an expectation*

$$\langle Tf, g\rangle = C_T\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2} \sum_{\substack{i=(i_1,i_2)\in\mathbb{N}^2\\ j=(j_1,j_2)\in\mathbb{N}^2}} 2^{-\max(i_1,i_2)\frac{\delta}{2}}2^{-\max(j_1,j_2)\frac{\delta}{2}}\big\langle S^{i,j}_{\mathcal{D}^1_{\omega_1},\mathcal{D}^2_{\omega_2}}f, g\big\rangle,$$

where $S^{i,j}_{\mathcal{D}^1_{\omega_1},\mathcal{D}^2_{\omega_2}}$ are bi-parameter dyadic model operators (detailed below) associated to the randomized dyadic grids $\mathcal{D}^1_{\omega_1}$ and $\mathcal{D}^2_{\omega_2}$.

By Theorem 4.2 to have estimates for $[b,T]$, it is enough to have them for $[b, S^{i,j}]$, where $S^{i,j}$ is a dyadic model operator, and with constants of at most polynomial growth in the parameters i, j, namely, it is enough to prove the following

4.3. Proposition. *Let $p_1 = q_1$ and $p_2 < q_2$, let $b(x_1, \cdot) \in \dot{C}^{0,\alpha_2}(\mathbb{R}^{d_2})$ and $b(\cdot, x_2) =$ constant. Then, we have*

$$\big\|[b, S^{i,j}]f\big\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}} \lesssim \|b(x_1,\cdot)\|_{\dot{C}^{0,\alpha_2}(\mathbb{R}^{d_2})}\|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}$$

with an implied constant of at most polynomial growth in i, j.

We have that $S^{i,j}$ is either a shift, a partial paraproduct or a full paraproduct, to detail each of which we first recall few basic facts about martingale differences and Haar functions, the reader familiar with these may skip to Section 4.1.4.

4.0.1. Haar functions, basic facts. Given a dyadic grid $\mathcal{D}$ and a cube $I \in \mathcal{D}$ the martingale difference on I is

$$\Delta_I f = \sum_{P\in \mathrm{ch}(I)} \big(\langle f\rangle_P - \langle f\rangle_I\big)1_P.$$

These are naturally useful as $f = \sum_{I\in\mathcal{D}}\Delta_I f$, where each element is nicely localized and has zero mean. For a given interval $I = I_l \bigcup I_r \subset \mathbb{R}$, with a left- and a right half, respectively the cancellative and non-cancellative Haar functions supported I are

$$h_I = \frac{1_{I_l} - 1_{I_r}}{|I|^{1/2}}, \quad h^0_I = \frac{1_I}{|I|^{1/2}}.$$

Given a rectangle $I = I_1 \times \cdots \times I_d \subset \mathbb{R}^d$, the Haar functions on I are

$$h_I = \otimes_{i=1}^d \widetilde{h}_{I_i}, \quad \widetilde{h}_{I_i} \in \{h_{I_i}, h^0_{I_i}\}$$

on the condition that at least one component is a cancellative Haar function. Hence, all in all, there are 2^d-1 Haar functions on any given rectangle of dimension d, along with the non-cancellative Haar function $1_I/|I|^{1/2}$. It is a basic fact that

$$\Delta_I f = \sum_{i=1}^{2^d-1} \langle f, h_I^i\rangle h_I^i \tag{4.4}$$

where we enumerate all 2^d-1 cancellative Haar functions on the rectangle I. Hence, when proving upper bounds we just write $h_I = h_I^i$ for a generic cancellative Haar function on I and it is customary to ignore the $i = 1, \dots, 2^d-1$ summation in (4.4).

Fix a rectangle $R = I \times J \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Fully cancellative Haar functions of product form are the tensor products $h_R = h_I \otimes h_J$, where both h_I, h_J are cancellative Haar functions respectively on I and J. Then, simply,

$$\Delta_{I\times J} f = \Delta_I\big(\Delta_J f\big) = \sum_{i=1}^{2^{d_1}-1}\sum_{j=1}^{2^{d_2}-1} \big\langle \big\langle f, h_J^j\big\rangle h_J^j, h_I^i\big\rangle h_I^i = \sum_{i=1}^{2^{d_1}-1}\sum_{j=1}^{2^{d_2}-1} \big\langle f, h_I^i\otimes h_J^j\big\rangle h_I^i\otimes h_J^j,$$

and again each Haar function $h_R = h_I^i \otimes h_J^j$ carries enough cancellation for boundedness of bi-parameter square functions etc.

4.1. Model operators. A pair of intervals we denote $(I) = (I_1, I_2)$ and with $I^k = I^{(k)} = Q$ we mean that $I, Q \in \mathcal{D}$, $I \subset Q$ and $\ell(I) = 2^{-k}\ell(Q)$. Now, the bi-parameter dyadic model operators of Theorem 4.2 have the generic form

$$\big\langle S^{i,j} f, g\big\rangle = \sum_{\substack{K\in\mathcal{D}^1\\ I_1^{i_1}=I_2^{i_2}=K}} \sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}} \alpha_{(I)(J)KV} \langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle,$$

where the coefficients $\alpha_{(I)(J)KV}$ have sizes according to which dyadic model operator we have: There are in total three different kinds of model operators that appear in 4.2.

4.1.1. Shifts. We have

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \langle f, h_{I_1\times J_1}\rangle\langle g, h_{I_2\times J_2}\rangle$$

where each of the Haar functions is cancellative and the coefficients have the size

$$|\alpha_{(I)(J)KV}| \lesssim \frac{(|I_1||I_2||J_1||J_2|)^{1/2}}{|K\times V|}.$$

4.1.2. Partial paraproducts. We have $i_1 = i_2 = 0$ and

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \left\langle f, \frac{1_K}{|K|}\otimes h_{J_1}\right\rangle \langle g, h_K\otimes h_{J_2}\rangle,$$

or the symmetric case,

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \langle f, h_K\otimes h_{J_1}\rangle \left\langle g, \frac{1_K}{|K|}\otimes h_{J_2}\right\rangle,$$

and in both cases the coefficients have the size

$$\|(\alpha_{(J)KV})_K\|_{\mathrm{BMO}_2(\mathbb{R}^{d_1})} = \sup_{K_0\in\mathcal{D}^1} \frac{1}{|K_0|^{1/2}} \left\| \left(\sum_{\substack{K\in\mathcal{D}\\ K\subset K_0}} |\alpha_{(J)KV}|^2 \frac{1_K}{|K|}\right)^{1/2}\right\|_{L^2(\mathbb{R}^{d_1})}$$

$$\lesssim \frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}.$$

There is also the other symmetry of $j_1 = j_2 = 0$, and then

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \left\langle f, h_{I_1}\otimes\frac{1_V}{|V|}\right\rangle\langle g, h_{I_2}\otimes h_V\rangle,$$

and its symmetric case

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \langle f, h_{I_1}\otimes h_V\rangle\left\langle g, h_{I_2}\otimes\frac{1_V}{|V|}\right\rangle,$$

and in both of these two cases the coefficients have the size

$$\|(\alpha_{(I)KV})_V\|_{\mathrm{BMO}_2(\mathbb{R}^{d_2})} = \sup_{V_0\in\mathcal{D}^1}\frac{1}{|V_0|^{1/2}}\left\|\left(\sum_{\substack{V\in\mathcal{D}\\ V\subset V_0}}|\alpha_{(I)KV}|^2\frac{1_V}{|V|}\right)^{1/2}\right\|_{L^2(\mathbb{R}^{d_2})}$$
$$\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}.$$

4.1.3. Full paraproducts. We have $i_1 = i_2 = j_1 = j_2 = 0$ and

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \langle f\rangle_{K\times V}\langle g, h_K\otimes h_V\rangle$$

or the symmetric case

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \langle f, h_K\otimes h_V\rangle\langle g\rangle_{K\times V},$$

or we have the other symmetry

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \left\langle f, h_K\otimes\frac{1_V}{|V|}\right\rangle\left\langle g, \frac{1_K}{|K|}\otimes h_V\right\rangle$$

and its symmetric case

$$\langle f, \widetilde{h}_{I_1\times J_1}\rangle\langle g, \widetilde{h}_{I_2\times J_2}\rangle = \left\langle f, \frac{1_K}{|K|}\otimes h_V\right\rangle\left\langle g, h_K\otimes\frac{1_V}{|V|}\right\rangle.$$

The boundedness of full paraproducts bootstraps directly to Proposition 4.5 below and to the boundedness of fractional operators. Hence, we will not record their coefficient size, nonetheless, we mention that the coefficient size is measured by the product BMO space of Chang and Fefferman, see e.g. Section 7 in [1].

The following Proposition 4.5 is e.g. contained as a part of Hytönen–Martikainen–Vuorinen [8].

4.5. Proposition. *All the above described dyadic model operators, the shifts, the partial paraproducts and the full paraproducts, are bounded*

$$\|S^{i,j}f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}} \lesssim \|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}$$

with an implied constant of at most polynomial growth in i, j.

4.1.4. Decomposition of products. Notice that as the function b bears no important information in the first variable, we only need to analyse it carefully in the second parameter, which we do according to the commutator decomposition strategy from [12]:

(i) Whenever a product bf (or bg) is paired against a cancellative Haar function h_J and $J \in \mathcal{D}^2$, we expand with respect to the dyadic grid $\mathcal{D}^2$ as

$$\begin{aligned} bf &= \sum_{J\in\mathcal{D}^2} \Delta_J b \Delta_J f + \sum_{J\in\mathcal{D}^2} \Delta_J b E_J f + \sum_{J\in\mathcal{D}^2} E_J b \Delta_J f \\ &= A_1(b,f) + A_2(b,f) + A_3(b,f), \end{aligned}$$

where we denote $E_J b = \langle b\rangle_J 1_J$. It should be understood that the operators A_i depend on the fixed dyadic grid $\mathcal{D}^2$ even though we omit this detail from the notation. Especially, if our model operators $S^{i,j}$ are defined on the grid $\mathcal{D}^1 \times \mathcal{D}^2$, then we will expand in the grid $\mathcal{D}^2$.

(ii) If bf is averaged in the second parameter, then we add and subtract $\langle bf\rangle_J 1_J$,

$$bf 1_J = (bf - \langle bf\rangle_J)1_J + \langle bf\rangle_J 1_J.$$

The proof of Proposition 4.3 splits into several cases of which some are symmetric; as there are too many cases to present here fully, we go through a proof of each representative case for each model operator after which it is clear how to carry through the remaining cases.

The first step is to establish the boundedness for the auxiliary operators.

4.6. Proposition. *Let $1 < p < q < \infty$ and $\alpha = d(1/p - 1/q)$. Then,*

$$\|A_i(b,f)\|_{L^q(\mathbb{R}^d)} \lesssim \|b\|_{\dot{C}^{0,\alpha}(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)}. \tag{4.7}$$

Proof. Let us first estimate

$$\begin{aligned} |A_i(b,f)| &\le \sum_{Q\in\mathcal{D}} \frac{|\langle b, h_Q\rangle|}{|Q|^{1/2}} \langle |f|\rangle_Q 1_Q = \sum_{Q\in\mathcal{D}} \frac{|\langle b - \langle b\rangle_Q, h_Q\rangle|}{|Q|^{1/2}} \langle |f|\rangle_Q 1_Q \\ &\le \sum_{Q\in\mathcal{D}} \langle |b - \langle b\rangle_Q|\rangle_Q \langle |f|\rangle_Q 1_Q \le \|b\|_{\dot{C}^{0,\alpha}(\mathbb{R}^d)} \sum_{Q\in\mathcal{D}} \ell(Q)^\alpha \langle |f|\rangle_Q 1_Q. \end{aligned} \tag{4.8}$$

Then, we show that the positive operator

$$\mathsf{A}^\alpha_{\mathcal{D}} f = \sum_{Q\in\mathcal{D}} \ell(Q)^\alpha \langle |f|\rangle_Q 1_Q \tag{4.9}$$

satisfies the desired bound. Fix a top cube $Q_0 \in \mathcal{D}$ and let $\mathscr{S} \subset \mathcal{D}(Q_0)$ be the stopping time sparse collection inside the cube Q_0 as described in the beginning of Section 3.4. By the stopping condition and sparseness of $\mathscr{S}$, we estimate

$$\begin{aligned} \|\mathsf{A}^\alpha_{\mathcal{D}_{Q_0}} f\|_{L^q(\mathbb{R}^d)} &= \left\| \sum_{P\in\mathscr{S}} \sum_{\Pi Q = P} \ell(Q)^\alpha \langle |f|\rangle_Q 1_Q \right\|_{L^q(\mathbb{R}^d)} \lesssim \left\| \sum_{P\in\mathscr{S}} \langle |f|\rangle_P \sum_{\substack{Q\in\mathcal{D}_{Q_0} \\ Q\subset P}} \ell(Q)^\alpha 1_Q \right\|_{L^q(\mathbb{R}^d)} \\ &\le \left\| \sum_{P\in\mathscr{S}} \langle |f|\rangle_P \left(\sum_{k=0}^\infty 2^{-k\alpha} \right) \ell(P)^\alpha 1_P \right\|_{L^q(\mathbb{R}^d)} \lesssim \left\| \sum_{P\in\mathscr{S}} \ell(P)^\alpha \langle |f|\rangle_P 1_P \right\|_{L^q(\mathbb{R}^d)} \\ &\overset{*}{\lesssim} \left\| \sum_{P\in\mathscr{S}} \ell(P)^\alpha \langle |f|\rangle_P 1_{E_P} \right\|_{L^q(\mathbb{R}^d)} = \left(\sum_{P\in\mathscr{S}} \int_{E_P} (\ell(P)^\alpha \langle |f|\rangle_P)^q \right)^{1/q} \\ &\le \|\mathsf{M}^\alpha f\|_{L^q(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}, \end{aligned}$$

where at the estimate marked with $*$ we used the sparseness of $\mathscr{S}$ to get the norm estimate (for details, see the similar estimate in the proof of Proposition 3.13), and

where the boundedness of the fractional maximal operator,

$$\mathsf{M}^\alpha f(x) = \sup_Q 1_Q(x)\ell(Q)^\alpha \fint_Q |f|,$$

where the supremum is taken over all cubes $Q \subset \mathbb{R}^d$, was used. As the demonstrated bound is independent of the choice of the top cube Q_0, we get the boundedness for $\mathsf{A}^\alpha_{\mathcal{D}}$ and hence (4.7). □

We will also have use of the following fractional Fefferman–Stein inequality due to Blasco and Hernandez [16].

4.10. Lemma. *Let $1 < p < q < \infty$, $\alpha = d(1/p - 1/q) < d$, and $1 < r < \infty$. Then, there holds that*

$$\left\| \left(\sum_k (\mathsf{M}^\alpha f_k)^r \right)^{1/r} \right\|_{L^q(\mathbb{R}^d)} \lesssim_{d,p,q,r} \left\| \left(\sum_k |f_k|^r \right)^{1/r} \right\|_{L^p(\mathbb{R}^d)}.$$

4.11. Remark. Lemma 4.10 becomes Fefferman–Stein inequality when $p = q$.

For the following two lemmas see e.g. [8].

4.12. Lemma. *Let $1 < p_1, p_2 < \infty$. Then, there holds that*

$$\big\|\mathsf{S}^1 f\big\|_{L^{p_1}_{x_1} L^{p_2}_{x_2}} \sim \big\|\mathsf{S}^2 f\big\|_{L^{p_1}_{x_1} L^{p_2}_{x_2}} \sim \big\|\mathsf{S} f\big\|_{L^{p_1}_{x_1} L^{p_2}_{x_2}} \sim \big\| f \big\|_{L^{p_1}_{x_1} L^{p_2}_{x_2}},$$

hold, where

$$\mathsf{S}^i f = \left(\sum_{L \in \mathcal{D}^i} |\langle f, h_L \rangle|^2 \frac{1_L}{|L|} \right)^{1/2}, \quad \mathsf{S} f = \left(\sum_{\substack{I \in \mathcal{D}^1 \\ J \in \mathcal{D}^2}} |\langle f, h_I \otimes h_J \rangle|^2 \frac{1_{I \times J}}{|I \times J|} \right)^{1/2}.$$

4.13. Lemma. *Let $1 < s, t, r < \infty$. Then, there holds that*

$$\left\| \left(\sum_k \mathsf{M}^1 \mathsf{M}^2 f_k \right)^{1/r} \right\|_{L^s_{x_1} L^t_{x_2}} \lesssim_{s,t,r} \left\| \left(\sum_k |f_k| \right)^{1/r} \right\|_{L^s_{x_1} L^t_{x_2}}.$$

Proof of Theorem 4.3, part 1/3, shifts. Let $S^{i,j}$ stand for the model operator

$$\big\langle S^{i,j} f, g \big\rangle = \sum_{K \in \mathcal{D}^1} \sum_{\substack{V \in \mathcal{D}^2 \\ I_1^{i_1} = I_2^{i_2} = K \\ J_1^{j_1} = J_2^{j_2} = V}} \alpha_{(I)(J)KV} \langle f, h_{I_1 \times J_1} \rangle \langle g, h_{I_2 \times J_2} \rangle. \tag{4.14}$$

By the above described decomposition strategy, we find that the summand (without the scaling factor $\alpha_{(I)(J)KV}$ in front) in (4.14) writes out as

$$\begin{aligned}
&\big[\langle f, h_{I_1} \otimes h_{J_1} \rangle \langle bg, h_{I_2} \otimes h_{J_2} \rangle - \langle bf, h_{I_1} \otimes h_{J_1} \rangle \langle g, h_{I_2} \otimes h_{J_2} \rangle \big] \\
&= \sum_{i=1,2} \langle f, h_{I_1} \otimes h_{J_1} \rangle \langle A_i(b, g), h_{I_2} \otimes h_{J_2} \rangle - \sum_{i=1,2} \langle A_i(b, f), h_{I_1} \otimes h_{J_1} \rangle \langle g, h_{I_2} \otimes h_{J_2} \rangle \\
&\quad + \left[\langle f, h_{I_1} \otimes h_{J_1} \rangle \langle A_3(b, g), h_{I_2} \otimes h_{J_2} \rangle - \langle A_3(b, f), h_{I_1} \otimes h_{J_1} \rangle \langle g, h_{I_2} \otimes h_{J_2} \rangle \right].
\end{aligned}$$

The terms with the first four summands are bounded by the mixed norm estimates of bi-parameter model operators and Proposition 4.6. Indeed, for the first two terms we use that $A_i(b, \cdot)\colon L^{q_2'}_{x_2} \to L^{p_2'}_{x_2}$ boundedly, and for the following two terms directly

Proposition 4.6. For the bracketed difference on the last line we utilise the cancellation of the commutator, hence writing it out as

$$\begin{aligned}(4.15)\qquad &\langle f, h_{I_1}\otimes h_{J_1}\rangle\langle b\rangle_{J_2}\langle g, h_{I_2}\otimes h_{J_2}\rangle - \langle b\rangle_{J_1}\langle f, h_{I_1}\otimes h_{J_1}\rangle\langle g, h_{I_2}\otimes h_{J_2}\rangle\\ &= (\langle b\rangle_{J_2} - \langle b\rangle_{J_1})\langle f, h_{I_1}\otimes h_{J_1}\rangle\langle g, h_{I_2}\otimes h_{J_2}\rangle.\end{aligned}$$

Recall, that we may assume the slice $b(\cdot, x_2)\colon \mathbb{R}^{d_1} \to \mathbb{C}$ to be a constant for all $x_2 \in \mathbb{R}^{d_2}$.

Then, similarly as in e.g. (4.8), we estimate $|\langle b\rangle_{J_2} - \langle b\rangle_{J_1}| \le \|b(x_1,\cdot)\|_{\dot{C}^{0,\alpha_2}_{x_2}}\ell(V)^{\alpha_2}$ for any $x_1 \in \mathbb{R}^{d_1}$. Let us simply notate $\|b(x_1,\cdot)\|_{\dot{C}^{0,\alpha_2}_{x_2}} = \|b\|_{\dot{C}^{0,\alpha_2}_{x_2}}$. Then, we estimate the remaining part of the commutator,

$$\begin{aligned}&\left|\sum_{\substack{K\in\mathcal{D}^1\\ I_1^{i_1}=I_2^{i_2}=K}}\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}} \alpha_{(I)(J)KV}(\langle b\rangle_{J_2} - \langle b\rangle_{J_1})\langle f, h_{I_1}\otimes h_{J_1}\rangle\langle g, h_{I_2}\otimes h_{J_2}\rangle\right|\\
&\le \|b\|_{\dot{C}^{0,\alpha_2}_{x_2}}\int \sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} \ell(V)^{\alpha_2}\big\langle|\Delta^{i_1,j_1}_{K,V}f|\big\rangle_{K\times V}\big\langle|\Delta^{i_2,j_2}_{K,V}g|\big\rangle_{K\times V}1_K\otimes 1_V\\
&\le \|b\|_{\dot{C}^{0,\alpha_2}_{x_2}}\left\|\left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}}\Big(\ell(V)^{\alpha_2}\big\langle|\Delta^{i_1,j_1}_{K,V}f|\big\rangle_{K\times V}\Big)^2 1_K\otimes 1_V\right)^{1/2}\right\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}}\\
&\quad\times\left\|\left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}}\big\langle|\Delta^{i_1,j_1}_{K,V}g|\big\rangle^2_{K\times V}1_K\otimes 1_V\right)^{1/2}\right\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}}\\
&\lesssim \|b\|_{\dot{C}^{0,\alpha_2}_{x_2}}\|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}\|g\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}},\end{aligned}$$

where in the last step we estimate as follows: first, for the fractional term, we note that as

$$\Big(\ell(V)^{\alpha_2}\big\langle|\Delta^{i_1,j_1}_{K,V}f|\big\rangle_{K\times V}\Big)^2 1_K\otimes 1_V \lesssim \Big(\mathsf{M}^{\alpha_2}\big(\langle|\Delta^{i_1,j_1}_{K,V}f|\rangle_{K\times V}1_K\otimes 1_V\big)\Big)^2,$$

by applying Lemma 4.10, we have

$$\begin{aligned}&\left\|\left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}}\Big(\ell(V)^{\alpha_2}\langle|\Delta^{i_1,j_1}_{K,V}f|\rangle_{K\times V}\Big)^2 1_K\otimes 1_V\right)^{1/2}\right\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}}\\
&\lesssim\left\|\left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}}\langle|\Delta^{i_1,j_1}_{K,V}f|\rangle^2_{K\times V}1_K\otimes 1_V\right)^{1/2}\right\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}\end{aligned}$$

$$\lesssim^{*} \left\| \left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} |\Delta_{K,V}^{i_1,j_1} f|^2 1_K\otimes 1_V \right)^{1/2} \right\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}} \overset{**}{\leq} \left\| \left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} |\Delta_{K,V}^{0,0} f|^2 1_K\otimes 1_V \right)^{1/2} \right\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}$$
$$\lesssim \big\|\mathsf{S} f\big\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}} \lesssim \big\|f\big\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}},$$

where the $*$-estimate follows by Lemma 4.13, and the $**$-estimate follows as

$$\left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} |\Delta_{K,V}^{i_1,j_1} f|^2 1_K\otimes 1_V \right)^{\frac12} = \left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} \left| \sum_{\substack{I^{i_1}=K\\ J^{j_1}=V}} \Delta_{I^{i_1},J^{j_1}}^{0,0} f \right|^2 1_K\otimes 1_V \right)^{\frac12}$$
$$\leq \left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} \sum_{\substack{I^{i_1}=K\\ J^{j_1}=V}} |\Delta_{I^{i_1},J^{j_1}}^{0,0} f|^2 1_K\otimes 1_V \right)^{\frac12} = \left(\sum_{\substack{K\in\mathcal{D}^1\\ V\in\mathcal{D}^2}} |\Delta_{K,V}^{0,0} f|^2 1_K\otimes 1_V \right)^{\frac12}.$$

The remaining non-fractional term estimates in the same fashion and we leave the details to the reader. □

With partial paraproducts we will use the following side of the classical and well-known H^1-BMO -duality.

4.16. Lemma. *Let $\mathcal{D}$ be a dyadic grid. Then, for any arbitrary sequences (α_Q), (β_Q) there holds that*

$$\sum_{Q\in\mathcal{D}} |\alpha_Q||\beta_Q| \lesssim \|(\alpha_Q)\|_{\mathrm{BMO}} \big\|\mathsf{S}_{\mathcal{D}}(\beta_Q)\big\|_{L^1(\mathbb{R}^d)},$$

where,

$$\|(\alpha_Q)\|_{\mathrm{BMO}} = \sup_{Q_0\in\mathcal{D}} \frac{1}{|Q_0|^{1/2}} \left\| \left(\sum_{\substack{Q\in\mathcal{D}\\ Q\subset Q_0}} |\alpha_Q|^2 \frac{1_Q}{|Q|} \right)^{\frac12} \right\|_{L^2(\mathbb{R}^d)}, \quad \mathsf{S}_{\mathcal{D}}(\beta_Q) = \left(\sum_{Q\in\mathcal{D}} |\beta_Q|^2 \frac{1_Q}{|Q|} \right)^{\frac12}.$$

Proof of Theorem 4.3, part 2/3, partial paraproducts. We choose the symmetry $i_1 = i_2 = 0$ and consider the model operator

$$\langle S^{(0,0),j} f, g\rangle = \sum_{K\in\mathcal{D}^1} \sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}} \alpha_{(J)KV} \left\langle f, \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle g, h_K\otimes h_{J_2}\rangle.$$

Writing out the main term, we find out that the summand (without the scaling factor $\alpha_{(I)(J)KV}$ in front) in $\langle [b, S^{i,j}] f, g\rangle$ is

$$\left[\left\langle f, \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle bg, h_K\otimes h_{J_2}\rangle - \left\langle bf, \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle g, h_K\otimes h_{J_2}\rangle \right]$$
$$= \sum_{i=1,2} \left\langle f, \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle A_i(b,g), h_K\otimes h_{J_2}\rangle - \sum_{i=1,2} \left\langle A_i(b,f), \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle g, h_K\otimes h_{J_2}\rangle$$
$$+ \left[\left\langle f, \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle A_3(b,g), h_K\otimes h_{J_2}\rangle - \left\langle A_3(b,f), \frac{1_K}{|K|}\otimes h_{J_1} \right\rangle \langle g, h_K\otimes h_{J_2}\rangle \right].$$

The terms with the first four summands are bounded by the mixed norm estimates of bi-parameter model operators and Lemma 4.6, as in the previous case, and the difference on the last line writes out to reduce us to bounding the form

$$\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}}\sum_{K\in\mathcal{D}^1}\alpha_{(J)KV}(\langle b\rangle_{J_2}-\langle b\rangle_{J_1})\left\langle f,\frac{1_K}{|K|}\otimes h_{J_1}\right\rangle\langle g,h_K\otimes h_{J_2}\rangle.$$

Then, as above, we estimate $|\langle b\rangle_{J_2}-\langle b\rangle_{J_1}|\le\|b\|_{\dot{C}^{0,\alpha_2}_{x_2}}\ell(V)^{\alpha_2}$ and this gives the desired factor $\|b\|_{\dot{C}^{0,\alpha_2}_{x_2}}$ in front. It remains to estimate as follows. By Lemma 4.16 and the coefficient size of the partial paraproduct, we find the first estimate in the following, with the rest being straightforward or follow by Lemmas 4.10 and 4.12,

$$\begin{aligned}
&\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}}\sum_{K\in\mathcal{D}^1}\left|\alpha_{(J)KV}\ell(V)^{\alpha_2}\left\langle f,\frac{1_K}{|K|}\otimes h_{J_1}\right\rangle\langle g,h_K\otimes h_{J_2}\rangle\right|\\
&\lesssim\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}}\frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}\ell(V)^{\alpha_2}\left\|\left(\sum_{K\in\mathcal{D}^1}\left|\left\langle f,\frac{1_K}{|K|}\otimes h_{J_1}\right\rangle\langle g,h_K\otimes h_{J_2}\rangle\right|^2\frac{1_K}{|K|}\right)^{\frac12}\right\|_{L^1(\mathbb{R}^{d_1})}\\
&\le\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}}\frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}\ell(V)^{\alpha_2}\int_{\mathbb{R}^{d_1}}\mathsf{M}^1(\langle f,h_{J_1}\rangle)\mathsf{S}^1(\langle g,h_{J_2}\rangle)\\
&=\int_{\mathbb{R}^{d_1}}\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}}\frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}\ell(V)^{\alpha_2}\mathsf{M}^1\left(\int_{J_1}\Delta_V^{j_1}fh_{J_1}\right)\mathsf{S}^1\left(\int_{J_2}\Delta_V^{j_2}gh_{J_2}\right)\\
&\le\int_{\mathbb{R}^{d_1}}\sum_{\substack{V\in\mathcal{D}^2\\ J_1^{j_1}=J_2^{j_2}=V}}\frac{\ell(V)^{\alpha_2}}{|V|}\int_{J_1}\mathsf{M}^1\Delta_V^{j_1}f\int_{J_2}\mathsf{S}^1\Delta_V^{j_2}g\\
&=\int_{\mathbb{R}^{d_1}}\sum_{V\in\mathcal{D}^2}\frac{\ell(V)^{\alpha_2}}{|V|}\int_V\mathsf{M}^1\Delta_V^{j_1}f\int_V\mathsf{S}^1\Delta_V^{j_2}g\\
&=\int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_2}}\sum_{V\in\mathcal{D}^2}\ell(V)^{\alpha_2}\langle\mathsf{M}^1\Delta_V^{j_1}f\rangle_V\langle\mathsf{S}^1\Delta_V^{j_2}g\rangle_V1_V\\
&\le\left\|\left(\sum_{V\in\mathcal{D}^2}\left(\ell(V)^{\alpha_2}\langle\mathsf{M}^1\Delta_V^{j_1}f\rangle_V\right)^2\!1_V\right)^{\frac12}\right\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}}\left\|\left(\sum_{V\in\mathcal{D}^2}\langle\mathsf{S}^1\Delta_V^{j_2}g\rangle_V^2\,1_V\right)^{\frac12}\right\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}}\\
&\le\left\|\left(\sum_{V\in\mathcal{D}^2}\left(\mathsf{M}^{\alpha_2}\left(\mathsf{M}^1\Delta_V^{j_1}f1_V\right)\right)^2\right)^{\frac12}\right\|_{L^{p_1}_{x_1}L^{q_2}_{x_2}}\left\|\left(\sum_{V\in\mathcal{D}^2}\left(\mathsf{M}^2\left(\mathsf{S}^1\Delta_V^{j_2}g1_V\right)\right)^2\right)^{\frac12}\right\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}}\\
&\lesssim\left\|\left(\sum_{V\in\mathcal{D}^2}|\Delta_V^{j_1}f|^2\right)^{\frac12}\right\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}\left\|\left(\sum_{V\in\mathcal{D}^2}\left(\mathsf{S}^1\Delta_V^{j_2}g\right)^2\right)^{\frac12}\right\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}}\\
&\lesssim\|\mathsf{S}^2f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}\|\mathsf{S}g\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}}\lesssim\|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}\|g\|_{L^{p_1'}_{x_1}L^{q_2'}_{x_2}}.
\end{aligned}$$

□

Proof of Theorem 4.3, part 3/3, full paraproducts. Now, let $i = j = (0,0)$ and we consider the paraproduct

$$\langle S^{(0,0),(0,0)} f, g\rangle = \sum_{K\in\mathcal{D}^1}\sum_{V\in\mathcal{D}^2} \alpha_{KV}\langle f\rangle_{K\times V}\langle g, h_K\otimes h_V\rangle.$$

Writing out the main term we find out that the summand (without the scaling factor $\alpha_{(I)(J)KV}$ in front) in $\langle [b, S^{i,j}]f, g\rangle$ is

$$\begin{aligned}&\big[\langle f\rangle_{K\times V}\langle bg, h_K\otimes h_V\rangle - \langle bf\rangle_{K\times V}\langle g, h_K\otimes h_V\rangle\big]\\ &= \sum_{i=1,2}\langle f\rangle_{K\times V}\langle A_i(b,g), h_K\otimes h_V\rangle\\ &\quad + \big[\langle f\rangle_{K\times V}\langle A_3(b,g), h_K\otimes h_V\rangle - \langle bf\rangle_{K\times V}\langle g, h_K\otimes h_V\rangle\big].\end{aligned}$$

The terms with the first two summands are bounded by the mixed norm estimates of bi-parameter model operators and Lemma 4.6, as before, and the bracketed difference on the last line writes out to reduce us to bounding the form

$$\sum_{K\in\mathcal{D}^1}\sum_{V\in\mathcal{D}^2} \alpha_{KV}\big\langle (\langle b\rangle_V - b)f\big\rangle_{K\times V}\langle g, h_K\otimes h_V\rangle.$$

This is bounded by the known boundedness of the model operator and M^{α_2} and the observation that

$$\begin{aligned}\Big|\big\langle (\langle b\rangle_V - b)f\big\rangle_{K\times V}\Big| &\le \big\langle |b-\langle b\rangle_V||f|\big\rangle_{K\times V} \le \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\langle \ell(V)^{\alpha_2}|f|\rangle_{K\times V}\\ &= \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\langle \ell(V)^{\alpha_2}\langle |f|\rangle_V\rangle_{K\times V} \le \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\langle \mathsf{M}^{\alpha_2} f\rangle_{K\times V}.\end{aligned}$$

Now consider the commutator taken with the other paraproduct with the summands being

$$\left\langle f, h_K\otimes\frac{1_V}{|V|}\right\rangle\left\langle bg, \frac{1_K}{|K|}\otimes h_V\right\rangle - \left\langle bf, h_K\otimes\frac{1_V}{|V|}\right\rangle\left\langle g, \frac{1_K}{|K|}\otimes h_V\right\rangle.$$

Again, going through with our decomposition strategy, we reduce to the operator that originates as a difference through the auxiliary operator A_3,

$$\sum_{K\in\mathcal{D}^1}\sum_{V\in\mathcal{D}^2} \alpha_{KV}\left\langle (\langle b\rangle_V - b)f, h_K\otimes\frac{1_V}{|V|}\right\rangle\left\langle g, \frac{1_K}{|K|}\otimes h_V\right\rangle.$$

Again, this is bounded by the known boundedness of the model operator and the following observations, we have

$$\begin{aligned}\left|\left\langle (\langle b\rangle_V - b)f, h_K\otimes\frac{1_V}{|V|}\right\rangle\right| &\le \left\langle |\langle b\rangle_V - b||\langle f, h_K\rangle|h_K, h_K\otimes\frac{1_V}{|V|}\right\rangle\\ &\le \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\left\langle \ell(V)^{\alpha_2}|\langle f, h_K\rangle|h_K, h_K\otimes\frac{1_V}{|V|}\right\rangle\\ &= \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\left\langle \ell(V)^{\alpha_2}\big\langle |\langle f, h_K\rangle|\big\rangle_V h_K, h_K\otimes\frac{1_V}{|V|}\right\rangle\\ &= \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\left\langle \sum_{L\in\mathcal{D}^1}\ell(V)^{\alpha_2}\big\langle |\langle f, h_L\rangle|\big\rangle_V h_L, h_K\otimes\frac{1_V}{|V|}\right\rangle\\ &\le \|b\|_{\dot C^{0,\alpha_2}_{x_2}(\mathbb{R}^{d_2})}\left\langle \sum_{L\in\mathcal{D}^1} h_L\otimes\mathsf{M}^{\alpha_2}\langle f, h_L\rangle, h_K\otimes\frac{1_V}{|V|}\right\rangle,\end{aligned}$$

and this time we are done as soon as we show that $\Phi f = \sum_{L\in\mathcal{D}^1} h_L \otimes \mathsf{M}^{\alpha_2}\langle f, h_L\rangle$ satisfies the correct bound. For this, by duality it is enough to estimate as follows

$$\begin{aligned}
|\langle \Phi f, g\rangle| &\le \left\| \left(\sum_{L\in\mathcal{D}^1} (\mathsf{M}^{\alpha_2}\langle f, h_L\rangle)^2 \frac{1_L}{|L|} \right)^{1/2} \right\|_{L^{p_1}_{x_2} L^{q_2}_{x_2}} \left\| \left(\sum_{L\in\mathcal{D}^1} |\langle g, h_L\rangle|^2 \frac{1_L}{|L|} \right)^{1/2} \right\|_{L^{p_1'}_{x_2} L^{q_2'}_{x_2}} \\
&\lesssim \left\| \left(\sum_{L\in\mathcal{D}^1} |\langle f, h_L\rangle|^2 \frac{1_L}{|L|} \right)^{1/2} \right\|_{L^{p_1}_{x_2} L^{p_2}_{x_2}} \left\| \left(\sum_{L\in\mathcal{D}^1} |\langle g, h_L\rangle|^2 \frac{1_L}{|L|} \right)^{1/2} \right\|_{L^{p_1'}_{x_2} L^{q_2'}_{x_2}} \\
&= \big\|\mathsf{S}^1 f\big\|_{L^{p_1}_{x_2} L^{p_2}_{x_2}} \big\|\mathsf{S}^1 g\big\|_{L^{p_1'}_{x_2} L^{q_2'}_{x_2}} \lesssim \big\|f\big\|_{L^{p_1}_{x_2} L^{p_2}_{x_2}} \big\|g\big\|_{L^{p_1'}_{x_2} L^{q_2'}_{x_2}},
\end{aligned}$$

where we again used lemmas 4.10 and 4.12. □

5. Upper bound for the case $p_1 < q_1$, $p_2 = q_2$

To treat this case, it is better to work with an alternative definition of bi-parameter CZOs. By Grau de la Herrán [5] an equivalent way to defining bi-parameter Calderón–Zygmund operators as by Martikainen [14] is the one by Journé [11]. The definition of Martikainen follows quickly from Journe's and the main result in [5] is the reverse direction.

5.1. Definition. (Journé) A pair $K = (K_1, K_2)$ of kernels is said to be a bi-parameter CZ-kernel if for $j \in \{1,2\}$ and $i \in \{1,2\} \setminus \{j\}$ the kernels map

$$K_j(x_i, y_i)\colon \mathbb{R}^{d_i} \times \mathbb{R}^{d_i} \setminus \Delta \to \mathrm{CZO}(d_j, \delta),$$

satisfy the bounds

$$\big\|K_j(x_i, y_i)\big\|_{\mathrm{CZO}(d_j,\delta)} \le C|x_i - y_i|^{-d_i},$$

and

$$\big\|K_j(x_i, y_i) - K_j(x_i', y_i)\big\|_{\mathrm{CZO}(d_j,\delta)} + \big\|K_j(y_i, x_i) - K_j(y_i, x_i')\big\|_{\mathrm{CZO}(d_j,\delta)} \le C\frac{|x_i - x_i'|^\delta}{|x_i - y_i|^{d_i+\delta}},$$

whenever $|x_i - x_i'| \le 1/2|x_i - y_i|$.

An operator T with a bi-parameter CZ-kernel is said to be a bi-parameter SIO if for $i \in \{1,2\}$ and $j \in \{1,2\} \setminus \{i\}$ we have

$$\big\langle T(f_1 \otimes f_2), g_1 \otimes g_2\big\rangle = \int_{\mathbb{R}^{d_j}} \int_{\mathbb{R}^{d_j}} \langle K_i(x_j, y_j) f_i, g_i\rangle f_j(y_j) g_j(x_j)\,\mathrm{d}y_j\,\mathrm{d}x_j,$$

whenever $\mathrm{spt}(f_j) \cap \mathrm{spt}(g_j) = \emptyset$ and $f_k, g_k \in \Sigma_k$ for $k \in \{i, j\}$.

The dual T^{1*} of T is given by the identity $\big\langle T^{1*}(f_1 \otimes f_2), g_1 \otimes g_2\big\rangle = \big\langle T(g_1 \otimes f_2), f_1 \otimes g_2\big\rangle$. It is straightforward to see that T^{1*} is a bi-parameter SIO if T is and that the kernels of T^{1*} are given by $K_1^{1*}(x_2, y_2) = K_1^*(x_2, y_2)$ and $K_2^{1*}(x_1, y_1) = K_2(x_1, y_1)$.

5.2. Definition. A bi-parameter SIO as in Definition 5.1 is a bi-parameter CZO if T and T^{1*} are bounded on $L^2(\mathbb{R}^d)$.

The advantage with this setup is that we can now easily prove the following.

5.3. Lemma. *Let T be a bi-parameter CZO. Suppose that $b(x_1, \cdot) =$ constant and $b(\cdot, x_2) \in L^\infty_{\mathrm{loc},x_1}$. Then, for all $f, g \in \Sigma$ we have*

$$\big\langle [b, T] f, g\big\rangle = \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_1}} (b(x_1) - b(y_1))\big\langle K_2(x_1, y_1) f(y_1, \cdot), g(x_1, \cdot)\big\rangle\,\mathrm{d}y_1\,\mathrm{d}x_1,$$

where we denote $b(x_1) = b(x_1, v)$ for some choice of $v \in \mathbb{R}^{d_2}$.

Proof. We first consider the one-parameter setting with the one-parameter space $\mathbb{R}^n$ and let $b \in L^\infty_{\mathrm{loc}}(\mathbb{R}^n)$. It is a basic part of the one-parameter theory (see e.g. Grafakos [4], Proposition 4.1.11.) that for each $T \in \mathrm{CZO}(n, \delta)$ there exists $\widetilde{T} \in \mathrm{CZO}(n, \delta)$ and a function $m \in L^\infty$ so that

$$(T - m)h = \widetilde{T}h, \quad \widetilde{T}h(x) = \lim_{k\to\infty} \int_{|x-y|>\varepsilon_k} K(x, y)h(y)\,\mathrm{d}y$$

where K is the kernel of T and the limit holds along some sequence $\varepsilon_k \to 0$ and for all bounded and compactly supported functions h.

The above immediately gives the following: suppose that $b \in L^\infty_{\mathrm{loc}}(\mathbb{R}^n)$ and $f, g \in \Sigma_n$, so that

$$\begin{aligned} \big\langle [b, T]f, g\big\rangle &= \big\langle [b, T - m]f, g\big\rangle = \big\langle [b, \widetilde{T}]f, g\big\rangle \\ &= \int_{\mathbb{R}^n} \lim_{\varepsilon_k\to 0} \int_{|x-y|>\varepsilon_k} (b(x) - b(y))K(x, y)f(y)g(x)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (b(x) - b(y))K(x, y)f(y)g(x)\,\mathrm{d}y\,\mathrm{d}x. \end{aligned} \tag{5.4}$$

The last step follows by the dominated convergence theorem after the following estimate uniform in ε_k,

$$\int |(b(x) - b(y))K(x, y)f(y)|\,\mathrm{d}y \lesssim \|b\|_{\dot{C}^{0,\alpha}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x - y|^{\alpha-d}|f(y)|\,\mathrm{d}y;$$

since f is bounded and compactly supported, we see that the right-hand side is finite. Now with this one-parameter result at hand, we turn to the claim itself.

By linearity it is enough to prove the claim for functions $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ of tensor form. If T is an SIO as by Journé, then the size estimate

$$|\langle K_2(x_1, y_1)f_2, g_2\rangle| \lesssim |x_1 - y_1|^{d_1}\|f_2\|_{L^p}\|g_2\|_{L^{p'}}$$

is satisfied, and similarly immediately from the definitions the regularity estimates also hold. Consequently, since T is bounded, the function $(x_1, y_1) \mapsto \langle K_2(x_1, y_1)f_2, g_2\rangle$ is a kernel of the one-parameter CZO defined by

$$\langle T_{f_2,g_2}f_1, g_1\rangle = \langle T(f_1 \otimes f_2), g_1 \otimes g_2\rangle.$$

Then, it follows by the one-parameter result (5.4) that

$$\begin{aligned} &\big\langle [b(\cdot, v), T_{f_2,g_2}]f_1, g_1\big\rangle \\ &= \int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}} (b(x_1, v) - b(y_1, v))\big\langle K_2(x_1, y_1)f_2, g_2\big\rangle f_1(y_1)g_1(x_1)\,\mathrm{d}y_1\,\mathrm{d}x_1, \end{aligned}$$

where we note that $b(\cdot, v) \in L^\infty_{\mathrm{loc},x_1}$. However, we also have

$$\begin{aligned} \big\langle [b, T](f_1 \otimes f_2), g_1 \otimes g_2\big\rangle &= \big\langle b(\cdot, v)T(f_1 \otimes f_2) - T(b(\cdot, v)(f_1 \otimes f_2), g_1 \otimes g_2\big\rangle \\ &= \big\langle T_{f_2,g_2}f_1, b(\cdot, v)g_1\big\rangle - \big\langle T_{f_2,g_2}(b(\cdot, v)f_1), g_1\big\rangle \\ &= \big\langle [b(\cdot, v), T_{f_2,g_2}]f_1, g_1\big\rangle, \end{aligned}$$

and thus the claim follows. □

5.5. Proposition. *Let $p_1 < q_1$ and $p_2 = q_2$, let T be a bi-parameter CZO and suppose that $b(x_1, \cdot) = $ constant and $b(\cdot, x_2) \in \dot{C}^{0,\alpha_1}_{x_1}$. Then, we have*

$$\|[b,T]f\|_{L^{q_1}_{x_1}L^{p_2}_{x_2}} \lesssim \|b(\cdot,x_2)\|_{\dot{C}^{0,\alpha}_{x_1}} \|f\|_{L^{p_1}_{x_1}L^{p_2}_{x_2}}.$$

Proof. It is enough to prove the claim for functions in a dense subset of the space $L^{p_1}_{x_1}L^{p_2}_{x_2}$ and clearly Σ is such a subset. As $b(\cdot,x_2) \in \dot{C}^{0,\alpha_1}_{x_1}$, especially $b(\cdot,x_2) \in L^\infty_{\mathrm{loc},x_1}$ and thus by Lemma 5.3 we can estimate

$$\begin{aligned}
|\langle [b,T]f, g\rangle| &\le \int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}} \Big|(b(x_1,v)-b(y_1,v))\Big\langle K_2(x_1,y_1)f(y_1,\cdot)(z), g(x_1,z)\Big\rangle_z\Big| \,\mathrm{d}y_1\,\mathrm{d}x_1 \\
&\lesssim \int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}} \frac{|b(x_1,v)-b(y_1,v)|}{|x_1-y_1|^{d_1}} \big\|f(y_1,z)\big\|_{L^{p_2}_z} \big\|g(x_1,z)\big\|_{L^{p_2'}_z} \,\mathrm{d}y_1\,\mathrm{d}x_1 \\
&\le \big\|b(\cdot,v)\big\|_{\dot{C}^{0,\alpha_1}_{x_1}} \int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}} |x_1-y_1|^{\alpha_1-d_1} \big\|f(y_1,z)\big\|_{L^{p_2}_z} \big\|g(x_1,z)\big\|_{L^{p_2'}_z} \,\mathrm{d}y_1\,\mathrm{d}x_1 \\
&= \big\|b(\cdot,v)\big\|_{\dot{C}^{0,\alpha_1}_{x_1}} \int_{\mathbb{R}^{d_1}} \mathsf{I}_{\alpha_1}\Big(\big\|f(\cdot,z)\big\|_{L^{p_2}_z}\Big)(x_1)\cdot \big\|g(x_1,z)\big\|_{L^{p_2'}_z} \,\mathrm{d}x_1 \\
&\lesssim \big\|b(\cdot,v)\big\|_{\dot{C}^{0,\alpha_1}_{x_1}} \big\|f(y_1,z)\big\|_{L^{p_1}_{y_1}L^{p_2}_z} \big\|g(x_1,z)\big\|_{L^{q_1'}_{x_1}L^{p_2'}_z},
\end{aligned}$$

where in the last step we used the boundedness of the fractional integral. □

References

[1] Airta, E., T. Hytönen, K. Li, H. Martikainen, and T. Oikari: Off-diagonal estimates for bi-commutators. - Int. Math. Res. Not. IMRN 2022:23, 2022, 18766–18832.

[2] Coifman, R., P.-L. Lions, Y. Meyer, and S. Semmes: Compensated compactness and Hardy spaces. - J. Math. Pures Appl. (9) 72:3, 1993, 247–286.

[3] Coifman, R., R. Rochberg, and G. Weiss: Factorization theorems for Hardy spaces in several variables. - Ann. of Math. (2) 103:3, 1976, 611–635.

[4] Grafakos, L.: Modern Fourier analysis. - Grad. Texts in Math. 250, 3rd edition, New York, NY, 2014.

[5] Grau de la Herrán, A.: Comparison of $T1$ conditions for multi-parameter operators. - Proc. Amer. Math. Soc. 144:6, 2016, 2437–2443.

[6] Holmes, I., S. Petermichl, and B. D. Wick: Weighted little bmo and two-weight inequalities for Journé commutators. - Anal. PDE 11:7, 2018, 1693–1740.

[7] Hytönen, T.: The L^p-to-L^q boundedness of commutators with applications to the Jacobian operator. - J. Math. Pures Appl. 156, 2021, 351–391.

[8] Hytönen, T., H. Martikainen, and E. Vuorinen: Multi-parameter estimates via operator-valued shifts. - Proc. Lond. Math. Soc. 119:6, 2019, 1560–1597.

[9] Iwaniec, T.: Nonlinear commutators and Jacobians. - J. Fourier Anal. Appl. 3, 1997, 775–795.

[10] Janson, S.: Mean oscillation and commutators of singular integral operators. - Ark. Mat. 16, 1978, 263–270.

[11] Journé, J.-L.: Calderón–Zygmund operators on product spaces. - Rev. Mat. Iberoam. 1, 1985, 55–91.

[12] Li, K., H. Martikainen, and E. Vuorinen: Bloom-type inequality for bi-parameter singular integrals: efficient proof and iterated commutators. - Int. Math. Res. Not. IMRN 2021:11, 2021, 8153–8187.

[13] Lindberg, S.: On the Hardy space theory of compensated compactness quantities. - Arch. Ration. Mech. Anal. 224, 2017, 709–742.

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

[15] Nehari, Z.: On bounded bilinear forms. - Ann. of Math. (2) 65, 1957, 153–162.

[16] Ruiz Blasco, F. J., and J. L. Torrea Hernandez: Weighted and vector-valued inequalities for potential operators. - Trans. Amer. Math. Soc. 295:1, 1986, 213–232.

[17] Uchiyama, A.: On the compactness of operators of Hankel type. - Tohoku Math. J. 30:1, 1978, 163–171.

Received 2 February 2022 • Accepted 12 December 2022 • Published online 31 December 2022

Tuomas Oikari
University of Jyväskylä
Department of Mathematics and Statistics
PO Box 35, FI-40014 University of Jyväskylä, Finland
tuomas.oikari@gmail.com