The Journal of Geometric Analysis (2023) 33:157
https://doi.org/10.1007/s12220-023-01216-x

Exotic Calderón–Zygmund Operators

Tuomas Hytönen[1] · **Kangwei Li**[2] · **Henri Martikainen**[3] · **Emil Vuorinen**[1]

Received: 2 August 2022 / Accepted: 2 February 2023 / Published online: 28 February 2023

Abstract

We study singular integral operators with kernels that are more singular than standard Calderón–Zygmund kernels, but less singular than bi-parameter product Calderón–Zygmund kernels. These kernels arise as restrictions to two dimensions of certain three-dimensional kernels adapted to so-called Zygmund dilations, which is part of our motivation for studying these objects. We make the case that such kernels can, in many ways, be seen as part of the extended realm of standard kernels by proving that they satisfy both a $T1$ theorem and commutator estimates in a form reminiscent of the corresponding results for standard Calderón–Zygmund kernels. However, we show that one-parameter weighted estimates, in general, fail.

Keywords Singular integrals · Multi-parameter analysis · Zygmund dilations · Multiresolution analysis

Mathematics Subject Classification 42B20

✉ Emil Vuorinen
emil.vuorinen@helsinki.fi

Tuomas Hytönen
tuomas.hytonen@helsinki.fi

Kangwei Li
kli@tju.edu.cn

Henri Martikainen
henri@wustl.edu

1 Department of Mathematics and Statistics, University of Helsinki, P.O.B. 68, 00014 Helsinki, Finland

2 Center for Applied Mathematics, Tianjin University, Weijin Road 92, 300072 Tianjin, China

3 Department of Mathematics and Statistics, Washington University in St. Louis, 1 Brookings Drive, St. Louis, MO 63130, USA

1 Introduction

Working on the Euclidean product space $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, we define for $x = (x^1, x^2)$ and $y = (y^1, y^2)$ the decay factor

$$D_\theta(x, y) := \left(\frac{|x^1 - y^1|}{|x^2 - y^2|} + \frac{|x^2 - y^2|}{|x^1 - y^1|} \right)^{-\theta} < 1, \qquad \theta \in (0, 1], \tag{1.1}$$

whenever $x^1 \neq y^1$ and $x^2 \neq y^2$. Notice that this decay factor becomes larger and larger as θ shrinks. The point is that when $\theta = 1$ it is at its smallest, and then

$$\frac{1}{|x^1 - y^1|} \frac{1}{|x^2 - y^2|} D_1(x, y) = \frac{1}{|x^1 - y^1|^2 + |x^2 - y^2|^2} = \frac{1}{|x - y|^2}.$$

That is, in this case, the bi-parameter size estimate multiplied with this decay factor yields the usual one-parameter size estimate. When $\theta < 1$, the decay factor is larger and the corresponding product is something between the bi-parameter and one-parameter size estimate.

We say that kernels that decay like

$$\frac{1}{|x^1 - y^1|} \frac{1}{|x^2 - y^2|} D_\theta(x, y)$$

for some θ and satisfy some similar continuity estimates are CZX kernels—one can pronounce the "X" in "CZX" as "exotic". Such kernels are more singular than the standard Calderón–Zygmund kernels, but less singular than the product Calderón–Zygmund(–Journé) kernels [5, 13, 18]. Even with $\theta = 1$, they are different from the standard Calderón–Zygmund kernels—in this case, the difference is only in the Hölder estimates (see Sect. 2). The CZX kernels can, for example, be motivated by looking at Zygmund dilations [4, 19–21]. Zygmund dilations are a group of dilations lying in between the standard product theory and the one-parameter setting—in $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R}^2$ they are the dilations $(x_1, x_2, x_3) \mapsto (\delta_1 x_1, \delta_2 x_2, \delta_1 \delta_2 x_3)$. Recently, in [8] and subsequently in [3, 9] general convolution form singular integrals invariant under Zygmund dilations were studied. In these papers the decay factor

$$t \mapsto \left(t + \frac{1}{t} \right)^{-\theta}$$

controls the additional, compared to the product setting, decay with respect to the Zygmund ratio

$$\frac{|x_1 x_2|}{|x_3|}.$$

See also our recent paper [12] which attacks the Zygmund setting from the point of view of new multiresolution methods. Essentially, in the current paper, we isolate the conditions on the lower-dimensional kernels obtained by fixing the variables x^1, y^1 in the Zygmund setting [8, 12] and ignoring the dependence on these variables. A class of CZX operators is also induced by the Fefferman–Pipher multipliers [4]—importantly, they satisfy $\theta = 1$ but with an additional logarithmic growth factor. This subtle detail has a key relevance for the weighted estimates as we explain below.

There is a useful operator-valued viewpoint to multi-parameter analysis—Journé [13] views, e.g. bi-parameter operators as "operator-valued one-parameter operators". For recent work using this viewpoint see e.g. [11]. Developing such an approach to Zygmund SIOs is interesting. The operator-valued viewpoint is useful for example when proving the necessity of $T1$ type assumptions in the product setting, see e.g. [6], and the full product BMO type $T1$ theory of Zygmund SIOs is still to be developed. The operator-valued approach will necessarily be complicated in the Zygmund setting, since the parameters are tied and it is not as simple as fixing a single variable. Our new exotic operators are pertinent to the operator-valued viewpoint, where Zygmund SIOs could partly be seen as operator-valued one-parameter operators the values being exotic operators.

It has been known for a long time that Calderón–Zygmund operators act boundedly in the weighted spaces $L^p(w)$ whenever w belongs to the Muckenhoupt class A_p, defined by the finiteness of the weight constant

$$[w]_{A_p} := \sup_J \langle w \rangle_J \langle w^{-1/(p-1)} \rangle_J^{p-1},$$

where the supremum is over all cubes J. On the other hand, the more singular multi-parameter Calderón–Zygmund(–Journé) operators in general satisfy such bounds only for the smaller class of strong A_p weights, defined by $[w]_{A_p^*}$, where the supremum is over all axes-parallel rectangles. While on a general level, the CZX operators behave quite well with any θ, even with $\theta < 1$, for one-parameter weighted estimates it is critical that $\theta = 1$, the aforementioned logarithmic extra growth being allowed.

1.2 Theorem *Let $T \in \mathcal{L}(L^2(\mathbb{R}^2))$ be an operator with a CZX kernel.*

1. *If $\theta < 1$ in (1.1), one-parameter weighted estimates may fail.*
2. *If $\theta = 1$ in (1.1), possibly with a logarithmic growth factor, then for every $p \in (1, \infty)$ and every $w \in A_p(\mathbb{R}^2)$ the operator T extends boundedly to $L^p(w)$.*

In the paper [12], we also develop the corresponding counterexamples in the full Zygmund case. There the interest is whether Zygmund singular integrals are weighted bounded with respect to the Zygmund weights—a larger class than the strong A_p with the supremum running only over the so-called Zygmund rectangles satisfying the natural scaling. For $\theta < 1$, the situation parallels the one from the CZX world—they need not be weighted bounded with respect to the Zygmund weights.

Apart from the weighted estimates, we want to make the case that, in many ways, the CZX kernels with an arbitrary θ can be seen as part of the extended realm of standard kernels, rather than the more complicated product theory. In particular, the $T1$ theorem for CZX kernels takes the following form reminiscent of the standard $T1$ theorem [1].

1.3 Theorem *Let $B(f,g)$ be a bilinear form defined on finite linear combinations of indicators of cubes of $\mathbb{R}^2$, and such that*

$$B(f,g)=\iint K(x,y)f(y)g(x)\,\mathrm{d}x\,\mathrm{d}y$$

when $\{f\neq 0\}\cap\{g\neq 0\}=\varnothing$, where $K\in CZX(\mathbb{R}^2)$. Then the following are equivalent:

(1) *There is a bounded linear $T\in\mathcal{L}(L^2(\mathbb{R}^2))$ such that $\langle Tf,g\rangle=B(f,g)$.*
(2) *B satisfies*

- *the weak boundedness property $|B(1_I,1_I)|\lesssim|I|$ for all cubes $I\subset\mathbb{R}^2$, and*
- *the $T(1)$ conditions*

$$B(1,g)=\int b_1 g,\qquad B(f,1)=\int b_2 f$$

for some $b_1,b_2\in\mathrm{BMO}(\mathbb{R}^2)$ and all f,g with $\int f=0=\int g$.

Moreover, under these conditions,

(3) *T defines a bounded operator from $L^\infty(\mathbb{R}^2)$ to $\mathrm{BMO}(\mathbb{R}^2)$, from $L^1(\mathbb{R}^2)$ to $L^{1,\infty}(\mathbb{R}^2)$, and on $L^p(\mathbb{R}^2)$ for every $p\in(1,\infty)$.*

In fact, our proof also gives a representation of $B(f,g)$, Theorem 4.9, which includes both one-parameter [10] and bi-parameter [18] elements. The following commutator bounds follow from the representation; however, the argument is not entirely standard due to the hybrid nature of the model operators.

1.4 Theorem *Let $T\in\mathcal{L}(L^2(\mathbb{R}^2))$ be an operator associated with a CZX kernel K. Then*

$$\|[b,T]f\|_{L^p}\lesssim\|b\|_{\mathrm{BMO}}\|f\|_{L^p}$$

whenever $p\in(1,\infty)$. Here $[b,T]f:=bTf-T(bf)$.

Thus, the commutator estimate holds with the one-parameter BMO space. This is another purely one-parameter feature of these exotic operators. As the weighted estimates do not, in general, hold, the commutator estimate cannot be derived from the well-known Cauchy integral trick.

Over the past several years, a standard approach to weighted norm inequalities has been via the methods of sparse domination pioneered by Lerner. For $\theta=1$, we can derive our weighted estimates directly from our representation theorem. However, we also provide some additional sparse estimates that give a solid quantitative dependence on the A_p constant and yield two-weight commutator estimates for free.

1.5 Theorem *Let $T\in\mathcal{L}(L^2(\mathbb{R}^2))$ be an operator with a CZX kernel with $\theta=1$. Then for every $p\in(1,\infty)$ and every $w\in A_p(\mathbb{R}^2)$ the operator T extends boundedly to $L^p(w)$ with norm*

$$\|T\|_{\mathcal{L}(L^p(w))}\lesssim_p[w]_{A_p}^{p'}.$$

Moreover, if $\nu = w^{\frac{1}{p}}\lambda^{-\frac{1}{p}}$ *with* $w, \lambda \in A_p$ *and*

$$\|b\|_{\mathrm{BMO}_\nu} := \sup_I \frac{1}{\nu(I)} \int_I |b - \langle b \rangle_I| < \infty,$$

where the supremum is over cubes $I \subset \mathbb{R}^2$, *then*

$$\|[b, T]\|_{L^p(w) \to L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}_\nu}.$$

The quantitative bound (in particular quadratic in $[w]_{A_2}$ when $p = 2$) is worse than the linear A_2 theorem valid for classical Calderón–Zygmund operators [10].

We conclude the introduction with an outline of how the paper is organized. In Sect. 2, we define the CZX kernels and prove part of Theorem 1.3 in Proposition 2.4. Section 3 begins with the definition of CZX forms. Lemma 3.3 proves estimates for CZX forms acting on Haar functions, which will be used in the representation theorem, Theorem 4.9. In Proposition 4.4, we prove certain weighted maximal function estimates, which are at the heart of proving that CZX forms with decay parameter $\theta_2 = 1$ satisfy weighted estimates. The dyadic operators used to represent CZX forms are defined in Definition 4.5, and estimates for them are proved in Lemma 4.6. The representation identity and the $T1$ theorem, and the weighted estimates when $\theta_2 = 1$, for CZX forms are recorded in Theorem 4.9. Theorem 1.4 is proved in Sect. 5. In the beginning of Sect. 6, we construct the counterexamples required to prove (1) of Theorem 1.2. The sparse domination of CZX operators with $\theta_2 = 1$ is recorded in Corollary 6.7. Theorem 1.5 is proved in Corollary 6.8 and in the discussion after Proposition 6.10.

2 CZX Kernels

We work in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Let $\theta_1, \theta_2 \in (0, 1]$. For $x^1 \neq y^1$ and $x^2 \neq y^2$ define

$$D_{\theta_2}(x, y) := \left(\frac{|x^1 - y^1|}{|x^2 - y^2|} + \frac{|x^2 - y^2|}{|x^1 - y^1|} \right)^{-\theta_2} < 1.$$

We assume that the kernel $K : \mathbb{R}^2 \backslash \{x^1 = y^1 \text{ or } x^2 = y^2\} \to \mathbb{C}$ satisfies the size estimate

$$|K(x, y)| \lesssim \frac{1}{|x^1 - y^1|} \frac{1}{|x^2 - y^2|} D_{\theta_2}(x, y)$$

and the mixed Hölder and size estimate

$$|K(x, y) - K((w^1, x^2), y)| \lesssim \frac{|x^1 - w^1|^{\theta_1}}{|x^1 - y^1|^{1+\theta_1}} \frac{1}{|x^2 - y^2|} D_{\theta_2}(x, y)$$

whenever $|x^1 - w^1| \le |x^1 - y^1|/2$, together with the other three symmetric mixed Hölder and size estimates. If this is the case, we say that $K \in CZX(\mathbb{R}^2)$. Again, such kernels are more singular than standard Calderón–Zygmund kernels, but less singular than the product Calderón–Zygmund(–Journé) kernels. See Remark 4.10 for some additional logarithmic factors when $\theta_2 = 1$ and why they are relevant from the point of view of Fefferman–Pipher multipliers [4].

2.1 Lemma *Let $K \in CZX(\mathbb{R}^2)$ and $x^1, x^2, y^2 \in \mathbb{R}$. Then*

$$\int_{\mathbb{R}} |K(x,y)|\,\mathrm{d}y^1 \lesssim \frac{1}{|x^2 - y^2|}.$$

Also, for $L > 0$ there holds that

$$\int_{\{y^1 :\, |x^1-y^1| \lesssim L\}} |K(x,y)|\,\mathrm{d}y^1 \lesssim \frac{L^{\theta_2}}{|x^2 - y^2|^{1+\theta_2}},$$

which is a useful estimate if $L \lesssim |x^2 - y^2|$.

Proof By elementary calculus

$$\begin{aligned}
\int_{\mathbb{R}} |K(x,y)|\,\mathrm{d}y^1 &\lesssim \frac{1}{|x^2 - y^2|} \int_0^\infty \frac{1}{u}\Big(\frac{u}{|x^2-y^2|} + \frac{|x^2-y^2|}{u}\Big)^{-\theta_2}\,\mathrm{d}u \\
&\lesssim \frac{1}{|x^2-y^2|}\Big(\int_0^{|x^2-y^2|} \frac{\mathrm{d}u}{u^{1-\theta_2}|x^2-y^2|^{\theta_2}} + \int_{|x^2-y^2|}^\infty \frac{\mathrm{d}u}{u^{1+\theta_2}|x^2-y^2|^{-\theta_2}}\Big) \\
&\lesssim \frac{1}{|x^2-y^2|},
\end{aligned}$$

and the logic for the second estimate is also clear from this. □

The first sharper estimate in the next lemma is only needed to derive the weighted estimates in the case $\theta_2 = 1$.

2.2 Lemma *Let $K \in CZX(\mathbb{R}^2)$ and $J = J^1 \times J^2 \subset \mathbb{R}^2$ be a square with centre $c_J = (c_{J^1}, c_{J^2})$. If $x \in J$ and $y \in (3J^1)^c \times (3J^2)^c$, then*

$$\begin{aligned}
|K(x,y) - K(c_J, y)| &\lesssim \prod_{i=1}^2 \frac{1}{\operatorname{dist}(y^i, J^i)} \times \frac{\ell(J)^{\theta_1}(\min_{i=1,2}\operatorname{dist}(y^i, J^i))^{\theta_2-\theta_1}}{(\max_{i=1,2}\operatorname{dist}(y^i, J^i))^{\theta_2}} \\
&\lesssim \prod_{i=1}^2 \frac{\ell(J)^\theta}{\operatorname{dist}(y^i, J^i)^{1+\theta}}, \qquad \theta := \frac{1}{2}\min(\theta_1, \theta_2).
\end{aligned}$$

Proof There holds that

$$\begin{aligned}|K(x,y)-K(c_J,y)| &\le |K(x^1,x^2,y)-K(c_{J^1},x^2,y)|\\ &+|K(c_{J^1},x^2,y)-K(c_{J^1},c_{J^2},y)|.\end{aligned}$$

Since $2|x^i - c_{J^i}| \le \ell(J) \le \operatorname{dist}(y^i, J^i) \le \min(|y^i - x^i|, |y^i - c_{J^i}|)$, we conclude

$$\begin{aligned}|K(x^1,x^2,y)-K(c_{J^1},x^2,y)| &\lesssim \frac{|x^1-c_{J^1}|^{\theta_1}}{|x^1-y^1|^{1+\theta_1}}\frac{1}{|x^2-y^2|}D_{\theta_2}(x,y),\\ |K(c_{J^1},x^2,y)-K(c_{J^1},c_{J^2},y)| &\lesssim \frac{1}{|c_{J^1}-y^1|}\frac{|x^2-c_{J^2}|^{\theta_1}}{|c_{J^2}-y^2|^{1+\theta_1}}D_{\theta_2}(c_J,y).\end{aligned}$$

Suppose for instance that $\operatorname{dist}(y^1, J^1) \ge \operatorname{dist}(y^2, J^2)$. Then the sum simplifies to

$$|K(x,y)-K(c_J,y)| \lesssim \frac{1}{\operatorname{dist}(y^1,J^1)}\frac{\ell(J)^{\theta_1}}{\operatorname{dist}(y^2,J^2)^{1+\theta_1}}\Big(\frac{\operatorname{dist}(y^1,J^1)}{\operatorname{dist}(y^2,J^2)}\Big)^{-\theta_2},$$

where further

$$\begin{aligned}\frac{\ell(J)^{\theta_1}}{\operatorname{dist}(y^2,J^2)^{\theta_1}}\Big(\frac{\operatorname{dist}(y^1,J^1)}{\operatorname{dist}(y^2,J^2)}\Big)^{-\theta_2} &= \frac{\ell(J)^{\theta_1}\operatorname{dist}(y^2,J^2)^{\theta_2-\theta_1}}{\operatorname{dist}(y^1,J^1)^{\theta_2}}\\ \le \Big(\frac{\ell(J)}{\operatorname{dist}(y^1,J^1)}\Big)^{\min(\theta_1,\theta_2)} &\le \prod_{i=1}^{2}\Big(\frac{\ell(J)}{\operatorname{dist}(y^i,J^i)}\Big)^{\theta}\end{aligned}$$

with $\theta := \frac{1}{2}\min(\theta_1, \theta_2)$. □

A combination of the previous two lemmas shows that CZX-kernels satisfy the Hörmander integral condition:

2.3 Lemma *Let $K \in CZX(\mathbb{R}^2)$, and $x \in J$ for some cube $J = J^1 \times J^2 \subset \mathbb{R}^2$ with centre c_J. Then*

$$\int_{(3J)^c} |K(x,y)-K(c_J,y)|\,\mathrm{d}y \lesssim 1.$$

Proof Notice that

$$(3J)^c = ((3J^1)^c \times 3J^2) \cup (3J^1 \times (3J^2)^c) \cup ((3J^1)^c \times (3J^2)^c),$$

where the first two components on the right-hand side are symmetric. For these, we simply estimate

$$\begin{aligned}\int_{(3J^1)^c\times 3J^2}|K(x,y)|\,\mathrm{d}y &= \int_{(3J^1)^c}\Big(\int_{3J^2}|K(x,y)|\,\mathrm{d}y^2\Big)\,\mathrm{d}y^1\\ &\lesssim \int_{(3J^1)^c}\frac{\ell(J)^{\theta_2}}{|x^1-y^1|^{1+\theta_2}}\,\mathrm{d}y^1 \lesssim 1,\end{aligned}$$

where the first $\lesssim$ was an application of Lemma 2.1. The estimate for $K(c_J, y)$ is of course a special case of this with $x = c_J$.

For the remaining component of the integration domain, there holds that

$$\begin{aligned}\int_{(3J^1)^c\times(3J^2)^c} |K(x,y) - K(c_J,y)|\,\mathrm{d}y &\lesssim \int_{(3J^1)^c\times(3J^2)^c} \prod_{i=1}^{2} \frac{\ell(J)^\theta}{\mathrm{dist}(y^i, J^i)^{1+\theta}}\,\mathrm{d}y \\ &= \prod_{i=1}^{2} \int_{(3J^i)^c} \frac{\ell(J)^\theta}{\mathrm{dist}(y^i, J^i)^{1+\theta}}\,\mathrm{d}y^i \lesssim 1,\end{aligned}$$

where the first $\lesssim$ was an application of Lemma 2.2. □

At this point, we can already provide a proof of part (1.3) of Theorem 1.3, which we restate as

2.4 Proposition *Let $T \in \mathcal{L}(L^2(\mathbb{R}^2))$ be an operator associated with a CZX kernel K. Then T extends boundedly from $L^\infty(\mathbb{R}^2)$ into* $\mathrm{BMO}(\mathbb{R}^2)$*, from $L^1(\mathbb{R}^2)$ into $L^{1,\infty}(\mathbb{R}^2)$, and from $L^p(\mathbb{R}^2)$ into itself for all $p \in (1,\infty)$.*

Proof By Lemma 2.3, the kernel K satisfies the Hörmander integral condition; the symmetry of the assumption on K ensures that it also satisfies the version with the roles of the first and second variable interchanged. It is well known that any $L^2(\mathbb{R}^2)$-bounded operator with a Hörmander kernel satisfies the mapping properties stated in the proposition. (See e.g. [22, §I.5] for the boundedness from $L^1(\mathbb{R}^2)$ into $L^{1,\infty}(\mathbb{R}^2)$, and from $L^p(\mathbb{R}^2)$ into itself for $p \in (1,2)$, and [22, §IV.4.1] for the boundedness from $L^\infty(\mathbb{R}^2)$ into $\mathrm{BMO}(\mathbb{R}^2)$. The latter is formulated for convolution kernels $K(x,y) = K(x-y)$, but an inspection of the proof shows that it extends to the general case with trivial modifications. The case of $p \in (2,\infty)$ can be inferred either by duality (observing that the adjoint T^* satisfies the same assumption) or by interpolation between the $L^2(\mathbb{R}^2)$ and the $L^\infty(\mathbb{R}^2)$-to-$\mathrm{BMO}(\mathbb{R}^2)$ estimates.) □

3 Haar Coefficients of CZX Forms

We recall the weak boundedness property and the $T1$ assumptions, which are just the same as in the classical theory for usual Calderón–Zygmund forms.

3.1 Definition Let $B(f,g)$ be a bilinear form defined on finite linear combinations of indicators of cubes of $\mathbb{R}^2$, and such that

$$B(f,g) = \iint K(x,y) f(y) g(x)\,\mathrm{d}x\,\mathrm{d}y$$

when $\{f \neq 0\} \cap \{g \neq 0\} = \varnothing$, where $K \in CZX(\mathbb{R}^2)$. We say that B is a $CZX(\mathbb{R}^2)$-form.

3.2 Definition A $CZX(\mathbb{R}^2)$-form satisfies the weak boundedness property if $|B(1_I, 1_I)| \lesssim |I|$ for all cubes $I \subset \mathbb{R}^2$. It satisfies the $T1$ conditions if

$$B(1, g) = \int b_1 g, \qquad B(f, 1) = \int b_2 f$$

for some $b_1, b_2 \in \mathrm{BMO}(\mathbb{R}^2)$ and all f, g with $\int f = 0 = \int g$. Here

$$\|b\|_{\mathrm{BMO}} = \|b\|_{\mathrm{BMO}(\mathbb{R}^2)} := \sup_I \frac{1}{|I|} \int_I |b - \langle b \rangle_I|,$$

where the supremum is over all cubes $I \subset \mathbb{R}^2$ and $\langle b \rangle_I = \frac{1}{|I|} \int_I b$.

For an interval $I \subset \mathbb{R}$, we denote by I_l and I_r the left and right halves of the interval I, respectively. We define $h_I^0 = |I|^{-1/2} 1_I$ and $h_I^1 = |I|^{-1/2}(1_{I_l} - 1_{I_r})$. Let now $I = I^1 \times I^2$ be a cube, and define the Haar function h_I^η, $\eta = (\eta^1, \eta^2) \in \{0, 1\}^2$, via

$$h_I^\eta = h_{I^1}^{\eta^1} \otimes h_{I^2}^{\eta^2}.$$

3.3 Lemma *Let B be a $CZX(\mathbb{R}^2)$-form satisfying the weak boundedness property. There holds that*

$$\begin{aligned} |B(h_I^\beta, h_J^\gamma)| &\lesssim \prod_{i=1}^{2} \Big(\frac{\ell(I)}{\ell(I) + \mathrm{dist}(I^i, J^i)} \Big) \times \frac{\ell(I)^{\theta_1} (\ell(I) + \min_{i=1,2} \mathrm{dist}(I^i, J^i))^{\theta_2 - \theta_1}}{(\ell(I) + \max_{i=1,2} \mathrm{dist}(I^i, J^i))^{\theta_2}} \\ &\lesssim \prod_{i=1}^{2} \Big(\frac{\ell(I)}{\ell(I) + \mathrm{dist}(I^i, J^i)} \Big)^{1+\theta}, \qquad \theta := \frac{1}{2} \min(\theta_1, \theta_2), \end{aligned}$$

whenever I, J are dyadic cubes with equal side lengths $\ell(I) = \ell(J)$ and at least $\beta \neq 0$ or $\gamma \neq 0$.

Proof We consider several cases. *Adjacent cubes:*

By this, we mean that $\mathrm{dist}(I, J) = 0$, but $I \neq J$. Here, we simply put absolute values inside. We are thus led to estimate

$$\int_I \int_J |K(x, y) h_I^\beta(x) h_J^\gamma(y)| \,\mathrm{d}y \,\mathrm{d}x \leq \frac{1}{|I|} \int_I \int_J |K(x, y)| \,\mathrm{d}y \,\mathrm{d}x. \tag{3.4}$$

By symmetry, we may assume for instance that $I^2 \neq J^2$. Lemma 2.1 gives that

$$\int_{J^1} |K(x, y)| \,\mathrm{d}y^1 \lesssim \frac{1}{|x^2 - y^2|}.$$

The assumption $I^2 \neq J^2$ implies that

$$\int_{I^2} \int_{J^2} \frac{\mathrm{d}x^2 \,\mathrm{d}y^2}{|x^2 - y^2|} \leq \int_{3J^2 \setminus J^2} \int_{J^2} \frac{\mathrm{d}x^2 \,\mathrm{d}y^2}{|x^2 - y^2|} \lesssim \ell(I).$$

The dependence on x^1 has already disappeared, and integration with respect to $x^1 \in I^1$ results in another $\ell(I)$. Then we are only left with observing that $\ell(I)^2/|I| = 1$.

Equal cubes: Now

$$B(h_I^\beta, h_I^\gamma) = \sum_{I', J' \in \mathrm{ch}(I)} \langle h_I^\beta \rangle_{I'} \langle h_I^\gamma \rangle_{J'} B(1_{I'}, 1_{J'}),$$

where $|\langle h_I^\beta \rangle_{I'} \langle h_I^\gamma \rangle_{J'}| = |I|^{-1}$. For $J' = I'$, the WBP implies that $|B(1_{I'}, 1_{I'})| \lesssim |I'| \le |I|$. For $J' \neq I'$, we can estimate the term as in the case of adjacent $I \neq J$, recalling that only the size and no cancellation of the Haar functions was used there.

Cubes separated in one direction:

By this, we mean that, say, $\mathrm{dist}(I^1, J^1) = 0 < \mathrm{dist}(I^2, J^2)$, or the same with 1 and 2 interchanged. We still apply only the non-cancellative estimate (3.4) (in contrast to what one would do with standard Calderón–Zygmund operators). From Lemma 2.1, we deduce that

$$\int_{J^1} |K(x, y)| \,\mathrm{d}y^1 \lesssim \frac{\ell(I)^{\theta_2}}{|x^2 - y^2|^{1+\theta_2}} \lesssim \frac{\ell(I)^{\theta_2}}{(\ell(I) + \mathrm{dist}(I^2, J^2))^{1+\theta_2}}.$$

There is no more dependence on the remaining variables x^1, x^2, y^2, so integrating over these gives the factor $\ell(I)^3$. After dividing by $|I| = \ell(I)^2$ in (3.4), we arrive at the bound

$$\Big(\frac{\ell(I)}{\ell(I) + \mathrm{dist}(I^2, J^2)}\Big)^{1+\theta_2}.$$

Cubes separated in both directions:

By this, we mean that $\mathrm{dist}(I^i, J^i) > 0$ for both $i = 1, 2$. It is only here that we make use of the assumed cancellation of at least one of the Haar functions, say h_I^β. Thus,

$$B(h_I^\beta, h_J^\gamma) = \int_I \int_J [K(x, y) - K(c_I, y)] h_I^\beta(x) h_J^\gamma(y) \,\mathrm{d}y \,\mathrm{d}x,$$

where $c_I = (c_{I^1}, c_{I^2})$ is the centre of I. Now $x \in I$ and $y^i \in J^i \subset (3I^i)^c$ for $i = 1, 2$, so Lemma 2.2 applies to give

$$|B(h_I^\beta, h_J^\gamma)| \lesssim \int_I \int_J \prod_{i=1}^{2} \frac{1}{(\ell(I) + \mathrm{dist}(I^i, J^i))} \times \frac{\ell(I)^{\theta_1} (\ell(I) + \min_{i=1,2} \mathrm{dist}(I_i, J_i))^{\theta_2 - \theta_1}}{(\ell(I) + \max_{i=1,2} \mathrm{dist}(I^i, J^i))^{\theta_2}} \times \frac{1}{|I|} \,\mathrm{d}y \,\mathrm{d}x,$$

which readily simplifies to the claimed bound after $|I|^2/|I| = \ell(I)^2$. □

4 Dyadic Representation and *T*1 Theorem

Let $\mathcal{D}_0$ be the standard dyadic grid in $\mathbb{R}$. For $\omega \in \{0,1\}^{\mathbb{Z}}$, $\omega = (\omega_i)_{i\in\mathbb{Z}}$, we define the shifted lattice

$$\mathcal{D}(\omega) := \Big\{ L + \omega := L + \sum_{i\,:\,2^{-i}<\ell(L)} 2^{-i}\omega_i : L \in \mathcal{D}_0 \Big\}.$$

Let $\mathbb{P}_\omega$ be the product probability measure on $\{0,1\}^{\mathbb{Z}}$. We recall the notion of k-good cubes from [7]. We say that $G \in \mathcal{D}(\omega, k)$, $k \ge 2$, if $G \in \mathcal{D}(\omega)$ and

$$d(G, \partial G^{(k)}) \ge \frac{\ell(G^{(k)})}{4} = 2^{k-2}\ell(G). \tag{4.1}$$

Notice that

$$\mathbb{P}_\omega(\{\omega \colon L + \omega \in \mathcal{D}(\omega, k)\}) = \frac{1}{2} \tag{4.2}$$

for all $L \in \mathcal{D}_0$ and $k \ge 2$.

For $\sigma = (\sigma^1, \sigma^2) \in \{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}$ and dyadic $\lambda > 0$ define

$$\begin{aligned}\mathcal{D}(\sigma) &:= \mathcal{D}(\sigma^1) \times \mathcal{D}(\sigma^2),\\ \mathcal{D}_\lambda(\sigma) &:= \{I = I^1 \times I^2 \in \mathcal{D}(\sigma) \colon \ell(I^1) = \lambda \ell(I^2)\},\\ \mathcal{D}_\square(\sigma) &:= \mathcal{D}_1(\sigma).\end{aligned}$$

Let $\mathbb{P}_\sigma := \mathbb{P}_{\sigma^1} \times \mathbb{P}_{\sigma^2}$. For $k = (k^1, k^2)$, $k^1, k^2 \ge 2$, we define $\mathcal{D}(\sigma, k) = \mathcal{D}(\sigma^1, k^1) \times \mathcal{D}(\sigma^2, k^2)$.

We will need an estimate for the maximal operator

$$M_{\mathcal{D}_\lambda(\sigma)} f(x) := \sup_{I \in \mathcal{D}_\lambda(\sigma)} 1_I(x) \langle |f| \rangle_I.$$

Before bounding it, we recall the following interpolation result due to Stein and Weiss, see [23, Theorem 2.11].

4.3 Proposition *Suppose that* $1 \le p_0, p_1 \le \infty$ *and let* w_0 *and* w_1 *be positive weights. Suppose that* T *is a sublinear operator that satisfies the estimates*

$$\|Tf\|_{L^{p_i}(w_i)} \le M_i \|f\|_{L^{p_i}(w_i)}, \quad i = 1, 2.$$

Let $t \in (0,1)$ *and define* $1/p = (1-t)/p_0 + t/p_1$ *and* $w = w_0^{p(1-t)/p_0} w_1^{pt/p_1}$. *Then* T *satisfies the estimate*

$$\|Tf\|_{L^p(w)} \le M_0^{1-t} M_1^t \|f\|_{L^p(w)}.$$

4.4 Proposition *For all* $p \in (1, \infty)$ *and all* $w \in A_p$*, there are constants* $C = C(p, w)$, $\eta = \eta(p, w) > 0$ *such that*

$$\|M_{\mathcal{D}_\lambda(\sigma)} f\|_{L^p(w)} \le C \cdot D(\lambda)^{1-\eta} \|f\|_{L^p(w)},$$

where $D(\lambda) := \max(\lambda, \lambda^{-1})$.

Proof The parameter σ plays no role in this argument, so we drop it from the notation. Since $\mathcal{D}_\lambda$ has the same nestedness structure as the usual $\mathcal{D}_\square$, the unweighted bound

$$\|M_{\mathcal{D}_\lambda} f\|_{L^s} \le s' \|f\|_{L^s}, \quad \forall s \in (1, \infty),$$

holds. On the other hand, for any $I \in \mathcal{D}_\lambda$, there is some $J \in \mathcal{D}_\square$ such that $I \subset J$ and $|J| \le D(\lambda)|I|$. Therefore, we conclude that

$$M_{\mathcal{D}_\lambda} f(x) = \sup_{I \in \mathcal{D}_\lambda} \langle |f| \rangle_I 1_I(x) \le D(\lambda) \sup_{J \in \mathcal{D}_\square} \langle |f| \rangle_J 1_J(x) = D(\lambda) M_{\mathcal{D}_\square} f(x),$$

and so

$$\|M_{\mathcal{D}_\lambda} f\|_{L^s(w)} \le C(s, w) D(\lambda) \|f\|_{L^s(w)}, \quad \forall s \in (1, \infty), \quad \forall w \in A_s.$$

Let us now consider $s \in (1, \infty)$ and $w \in A_s$ fixed. It is well known that we can find a $\delta = \delta(s, w) > 0$ such that $w^{1+\delta} \in A_s$, and thus

$$\|M_{\mathcal{D}_\lambda} f\|_{L^s(w^{1+\delta})} \le C(s, w^{1+\delta}) D(\lambda) \|f\|_{L^s(w^{1+\delta})}.$$

Now $w = (w^{1+\delta})^{1/(1+\delta)} \cdot 1^{\delta/(1+\delta)}$ and Proposition 4.3 shows that

$$\|M_{\mathcal{D}_\lambda} f\|_{L^s(w)} \le \big(C(s, w^{1+\delta}) D(\lambda)\big)^{1/(1+\delta)} (s')^{\delta/(1+\delta)} \|f\|_{L^s(w)}.$$

Set $\eta := \delta/(1+\delta)$. We have found $\eta = \eta(\delta) = \eta(s, w) > 0$ such that

$$\|M_{\mathcal{D}_\lambda} f\|_{L^s(w)} \le C(s, w) D(\lambda)^{1-\eta(s,w)} \|f\|_{L^s(w)}.$$

□

In addition to the usual Haar functions, we will need the functions $H_{I,J}$, where I and J are cubes with equal side length. The functions $H_{I,J}$ satisfy

(1) $H_{I,J}$ is supported on $I \cup J$ and constant on the children of I and J,
(2) $|H_{I,J}| \le |I|^{-1/2}$ and
(3) $\int H_{I,J} = 0$.

We denote (by slightly abusing notation) a general cancellative Haar function h_I^η, $\eta \ne (0, 0)$, simply by h_I.

4.5 Definition For $k = (k^1, k^2)$, $k^i \geq 0$, we define that the operator $Q_{k,\sigma}$ has either the form

$$\langle Q_{k,\sigma} f, g\rangle = \sum_{K \in \mathcal{D}_{2^{k^1-k^2}}(\sigma)} \sum_{\substack{I,J \in \mathcal{D}_\square(\sigma) \\ I^{(k)}=J^{(k)}=K}} a_{IJK} \langle f, H_{I,J}\rangle \langle g, h_J\rangle$$

or the symmetric form, and here $I^{(k)} = I^{(k^1)} \times I^{(k^2)}$ and the constants a_{IJK} satisfy

$$|a_{IJK}| \leq \frac{|I|}{|K|}.$$

4.6 Lemma *For $p \in (1, \infty)$ there holds that*

$$\|Q_{k,\sigma} f\|_{L^p} \lesssim (1 + \max(k^1, k^2))^{1/2} \|f\|_{L^p}.$$

Moreover, for $w \in A_p$, there is $\eta > 0$ such that

$$\|Q_{k,\sigma} f\|_{L^p(w)} \lesssim (1 + \max(k^1, k^2))^{1/2} 2^{|k^1-k^2|(1-\eta)} \|f\|_{L^p(w)}.$$

Proof We consider σ fixed here and drop it from the notation. Suppose, e.g. $k^1 \geq k^2$. We write

$$\langle f, H_{I,J}\rangle = \big\langle E_{\frac{\ell(I)}{2}} f - E_{\ell(K^1)} f, H_{I,J}\big\rangle, \qquad E_\lambda f := \sum_{\substack{L \in \mathcal{D}_\square \\ \ell(L)=\lambda}} E_L f, \; E_L f = \langle f\rangle_L 1_L.$$

Therefore, $\langle f, H_{I,J}\rangle = \langle \gamma_{K,k^1} f, H_{I,J}\rangle$, where

$$\gamma_{K,k^1} f := 1_K \sum_{\substack{L \in \mathcal{D}_\square \\ 2^{-k^1}\ell(K^1) \leq \ell(L) \leq \ell(K^1)}} \Delta_L f, \qquad \Delta_L f = \sum_{\substack{L' \in \mathcal{D}_\square \\ L' \subset L,\, \ell(L') = \frac{\ell(L)}{2}}} E_{L'} f - E_L f.$$

Notice now that for $w \in A_2$, there holds that

$$\begin{aligned}
\Big\| \Big(\sum_{K \in \mathcal{D}_{2^{k^1-k^2}}} |\gamma_{K,k^1} f|^2 \Big)^{\frac{1}{2}} \Big\|_{L^2(w)}^2 &= \sum_{G \in \mathcal{D}_\square} \Big\| \sum_{\substack{K \in \mathcal{D}_{2^{k^1-k^2}} \\ K^{(0,k^1-k^2)}=G}} \gamma_{K,k^1} f \Big\|_{L^2(w)}^2 \\
&= \sum_{G \in \mathcal{D}_\square} \Big\| \sum_{\substack{L \in \mathcal{D}_\square,\, L \subset G \\ \ell(L) \geq 2^{-k^1}\ell(G)}} \Delta_L f \Big\|_{L^2(w)}^2 \\
&\sim \sum_{G \in \mathcal{D}_\square} \sum_{\substack{L \in \mathcal{D}_\square,\, L \subset G \\ \ell(L) \geq 2^{-k^1}\ell(G)}} \|\Delta_L f\|_{L^2(w)}^2 \lesssim (1+k^1) \|f\|_{L^2(w)}^2,
\end{aligned}$$

where we used the standard weighted square function estimate

$$\sum_{L\in\mathcal{D}_\square} \|\Delta_L f\|_{L^2(w)}^2 \sim \|f\|_{L^2(w)}^2$$

twice in the end.

To bound $Q_k f$ we need to estimate

$$\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \frac{1}{|K|} \sum_{\substack{I,J\in\mathcal{D}_\square \\ I^{(k)}=J^{(k)}=K}} \langle|\gamma_{K,k^1} f|, 1_I + 1_J\rangle\langle|\Delta_{K,k} g|, 1_J\rangle,$$

$$\Delta_{K,k} g := \sum_{\substack{J\in\mathcal{D}_\square \\ J^{(k^1,k^2)}=K}} \Delta_J g.$$

We split this into two pieces according to $1_I + 1_J$. The first piece is bounded by

$$\begin{aligned}
\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \int \langle|\gamma_{K,k^1} f|\rangle_K |\Delta_{K,k} g| &\le \sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \int M_{\mathcal{D}_{2^{k^1-k^2}}} \gamma_{K,k^1} f \cdot |\Delta_{K,k} g| \\
&\le \Big\| \Big(\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} [M_{\mathcal{D}_{2^{k^1-k^2}}} \gamma_{K,k^1} f]^2 \Big)^{1/2} \Big\|_{L^2(w)} \Big\| \Big(\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} |\Delta_{K,k} g|^2 \Big)^{1/2} \Big\|_{L^2(w^{-1})} \\
&\lesssim 2^{(k^1-k^2)(1-\eta)} \Big\| \Big(\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} |\gamma_{K,k^1} f|^2 \Big)^{1/2} \Big\|_{L^2(w)} \|g\|_{L^2(w^{-1})} \\
&\lesssim 2^{(k^1-k^2)(1-\eta)} (1+k^1)^{1/2} \|f\|_{L^2(w)} \|g\|_{L^2(w^{-1})}
\end{aligned}$$

while the second piece is bounded by

$$\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \Big\langle \sum_{\substack{J\in\mathcal{D}_\square \\ J^{(k^1,k^2)}=K}} \langle|\gamma_{K,k^1} f|\rangle_J 1_J, |\Delta_{K,k} g| \Big\rangle \le \sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \int M_{\mathcal{D}_\square} \gamma_{K,k^1} f \cdot |\Delta_{K,k} g|,$$

which is estimated similarly except that the bound for $M_{\mathcal{D}_{2^{k^1-k^2}}}$ is replaced by the standard result for $M_{\mathcal{D}_\square}$. This proves the claimed bounds in $L^2(w)$, and the results for $L^p(w)$ follow by Rubio de Francia's extrapolation theorem (the correct $1-\eta$ dependence is maintained by the extrapolation, see Remark 4.7).

For the unweighted estimate in L^p (with better complexity dependence), simply run the above argument using the Fefferman–Stein $L^p(\ell^2)$ estimate for the strong maximal function instead of the $L^2(w)$ estimate of $M_{\mathcal{D}_{2^{k^1-k^2}}}$, and use the analogous $L^p(\ell^2)$ estimate for the square function involving γ_{K,k^1} that follows via Rubio de Francia extrapolation from the proved $L^2(w)$ estimate of the same square function. □

4.7 Remark It is clear that when $p = 2$, η depends only on the A_2 constant of w. In fact, in the proof of Proposition 4.4 we get $\eta \sim 1/[w]_{A_2}$. Thus,

$$\|Q_{k,\sigma} f\|_{L^2(w)} \le (1 + \max(k^1, k^2))^{1/2} 2^{|k^1-k^2|} N([w]_{A_2}) \|f\|_{L^2(w)}$$

with

$$N([w]_{A_2}) = K([w]_{A_2}) 2^{-c|k^1-k^2|/[w]_{A_2}},$$

where K is an increasing function. Hence N is also an increasing function. Then standard extrapolation (see e.g. [2, Theorem 3.1]) gives that the $L^p(w)$ bound of $Q_{k,\sigma}$ is

$$(1 + \max(k^1, k^2))^{1/2} 2^{|k^1-k^2|} N(c_p [w]_{A_p}^{\alpha(p)}).$$

Then we get the desired estimate with $\eta = c(c_p [w]_{A_p}^{\alpha(p)})^{-1}$.

4.8 Definition We say that π_b is a (one-parameter) paraproduct if it has the form

$$\langle \pi_b f, g \rangle = \sum_{I \in \mathcal{D}_\square} \langle b, h_I \rangle \langle f \rangle_I \langle g, h_I \rangle$$

or the symmetric form.

It is well known (and follows readily from H^1–BMO duality) that paraproducts are L^p bounded for $p \in (1, \infty)$ (and $L^p(w)$ bounded for $w \in A_p$) precisely when $b \in \mathrm{BMO}$.

4.9 Theorem *Let B be a $CZX(\mathbb{R}^2)$-form satisfying the weak boundedness property and the $T1$ conditions. Then*

$$B(f, g) = \mathbb{E}_\sigma \Big[C_T \sum_{k^1, k^2 \ge 0} 2^{-\theta_2 (k_{\max} - k_{\min})} 2^{-\theta_1 k_{\min}} \langle Q_{k,\sigma} f, g \rangle + \langle \pi_{b_1,\sigma} f, g \rangle + \langle \pi_{b_2,\sigma} f, g \rangle \Big],$$

where $k_{\max} = \max_{i=1,2} k^i$, $k_{\min} = \min_{i=1,2} k^i$. In particular, for $p \in (1, \infty)$ there holds that

$$|B(f, g)| \lesssim \|f\|_{L^p} \|g\|_{L^{p'}}.$$

If $\theta_2 = 1$, there also holds for all $w \in A_p$ that

$$|B(f, g)| \lesssim \|f\|_{L^p(w)} \|g\|_{L^{p'}(w^{1-p'})}.$$

Proof Write (by expanding $f = \sum_I \Delta_I f$, $g = \sum_J \Delta_J g$ and collapsing the off-diagonal)

$$B(f, g) = \mathbb{E}_\sigma \sum_{\ell(I) = \ell(J)} [B(E_I f, \Delta_J g) + B(\Delta_I f, E_J g) + B(\Delta_I f, \Delta_J g)],$$

where $I, J \in \mathcal{D}_\square(\sigma)$. We begin by writing

$$\begin{aligned}\Sigma_1 &:= \mathbb{E}_\sigma \sum_{\ell(I)=\ell(J)} B(E_I f, \Delta_J g)\\ &= \mathbb{E}_\sigma \sum_{\ell(I)=\ell(J)} B(h_I^0, h_J)\langle f, H_{I,J}\rangle\langle g, h_J\rangle\\ &\quad + \mathbb{E}_\sigma \sum_{J\in\mathcal{D}_\square(\sigma)} B(1, h_J)\langle f\rangle_J\langle g, h_J\rangle =: \Sigma_{1,1} + \Sigma_{1,2},\end{aligned}$$

where $H_{I,J} := h_I^0 - h_J^0$. As the term $\Sigma_{1,2}$ is readily a paraproduct, we only continue with $\Sigma_{1,1}$. This was a standard one-parameter start. Write

$$\begin{aligned}\Sigma_{1,1} &= \mathbb{E}_\sigma \sum_{m=(m^1,m^2)\in\mathbb{Z}^2\setminus\{0\}} \sum_{I\in\mathcal{D}_\square(\sigma)} \varphi_{I,I\dot{+}m},\\ \varphi_{I,I\dot{+}m} &= B(h_I^0, h_{I\dot{+}m})\langle f, H_{I,I\dot{+}m}\rangle\langle g, h_{I\dot{+}m}\rangle,\end{aligned}$$

where $I\dot{+}m := I + m\ell(I) \in \mathcal{D}_\square(\sigma)$. Next, write

$$\sum_{m=(m^1,m^2)\in\mathbb{Z}^2\setminus\{0\}} = \sum_{m^1\in\mathbb{Z}\setminus\{0\}}\sum_{m^2\in\mathbb{Z}\setminus\{0\}} + \sum_{\substack{m^2\in\mathbb{Z}\setminus\{0\}\\ m=(0,m^2)}} + \sum_{\substack{m^1\in\mathbb{Z}\setminus\{0\}\\ m=(m^1,0)}}.$$

Focusing, for now, on the part $m^1 \neq 0$ and $m^2 \neq 0$, write

$$\sum_{m^1\in\mathbb{Z}\setminus\{0\}}\sum_{m^2\in\mathbb{Z}\setminus\{0\}} = \sum_{k^1=2}^{\infty}\sum_{k^2=2}^{\infty}\sum_{|m^1|\in(2^{k^1-3},2^{k^1-2}]}\sum_{|m^2|\in(2^{k^2-3},2^{k^2-2}]}.$$

Independence and (4.2) imply that

$$\begin{aligned}&\mathbb{E}_\sigma \sum_{k^1=2}^{\infty}\sum_{k^2=2}^{\infty}\sum_{|m^1|\in(2^{k^1-3},2^{k^1-2}]}\sum_{|m^2|\in(2^{k^2-3},2^{k^2-2}]}\sum_{I\in\mathcal{D}_\square(\sigma)} \varphi_{I,I\dot{+}m}\\ &= 4\mathbb{E}_\sigma \sum_{k^1=2}^{\infty}\sum_{k^2=2}^{\infty}\sum_{|m^1|\in(2^{k^1-3},2^{k^1-2}]}\sum_{|m^2|\in(2^{k^2-3},2^{k^2-2}]}\sum_{I\in\mathcal{D}_\square(\sigma,k)} \varphi_{I,I\dot{+}m},\end{aligned}$$

where $k = (k^1, k^2)$, and the gist is that for $|m^1| \in (2^{k^1-3}, 2^{k^1-2}]$, $|m^2| \in (2^{k^2-3}, 2^{k^2-2}]$ and $I \in \mathcal{D}_\square(\sigma, k)$ there holds that

$$(I^1 \dot{+} m^1)^{(k^1)} = (I^1)^{(k^1)} =: K^1 \qquad \text{and} \qquad (I^2 \dot{+} m^2)^{(k^2)} = (I^2)^{(k^2)} =: K^2.$$

Notice that $K = I^{(k)} = (I \dotplus m)^{(k)} = K^1 \times K^2 \in \mathcal{D}_{2^{k^1-k^2}}(\sigma)$, since

$$\ell(K^1) = 2^{k^1}\ell(I^1) = 2^{k^1}\ell(I^2) = 2^{k^1-k^2}2^{k^2}\ell(I^2) = 2^{k^1-k^2}\ell(K^2).$$

Finally, notice that Lemma 3.3 implies that

$$|B(h_I^0, h_{I\dotplus m})| \lesssim 2^{-k^1}2^{-k^2}\frac{2^{k_{\min}(\theta_2-\theta_1)}}{2^{k_{\max}\theta_2}} = 2^{-\theta_2(k_{\max}-k_{\min})}2^{-\theta_1 k_{\min}}\frac{|I|}{|K|}.$$

The sums, where $m^1 = 0$ or $m^2 = 0$ are completely similar (just do the above in one of the parameters). We are done with Σ_1.

Of course, $\Sigma_2 := \mathbb{E}_\sigma \sum_{\ell(I)=\ell(J)} B(\Delta_I f, E_J g)$ is completely symmetric. The term $\Sigma_3 := \mathbb{E}_\sigma \sum_{\ell(I)=\ell(J)} B(\Delta_I f, \Delta_J g)$ does not produce a paraproduct and produces shifts with the simpler form $H_{I,J} = h_I$.

The unweighted boundedness follows immediately from the L^p bounds of the paraproducts and the bound $\|Q_{k,\sigma} f\|_{L^p} \lesssim (1+k_{\max})^{1/2}\|f\|_{L^p}$, since the exponentially decaying factor $2^{-\theta_2(k_{\max}-k_{\min})}2^{-\theta_1 k_{\min}}$ clearly make the series summable for any $\theta_1, \theta_2 > 0$.

Let us finally consider the weighted case with $\theta_2 = 1$. Then for some $\eta = \eta(p, w) > 0$, there holds that

$$\begin{aligned} &2^{-\theta_2(k_{\max}-k_{\min})}2^{-\theta_1 k_{\min}}|\langle Q_{k,\sigma} f, g\rangle| \\ &\quad\lesssim 2^{-(k_{\max}-k_{\min})}2^{-\theta_1 k_{\min}}2^{(k_{\max}-k_{\min})(1-\eta)}(1+k_{\max})^{1/2}\|f\|_{L^p(w)}\|g\|_{L^{p'}(w^{1-p'})} \\ &\quad= 2^{-\eta(k_{\max}-k_{\min})}2^{-\theta_1 k_{\min}}(1+k_{\max})^{1/2}\|f\|_{L^p(w)}\|g\|_{L^{p'}(w^{1-p'})}, \end{aligned}$$

and again we have exponential decay that makes the series over k^1, k^2 summable. □

4.10 Remark If $\theta_2 = 1$, we may redefine $D_1(x, y)$ to be the slightly larger quantity

$$D_1(x, y) := \left(\frac{|x^1-y^1|}{|x^2-y^2|} + \frac{|x^2-y^2|}{|x^1-y^1|}\right)^{-1}\log\left(\frac{|x^1-y^1|}{|x^2-y^2|} + \frac{|x^2-y^2|}{|x^1-y^1|}\right),$$

and still prove the weighted estimates essentially like above. This is pertinent in the sense that if we take a Fefferman–Pipher multiplier [4]—a singular integral of Zygmund type—and use it to induce a CZX operator, a logarithmic term appears. In this threshold a weighted estimate still holds. See also [12].

5 Commutator Estimates

We show that our exotic Calderón–Zygmund operators also satisfy the usual one-parameter commutator estimates. Since weighted estimates with one-parameter weights do not in general hold (see Sect. 6), this does not follow from the well-known Cauchy integral trick.

5.1 Theorem *Let $T \in \mathcal{L}(L^2(\mathbb{R}^2))$ be an operator associated with a CZX kernel K. Then*

$$\|[b,T]f\|_{L^p} \lesssim \|b\|_{\mathrm{BMO}}\|f\|_{L^p}$$

whenever $p \in (1,\infty)$. Here $[b,T]f := bTf - T(bf)$.

Proof By Theorem 4.9 and the well-known commutator estimates for the paraproducts π, we only need to prove that

$$\big|\langle Q_{k,\sigma}(bf), g\rangle - \langle Q_{k,\sigma} f, bg\rangle\big| \lesssim \varphi(k)\|b\|_{\mathrm{BMO}}\|f\|_{L^p}\|g\|_{L^{p'}},$$

where φ is some polynomial. We consider σ fixed and drop it from the notation. Recall the usual paraproduct decomposition of bf:

$$bf = a_1(b,f) + a_2(b,f) + a_3(b,f),$$

where

$$a_1(b,f) = \sum_{I\in\mathcal{D}_\square} \Delta_I b \Delta_I f, \quad a_2(b,f) = \sum_{I\in\mathcal{D}_\square} \Delta_I b \langle f\rangle_I, \quad a_3(b,f) = \sum_{I\in\mathcal{D}_\square} \langle b\rangle_I \Delta_I f.$$

Invoking the above decomposition, the well-known boundedness of paraproducts

$$\|a_i(b,f)\|_{L^p} \lesssim \|b\|_{\mathrm{BMO}}\|f\|_{L^p}, \qquad i = 1,2,$$

and Lemma 4.6, it suffices to control

$$\begin{aligned} &\big|\langle Q_k(a_3(b,f)), g\rangle - \langle Q_k f, a_3(b,g)\rangle\big| \\ &\quad = \Big|\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \sum_{\substack{I,J\in\mathcal{D}_\square \\ I^{(k)}=J^{(k)}=K}} a_{IJK}\Big[\langle a_3(b,f), H_{I,J}\rangle\langle g, h_J\rangle - \langle b\rangle_J\langle f, H_{I,J}\rangle\langle g, h_J\rangle\Big]\Big|. \end{aligned} \tag{5.2}$$

We may assume $k^1 \ge k^2$. There holds that

$$\begin{aligned} &\langle a_3(b,f), H_{I,J}\rangle\langle g, h_J\rangle - \langle b\rangle_J\langle f, H_{I,J}\rangle\langle g, h_J\rangle \\ &\quad = \sum_{\substack{Q\in\mathcal{D}_\square,\, Q\subset K^{(0,k^1-k^2)} \\ \ell(I)\le\ell(Q)\le 2^{k^1}\ell(I)}} (\langle b\rangle_Q - \langle b\rangle_J)\langle \Delta_Q f, H_{I,J}\rangle\langle g, h_J\rangle. \end{aligned}$$

Observe that $|\langle b\rangle_Q - \langle b\rangle_J| \lesssim k^1\|b\|_{\mathrm{BMO}}$. On the other hand, since we only need to consider those Q such that $\langle \Delta_Q f, H_{I,J}\rangle \ne 0$, i.e. either $I \subset Q$ or $J \subset Q$, there holds that

$$
\begin{aligned}
&\sum_{\substack{Q\in\mathcal{D}_\square, Q\subset K^{(0,k^1-k^2)}\\ \ell(I)\le\ell(Q)\le 2^{k^1}\ell(I)}} \big|(\langle b\rangle_Q-\langle b\rangle_J)\langle\Delta_Q f, H_{I,J}\rangle\langle g,h_J\rangle\big|\\
&\lesssim k^1\|b\|_{\mathrm{BMO}}\sum_{\ell^1=0}^{k^1}\Big(\big|\langle\Delta_{I^{(\ell^1)}}f, H_{I,J}\rangle\langle g,h_J\rangle\big|+\big|\langle\Delta_{J^{(\ell^1)}}f, H_{I,J}\rangle\langle g,h_J\rangle\big|\Big)\\
&\le 2k^1\|b\|_{\mathrm{BMO}}\sum_{\ell^1=0}^{k^1}\Big(\langle|\Delta_{I^{(\ell^1)}}f|, h_I^0\rangle|\langle g,h_J\rangle|+\langle|\Delta_{J^{(\ell^1)}}f|, h_J^0\rangle|\langle g,h_J\rangle|\Big),
\end{aligned}
$$

where we have used the simple observation

$$
\langle|\Delta_{I^{(\ell^1)}}f|, h_J^0\rangle\le\langle|\Delta_{J^{(\ell^1)}}f|, h_J^0\rangle,\quad \langle|\Delta_{J^{(\ell^1)}}f|, h_I^0\rangle\le\langle|\Delta_{I^{(\ell^1)}}f|, h_I^0\rangle.
$$

Now, returning to (5.2), for a fixed ℓ^1, we get

$$
\begin{aligned}
&\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}}\ \sum_{\substack{I,J\in\mathcal{D}_\square\\ I^{(k)}=J^{(k)}=K}}\frac{|I|}{|K|}\langle|\Delta_{I^{(\ell^1)}}f|, h_I^0\rangle|\langle g,h_J\rangle|\\
&\le\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}}\Big\langle\sum_{\substack{I\in\mathcal{D}_\square\\ I^{(k)}=K}}|\Delta_{I^{(\ell^1)}}f|1_I, M_{\mathcal{D}_{2^{k^1-k^2}}}|\Delta_{K,k}g|\Big\rangle.
\end{aligned}
$$

Using extrapolation we only need to show that

$$
\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}}\Big\|\sum_{\substack{I\in\mathcal{D}_\square\\ I^{(k)}=K}}|\Delta_{I^{(\ell^1)}}f|1_I\Big\|_{L^2(w)}^2\lesssim\|f\|_{L^2(w)}^2.
$$

However, this is clear because

$$
\begin{aligned}
\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}}\Big\|\sum_{\substack{I\in\mathcal{D}_\square\\ I^{(k)}=K}}|\Delta_{I^{(\ell^1)}}f|1_I\Big\|_{L^2(w)}^2&=\sum_{G\in\mathcal{D}_\square}\ \sum_{\substack{K\in\mathcal{D}_{2^{k^1-k^2}}\\ K^{(0,k^1-k^2)}=G}}\ \sum_{\substack{I\in\mathcal{D}_\square\\ I^{(k)}=K}}\|1_I\Delta_{I^{(\ell^1)}}f\|_{L^2(w)}^2\\
&=\sum_{G\in\mathcal{D}_\square}\ \sum_{\substack{I\in\mathcal{D}_\square\\ I^{(k^1,k^1)}=G}}\|1_I\Delta_{I^{(\ell^1)}}f\|_{L^2(w)}^2\\
&=\sum_{I\in\mathcal{D}_\square}\|1_I\Delta_{I^{(\ell^1)}}f\|_{L^2(w)}^2\\
&=\sum_{Q\in\mathcal{D}_\square}\|\Delta_Q f\|_{L^2(w)}^2.
\end{aligned}
$$

To conclude the proof of the proposition, we are left to deal with

$$\sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \sum_{\substack{I,J\in\mathcal{D}_\square \\ I^{(k)}=J^{(k)}=K}} \frac{|I|}{|K|}\langle|\Delta_{J^{(\ell^1)}} f|, h_J^0\rangle|\langle g, h_J\rangle|$$
$$\leq \sum_{K\in\mathcal{D}_{2^{k^1-k^2}}} \Big\langle \sum_{\substack{J\in\mathcal{D}_\square \\ J^{(k)}=K}} |\Delta_{J^{(\ell^1)}} f|1_J, M_{\mathcal{D}_\square}|\Delta_{K,k} g|\Big\rangle.$$

After the estimate above, this is clearly similar as the other term. We are done. □

6 Counterexample to Weighted Estimates and Sparse Bounds

We begin by showing that bounded $CZX(\mathbb{R}^2)$ operators need not be bounded with respect to the one-parameter weights if $\theta_2 < 1$.

6.1 Lemma *For scalars $\theta_2 \in (0, 1]$, $t_1, t_2 > 0$ and a bump function φ we define*

$$K(x) = K_{t_1,t_2,\theta_2}(x) = \Big(\frac{t_1}{t_2} + \frac{t_2}{t_1}\Big)^{-\theta_2} \prod_{i=1}^{2} \frac{1}{t_i}\phi\Big(\frac{x_i}{t_i}\Big).$$

Then, uniformly on t_1, t_2, $K \in CZX(\mathbb{R}^2)$ with $\theta_1 = 1$ and θ_2 appearing in the very definition of K, and

$$\|K * f\|_{L^2} \lesssim \|f\|_{L^2}.$$

Proof Suppose by symmetry that $t_1 \leq t_2$. Then, using the rapid decay of all the derivatives of ϕ, we conclude for all N that

$$\begin{aligned}
&|x_1^{1+\alpha_1} x_2^{1+\alpha_2}||\partial^\alpha K(x)| \lesssim \Big(\frac{t_1}{t_2}\Big)^{\theta_2} \prod_{i=1}^{2} \Big(\frac{|x_i|}{t_i}\Big)^{1+\alpha_i} \Big(1 + \frac{|x_i|}{t_i}\Big)^{-N} \\
&= \Big(\frac{t_1}{t_2}\Big)^{\theta_2} \prod_{i=1}^{2} \frac{|x_i|^{1+\alpha_i} t_i^{N-1-\alpha_i}}{(|x_i| + t_i)^N} \\
&= \frac{|x_1|^{1+\alpha_1-\theta_2} t_1^{N-\alpha_1-1+\theta_2}}{(|x_1| + t_1)^N} \frac{|x_2|^{1+\alpha_2+\theta_2} t_2^{N-\alpha_2-1-\theta_2}}{(|x_2| + t_2)^N} \Big(\frac{|x_1|}{|x_2|}\Big)^{\theta_2} \\
&= \Big(\frac{t_1}{t_2}\Big)^{2\theta_2} \frac{|x_1|^{1+\alpha_1+\theta_2} t_1^{N-\alpha_1-1-\theta_2}}{(|x_1| + t_1)^N} \frac{|x_2|^{1+\alpha_1-\theta_2} t_2^{N-\alpha_1-1+\theta_2}}{(|x_2| + t_2)^N} \Big(\frac{|x_2|}{|x_1|}\Big)^{\theta_2}.
\end{aligned} \tag{6.2}$$

In the last two lines of (6.2), if N is large enough, each factor in front of the last one can be bounded by one. Thus, we get that

$$|x_1^{1+\alpha_1} x_2^{1+\alpha_2}||\partial^\alpha K(x)| \lesssim \min\left\{\left(\frac{|x_1|}{|x_2|}\right)^{\theta_2}, \left(\frac{|x_2|}{|x_1|}\right)^{\theta_2}\right\} \sim \left(\frac{|x_1|}{|x_2|} + \frac{|x_2|}{|x_1|}\right)^{-\theta_2}.$$

From $\alpha \in \{(0,0), (0,1), (1,0)\}$, we get the desired kernel estimates.

For the boundedness, notice that

$$|\widehat{K}(\xi)| = \left(\frac{t_1}{t_2} + \frac{t_2}{t_1}\right)^{-\theta_2} \prod_{i=1}^{2} |\widehat{\phi}(t_i\xi_i)| \le \prod_{i=1}^{2} |\widehat{\phi}(t_i\xi_i)| \lesssim 1.$$

□

We fix t_1, t_2, θ_2 momentarily and denote $K = K_{t_1,t_2,\theta_2}$. For any rectangle R of sidelengths t_1, t_2, it is clear that

$$K * f \gtrsim \operatorname{ecc}(R)^{-\theta_2} 1_R \langle f \rangle_R$$

whenever $f \ge 0$ and

$$\operatorname{ecc}(R) := \max\left\{\frac{t_1}{t_2}, \frac{t_2}{t_1}\right\}$$

is the eccentricity of R. Suppose now that for $p \in (1,\infty)$ there holds that $\|K * f\|_{L^p(w)} \le C([w]_{A_p}) N \|f\|_{L^p(w)}$ for all $w \in A_p$ and $f \in L^p(w)$. Then

$$\operatorname{ecc}(R)^{-\theta_2} w(R)^{1/p} \langle f \rangle_R = \|\operatorname{ecc}(R)^{-\theta_2} 1_R \langle f \rangle_R\|_{L^p(w)} \lesssim C([w]_{A_p}) N \|f\|_{L^p(w)}.$$

If $f = 1_R\sigma$, where $\sigma = w^{-1/(p-1)}$, then $f^p w = 1_R \sigma$, and hence

$$\operatorname{ecc}(R)^{-\theta_2} w(R)^{1/p} \langle \sigma \rangle_R \lesssim C([w]_{A_p}) N \sigma(R)^{1/p} = C([w]_{A_p}) N \langle \sigma \rangle_R^{1/p} |R|^{1/p},$$

or

$$\langle w \rangle_R^{1/p} \langle \sigma \rangle_R^{1/p'} \lesssim C([w]_{A_p}) N \operatorname{ecc}(R)^{\theta_2}.$$

If all L^2 bounded CZX operators would satisfy the $L^p(w)$ boundedness for all $w \in A_p$ with a bound $C([w]_{A_p})N$, where N depends only on the kernel constants and boundedness constants of the operator, then for all rectangles $R \subset \mathbb{R}^2$, the estimate

$$\langle w \rangle_R^{p'/p} \langle \sigma \rangle_R \lesssim C([w]_{A_p}) \operatorname{ecc}(R)^{\theta_2 p'} \tag{6.3}$$

holds. This is because for the kernels $K = K_{t_1,t_2,\theta_2}$ the constant N is uniformly bounded on t_1, t_2.

Now consider $w(x) = |x|^\alpha$, which belongs to $A_p(\mathbb{R}^2)$ if $-2 < \alpha < 2(p-1)$. Fix some $\alpha \in (p-1, 2(p-1))$ for now. We let our implicit constants depend on α as it is not important for our argument. We consider a rectangle R of the form $(0,\epsilon) \times (\epsilon, 1)$ with eccentricity $\sim 1/\epsilon$. On R there holds that $|x| \sim x_2$, and then

$$\langle w \rangle_R \sim \epsilon^{-1} \int_0^\epsilon \Big(\int_\epsilon^1 x_2^\alpha \,\mathrm{d}x_2 \Big) \,\mathrm{d}x_1 \sim \epsilon^{-1} \int_0^\epsilon 1 \,\mathrm{d}x_1 = 1$$

and

$$\langle \sigma \rangle_R \sim \epsilon^{-1} \int_0^\epsilon \Big(\int_\epsilon^1 x_2^{-\frac{\alpha}{p-1}} \,\mathrm{d}x_2 \Big) \,\mathrm{d}x_1 \sim \epsilon^{-1} \int_0^\epsilon \epsilon^{1-\frac{\alpha}{p-1}} \,\mathrm{d}x_1 = \epsilon^{1-\frac{\alpha}{p-1}} \sim \operatorname{ecc}(R)^{\frac{\alpha}{p-1}-1}.$$

If (6.3) holds, then

$$\operatorname{ecc}(R)^{\frac{\alpha}{p-1}-1} \lesssim \operatorname{ecc}(R)^{p'\theta_2}.$$

Since we can let $\operatorname{ecc}(R) \to \infty$, it means that we must have $\alpha - (p-1) - p\theta_2 \le 0$. Letting $\alpha \to 2(p-1)$ in this inequality gives $\theta_2 \ge 1/p'$. Then $\theta_2 \ge 1$ follows by letting $p \to \infty$.

Thus, weighted boundedness cannot hold in general for CZX operators if $\theta_2 < 1$. Next, we prove some complementary sparse estimates, which refine the weighted estimates in the case $\theta_2 = 1$.

6.4 Definition We say that an operator T satisfies pointwise L^p-sparse domination with constants C and ϵ, if for every compactly supported $f \in L^p(\mathbb{R}^d)$, there exists an ϵ-sparse collection $\mathscr{S}$ of cubes such that

$$|Tf| \le C \sum_{S \in \mathscr{S}} \langle f \rangle_{S,p} 1_S,$$

where $\langle f \rangle_{S,p} := \langle |f|^p \rangle_S^{1/p}$, and a collection $\mathscr{S}$ of cubes is called ϵ-sparse, if there are pairwise disjoint sets $E(S) \subset S$ for every $S \in \mathscr{S}$ with $|E(S)| \ge \epsilon |S|$.

There are by now several approaches to proving sparse domination. We will use one by Lerner and Ombrosi [14], which depends on bounds on the following maximal function related to the operator T under investigation

$$\mathcal{M}^{\#}_{T,3} f(x) := \sup_{J \ni x} \operatorname*{ess\,sup}_{y,z \in J} |T(1_{(3J)^c} f)(y) - T(1_{(3J)^c} f)(z)|,$$

where the supremum, once again, is over all cubes J.

6.5 Lemma *Let T be an operator with a CZX kernel satisfying $\theta_2 = 1$. Then*

$$\mathcal{M}^{\#}_{T,3} f(x) \lesssim M_* f(x) := \sup_{R \ni x} \fint_R |f(y)| \,\mathrm{d}y,$$

where the right-hand side is the strong maximal function, with supremum over all (axes-parallel) rectangles $R \subset \mathbb{R}^2$ containing x.

Proof Let us fix a cube $J \subset \mathbb{R}^2$ with centre c_J, and some $x, y, z \in J$. Note that

$$\begin{aligned} &T(1_{(3J)^c} f)(y) - T(1_{(3J)^c} f)(z) \\ &\quad = [T(1_{(3J)^c} f)(y) - T(1_{(3J)^c} f)(c_J)] - [T(1_{(3J)^c} f)(z) - T(1_{(3J)^c} f)(c_J)] \end{aligned}$$

and

$$T(1_{(3J)^c} f)(y) - T(1_{(3J)^c} f)(c_J) = \int_{(3J)^c} [K(y,u) - K(c_J,u)] f(u) \,\mathrm{d}u.$$

As usual, we split

$$(3J)^c = [(3J^1)^c \times 3J^2] \cup [3J^1 \times (3J^2)^c] \cup [(3J^1)^c \times (3J^2)^c].$$

For the integral over the last component, Lemma 2.2 implies that

$$\begin{aligned} &\int_{(3J^1)^c \times (3J^2)^c} |[K(y,u) - K(c_J,u)] f(u)| \,\mathrm{d}u \\ &\quad \lesssim \int_{(3J^1)^c \times (3J^2)^c} \prod_{i=1}^{2} \frac{\ell(J)^\theta}{\operatorname{dist}(u^i, J^i)^{1+\theta}} |f(u)| \,\mathrm{d}u \\ &\quad \lesssim \sum_{k^1,k^2=1}^{\infty} \int_{(3^{k^1+1}J^1 \setminus 3^{k^1}J^1) \times (3^{k^2+1}J^2 \setminus 3^{k^2}J^2)} \prod_{i=1}^{2} \frac{\ell(J)^\theta}{(3^{k^i}\ell(J))^{1+\theta}} |f(u)| \,\mathrm{d}u \\ &\quad \lesssim \sum_{k^1,k^2=1}^{\infty} 3^{-(k^1+k^2)\theta} \fint_{3^{k^1+1}J^1 \times 3^{k^2+1}J^2} |f(u)| \,\mathrm{d}u \\ &\quad \lesssim \sum_{k^1,k^2=1}^{\infty} 3^{-(k^1+k^2)\theta} M_* f(x) \lesssim M_* f(x). \end{aligned}$$

For the other components, using that $\theta_2 = 1$ we get directly from the size estimate that

$$
\begin{aligned}
&\int_{(3J^1)^c\times 3J^2} |K(y,u)f(u)|\,\mathrm{d}u \\
&\lesssim \int_{(3J^1)^c\times 3J^2} \frac{1}{|y^1-u^1|^2}|f(u)|\,\mathrm{d}u \\
&\lesssim \sum_{k^1=1}^{\infty} \int_{(3^{k^1+1}J^1\setminus 3^{k^1}J^1)\times 3J^2} \frac{1}{(3^{k^1}\ell(J))^2}|f(u)|\,\mathrm{d}u \\
&\lesssim \sum_{k^1=1}^{\infty} 3^{-k^1} \fint_{3^{k^1+1}J^1\times 3J^2} |f(u)|\,\mathrm{d}u \\
&\lesssim \sum_{k^1=1}^{\infty} 3^{-k^1} M_* f(x) \lesssim M_* f(x).
\end{aligned}
$$

Altogether, we have checked that

$$|T(1_{(3J)^c}f)(y) - T(1_{(3J)^c}f)(z)| \lesssim M_* f(x),$$

and taking the supremum over $y, z \in J$ and then over $J \ni x$ we see that

$$\mathcal{M}^{\#}_{T,3} f(x) \lesssim M_* f(x).$$

□

We now quote a slight variant of a result of Lerner and Ombrosi [14, Theorem 1.1]:

6.6 Theorem (Lerner and Ombrosi [14]) *Let T be a sublinear operator that is bounded from $L^q(\mathbb{R}^d)$ to $L^{q,\infty}(\mathbb{R}^d)$, and such that $\mathcal{M}^{\#}_{T,3}$ is bounded from $L^r(\mathbb{R}^d)$ to $L^{r,\infty}(\mathbb{R}^d)$ for some $1 \le q, r < \infty$. Let $s = \max(q, r)$. Then T satisfies pointwise L^s-sparse domination with constants*

$$C = c_d(\|T\|_{L^q\to L^{q,\infty}} + \|\mathcal{M}^{\#}_{T,3}\|_{L^r\to L^{r,\infty}}), \qquad \epsilon = \epsilon_d.$$

Proof This is essentially [14, Theorem 1.1], except for some details mainly related to the constant C. Since this constant will be relevant to us below, we will explain the necessary changes. On a more trivial side, the statement in [14] involves an additional parameter α in the maximal operator $\mathcal{M}^{\#}_{T,\alpha}$; we simply take $\alpha = 3$. Also, in [14] the $L^q(\mathbb{R}^d)$-to-$L^{q,\infty}(\mathbb{R}^d)$ boundedness is replaced by a certain "W_q condition"; however, this follows from the $L^q(\mathbb{R}^d)$-to-$L^{q,\infty}(\mathbb{R}^d)$ boundedness, as pointed out shortly before [14, Theorem 1.1].

More seriously, the bound for C given in [14] has dependencies on additional parameters that we wish to avoid. For this, it is necessary to inspect the proof of [14, Theorem 1.1.]. The said proof provides the expression

$$C = (3 + c)A,$$

where c and A need to be chosen so that each of the sets

$$\{M(|f|^s 1_{3Q})^{1/s} > c\langle |f|\rangle_{3Q,s}\}, \quad \{|T(f1_{3Q})| > A\langle |f|\rangle_{3Q,s}\},$$
$$\{\mathcal{M}^{\#}_{T,3}(f1_{3Q}) > A\langle |f|\rangle_{3Q,s}\}$$

have measure at most $\epsilon_d|Q|$ for some small dimensional ϵ_d. However,

$$|\{M(|f|^s 1_{3Q}) > (c\langle |f|\rangle_{3Q,s})^s\}| \leq \frac{C_d}{(c\langle |f|\rangle_{3Q,s})^s}\||f|^s 1_{3Q}\|_1 = \frac{C_d}{c^s}|3Q| = \frac{C_d 3^d}{c^s}|Q|,$$

where $C_d = \|M\|_{L^1(\mathbb{R}^d)\to L^{1,\infty}(\mathbb{R}^d)} \geq 1$, so that we can take

$$c = \Big(\frac{C_d 3^d}{\epsilon_d}\Big) \geq \Big(\frac{C_d 3^d}{\epsilon_d}\Big)^{1/s},$$

since $s \geq 1$. Thus $c = c_d$. On the other hand, using among other things that $q \leq s$ and Hölder's inequality, there holds that

$$\begin{aligned}|\{|T(f1_{3Q})| > A\langle |f|\rangle_{3Q,s}\}| &\leq \frac{\|T\|^q_{L^q\to L^{q,\infty}}}{(A\langle |f|\rangle_{3Q,s})^q}\|f1_{3Q}\|^q_{L^q} \leq \frac{\|T\|^q_{L^q\to L^{q,\infty}}}{(A\langle |f|\rangle_{3Q,q})^q}\|f1_{3Q}\|^q_{L^q} \\ &= \frac{\|T\|^q_{L^q\to L^{q,\infty}}}{A^q}|3Q| = \frac{3^d\|T\|^q_{L^q\to L^{q,\infty}}}{A^q}|Q|,\end{aligned}$$

so to make this at most ϵ_d, it is enough to take

$$A \geq \frac{3^d}{\epsilon_d}\|T\|_{L^q\to L^{q,\infty}} \geq \Big(\frac{3^d}{\epsilon_d}\Big)^{1/q}\|T\|_{L^q\to L^{q,\infty}},$$

since $q \geq 1$. Similarly, with $\mathcal{M}^{\#}_{T,3}$ in place of T and r in place of q in order that $|\{\mathcal{M}^{\#}_{T,3} > A\langle |f|\rangle_{3Q,s}\}| \leq \epsilon_d|Q|$, it is enough to take

$$A \geq \frac{3^d}{\epsilon_d}\|\mathcal{M}^{\#}_{T,3}\|_{L^r\to L^{r,\infty}}.$$

So an admissible choice is $c = c_d$ and

$$A = \frac{3^d}{\epsilon_d}(\|T\|_{L^q\to L^{q,\infty}} + \|\mathcal{M}^{\#}_{T,3}\|_{L^r\to L^{r,\infty}}),$$

and hence

$$(3+c)A = C_d(\|T\|_{L^q\to L^{q,\infty}} + \|\mathcal{M}^{\#}_{T,3}\|_{L^r\to L^{r,\infty}}).$$

□

An immediate consequence of the previous results is the following:

6.7 Corollary *Let $T \in \mathcal{L}(L^2(\mathbb{R}^2))$ be an operator with a CZX kernel satisfying $\theta_2 = 1$. Then for every $p > 1$, the operator T satisfies pointwise L^p-sparse domination with constants $C \lesssim p'$ and an absolute $\epsilon > 0$.*

Proof We know from Theorem 1.3 that T is bounded from $L^1(\mathbb{R}^2)$ to $L^{1,\infty}(\mathbb{R}^2)$. From Lemma 6.5 we know that $\mathcal{M}^{\#}_{T,3} f \lesssim M_* f$. Since the strong maximal operator is bounded from $L^p(\mathbb{R}^2)$ to itself, and hence to $L^{p,\infty}(\mathbb{R}^2)$, the assumptions of Theorem 6.6 are satisfied. Since $d = 2$, it is immediate that $\epsilon = \epsilon_2$ provided by that theorem is absolute. In order to obtain the claim $C \lesssim p'$, we observe for completeness the following (probably well known) estimate for $M_* f \le M^2 M^1 f$, where M^i is the one-dimensional maximal operator with respect to the ith variable:

$$\begin{aligned} |\{M_* f > \lambda\}| &= \int_{\mathbb{R}} |\{y : M_* f(x, y) > \lambda\}| \, \mathrm{d}x \le \int_{\mathbb{R}} |\{y : M^2 M^1 f(x, y) > \lambda\}| \, \mathrm{d}x \\ &\le \int_{\mathbb{R}} \lambda^{-p} (\|M^2\|_{L^p \to L^{p,\infty}} \|M^1 f(x, \cdot)\|_{L^p(\mathbb{R})})^p \, \mathrm{d}x \\ &= \lambda^{-p} \|M^2\|^p_{L^p \to L^{p,\infty}} \|M^1 f\|^p_{L^p(\mathbb{R}^2)} \\ &\le \lambda^{-p} \|M^2\|^p_{L^p \to L^{p,\infty}} (\|M^1\|_{L^p \to L^p} \|f\|_{L^p(\mathbb{R}^2)})^p. \end{aligned}$$

It follows that

$$\|M_*\|_{L^p(\mathbb{R}^2) \to L^{p,\infty}(\mathbb{R}^2)} \le \|M^2\|_{L^p \to L^{p,\infty}} \|M^1\|_{L^p \to L^p} \lesssim 1 \cdot p',$$

using the fact that the L^p norm of the usual maximal operator is $O(p')$, while its L^p-to-$L^{p,\infty}$ norm can be estimated independently of p. In fact,

$$\begin{aligned} |\{Mf > \lambda\}| &= |\{(Mf)^p > \lambda^p\}| \le |\{M(|f|^p) > \lambda^p\}| \\ &\le \lambda^{-p} \|M\|_{L^1 \to L^{1,\infty}} \||f|^p\|_{L^1} = \lambda^{-p} \|M\|_{L^1 \to L^{1,\infty}} \|f\|^p_{L^p}, \end{aligned}$$

and hence

$$\|M\|_{L^p \to L^{p,\infty}} \le \|M\|^{1/p}_{L^1 \to L^{1,\infty}} \le \|M\|_{L^1 \to L^{1,\infty}}.$$

□

6.8 Corollary *Let $T \in \mathcal{L}(L^2(\mathbb{R}^2))$ be an operator with a CZX kernel satisfying $\theta_2 = 1$. Then for every $p \in (1, \infty)$ and every $w \in A_p(\mathbb{R}^2)$, the operator T extends boundedly to $L^p(w)$ with norm*

$$\|T\|_{\mathcal{L}(L^p(w))} \lesssim_p [w]^{p'}_{A_p}.$$

Proof This follows by the same reasoning as [17, Theorem 1.6]. The said theorem is stated for different operators, but its proof only uses a certain sparse domination estimate for these operators, which is a slightly weaker variant (so-called form domination) of what we proved (pointwise domination) for operators with CZX kernel in Corollary 6.7, and hence the same reasoning applies to the case at hand. □

A curious feature of the above proof of the weighted estimates is that it passes through estimates involving the strong maximal operator, which in principle should be forbidden in the theory of standard A_p weights; indeed, the strong maximal operator is bounded in $L^p(w)$ for strong A_p weights only. The resolution of this paradox is that we only use the strong maximal operator as an intermediate step, in a part of the argument with no weights yet present, to establish some sparse bounds, which in turn imply the weighted estimates.

To conclude this section, we discuss commutator estimates. In fact, combining the ideas from [15] and [14] we can establish the following sparse domination principle.

6.9 Proposition *Let T be a linear operator that is bounded from $L^q(\mathbb{R}^d)$ to $L^{q,\infty}(\mathbb{R}^d)$, and such that $\mathcal{M}^\#_{T,3}$ is bounded from $L^r(\mathbb{R}^d)$ to $L^{r,\infty}(\mathbb{R}^d)$ for some $1 \le q, r < \infty$. Let $s = \max(q, r)$. Then there exists an ϵ-sparse family $\mathscr{S}$ such that the commutator $[b, T]f := bTf - T(bf)$ satisfies*

$$|[b,T]f| \le C\left(\sum_{S\in\mathscr{S}} \langle|(b-\langle b\rangle_S)f|\rangle_{S,s} 1_S + \sum_{S\in\mathscr{S}} |b-\langle b\rangle_S|\langle|f|\rangle_{S,s} 1_S\right),$$

with constants

$$C = c_d(\|T\|_{L^q\to L^{q,\infty}} + \|\mathcal{M}^\#_{T,3}\|_{L^r\to L^{r,\infty}}), \qquad \epsilon = \epsilon_d.$$

Proof The proof is actually similar to [14, Theorem 1.1] but one should adapt ideas used in [15]. Let c, A, ϵ_d be the same as those in the proof of Theorem 6.6. Apart from the sets

$$\{M(|f|^s 1_{3Q})^{1/s} > c\langle|f|\rangle_{3Q,s}\},$$
$$\{|T(f1_{3Q})| > A\langle|f|\rangle_{3Q,s}\}, \quad \{\mathcal{M}^\#_{T,3}(f1_{3Q}) > A\langle|f|\rangle_{3Q,s}\},$$

for the same reason, we can also let each of the sets

$$\{M(|(b-\langle b\rangle_{3Q})f|^s 1_{3Q})^{1/s} > c\langle|(b-\langle b\rangle_{3Q})f|\rangle_{3Q,s}\},$$
$$\{|T((b-\langle b\rangle_{3Q})f1_{3Q})| > A\langle|(b-\langle b\rangle_{3Q})f|\rangle_{3Q,s}\},$$
$$\{\mathcal{M}^\#_{T,3}((b-\langle b\rangle_{3Q})f1_{3Q}) > A\langle|(b-\langle b\rangle_{3Q})f|\rangle_{3Q,s}\}$$

have measure at most $\epsilon_d|Q|$. Denote the union of the above six sets by E and set $\Omega = E \cap Q$. Manipulating in the same way as in [14, Theorem 1.1], we get a family of pairwise disjoint cubes $\{P_j\} \subset Q$ such that $\sum_j |P_j| \le \frac{1}{2}|Q|$ and $|\Omega \setminus \cup_j P_j| = 0$. The latter implies that for a.e. $x \in Q \setminus \cup_j P_j$ there holds that

$$|T(f1_{3Q})(x)| \le A\langle|f|\rangle_{3Q,s}, \quad |T((b-\langle b\rangle_{3Q})f1_{3Q})(x)| \le A\langle|(b-\langle b\rangle_{3Q})f|\rangle_{3Q,s}.$$

On the other hand, similarly as in [14, Theorem 1.1] we also have for a.e. $x \in P_j$ that

$$|T(f1_{3Q\setminus 3P_j})(x)| \le (2+c)A\langle |f|\rangle_{3Q,s},$$
$$|T((b-\langle b\rangle_{3Q})f1_{3Q\setminus 3P_j})(x)| \le (2+c)A\langle |(b-\langle b\rangle_{3Q})f|\rangle_{3Q,s}.$$

Thus

$$\begin{aligned}
|[b,T](f1_{3Q})|1_Q(x) &= |[b,T](f1_{3Q})|1_{Q\setminus\cup_j P_j}(x) \\
&\quad + \sum_j |[b,T](f1_{3Q\setminus 3P_j})|1_{P_j} \\
&\quad + \sum_j |[b,T](f1_{3P_j})|1_{P_j} \\
&= |[b-\langle b\rangle_{3Q},T](f1_{3Q})|1_{Q\setminus\cup_j P_j}(x) \\
&\quad + \sum_j |[b-\langle b\rangle_{3Q},T](f1_{3Q\setminus 3P_j})|1_{P_j} \\
&\quad + \sum_j |[b,T](f1_{3P_j})|1_{P_j} \\
&\le (3+c)A|b-\langle b\rangle_{3Q}|\langle |f|\rangle_{3Q,s}1_Q + (3+c)A\langle |(b-\langle b\rangle_{3Q})f|\rangle_{3Q,s}1_Q \\
&\quad + \sum_j |[b,T](f1_{3P_j})|1_{P_j}.
\end{aligned}$$

Note that the linearity of T is used in the second step. With the recursive inequality at hand, the rest is standard (see e.g. [14, Lemma 2.1]). And since the constants c and A are the same as those in Theorem 6.6, we get the desired constant in the sparse domination. □

Analogous to Corollary 6.7, we have the following result

6.10 Proposition *Let $T \in \mathcal{L}(L^2(\mathbb{R}^2))$ be an operator with a CZX kernel satisfying $\theta_2 = 1$. Then for every $p > 1$, the commutator $[b, T]$ satisfies*

$$|[b,T]f| \le Cp' \left(\sum_{S\in\mathscr{S}} \langle |(b-\langle b\rangle_S)f|\rangle_{S,p}1_S + \sum_{S\in\mathscr{S}} |b-\langle b\rangle_S|\langle |f|\rangle_{S,p}1_S \right),$$

where $\mathscr{S}$ is an ϵ-sparse family with $\epsilon > 0$ absolute and C is an absolute constant depending only on the operator T.

In [16, Theorem 5.2] a two-weight commutator estimate for rough homogeneous singular integrals was formulated. As our sparse forms are the same as there, the two-weight commutator estimate of Theorem 1.5 follows as a direct consequence of Proposition 6.10.

Acknowledgements T. Hytönen and E. Vuorinen were supported by the Academy of Finland through Project Numbers 314829 and 346314, and by the University of Helsinki internal grants for the Centre of Excellence in Analysis and Dynamics Research and the Finnish Centre of Excellence in Randomness and

Structures “FiRST”. K. Li was supported by the National Natural Science Foundation of China through Project Numbers 12222114 and 12001400.

Funding Open Access funding provided by University of Helsinki including Helsinki University Central Hospital.

References

1. David, G., Journé, J.-L.: A boundedness criterion for generalized Calderón–Zygmund operators. Ann. Math. **120**, 371–397 (1984)
2. Duoandikoetxea, J.: Extrapolation of weights revisited: new proofs and sharp bounds. J. Funct. Anal. **260**, 1886–1901 (2011)
3. Duong, X.T., Li, J., Ou, Y., Pipher, J., Wick, B.: Weighted estimates of singular integrals and commutators in the Zygmund dilation setting. arXiv:1905.00999 (2019)
4. Fefferman, R., Pipher, J.: Multiparameter operators and sharp weighted inequalities. Am. J. Math. **11**, 337–369 (1997)
5. Fefferman, R., Stein, E.: Singular integrals on product spaces. Adv. Math. **45**, 117–143 (1982)
6. Grau de la Herrán, A.: Comparison of $T1$ conditions for multi-parameter operators. Proc. Am. Math. Soc. **144**, 2437–2443 (2016)
7. Grau de la Herrán, A., Hytönen, T.: Dyadic representation and boundedness of non-homogeneous Calderón–Zygmund operators with mild kernel regularity. Michigan Math. J. **67**, 757–786 (2018)
8. Han, Y., Li, J., Lin, C.-C., Tan, C.: Singular integrals associated with Zygmund dilations. J. Geom. Anal. **29**, 2410–2455 (2019)
9. Han, Y., Li, J., Lin, C.-C., Tan, C., Wu, X.: Weighted endpoint estimates for singular integral operators associated with Zygmund dilations. Taiwanese J. Math. **23**, 375–408 (2019)
10. Hytönen, T.: The sharp weighted bound for general Calderón–Zygmund operators. Ann. Math. **175**, 1473–1506 (2012)
11. Hytönen, T., Martikainen, H., Vuorinen, E.: Multi-parameter estimates via operator-valued shifts. Proc. Lond. Math. Soc. **119**(6), 1560–1597 (2019)
12. Hytönen, T., Li, K., Martikainen, H., Vuorinen, E.: Multiresolution analysis and Zygmund dilations. arXiv:2203.15777 (2022)
13. Journé, J.-L.: Calderón–Zygmund operators on product spaces. Rev. Mat. Iberoam. **1**, 55–91 (1985)
14. Lerner, A.K., Ombrosi, S.: Some remarks on the pointwise sparse domination. J. Geom. Anal. **30**(1), 1011–1027 (2020)
15. Lerner, A.K., Ombrosi, S., Rivera-Ríos, I.: On pointwise and weighted estimates for commutators of Calderón–Zygmund operators. Adv. Math. **319**, 153–181 (2017)
16. Li, K.: Multilinear commutators in the two-weight setting. Bull. Lond. Math. Soc. **54**, 568–589 (2022)
17. Li, K., Pérez, C., Rivera-Ríos, I.P., Roncal, L.: Weighted norm inequalities for rough singular integral operators. J. Geom. Anal. **29**(3), 2526–2564 (2019)
18. Martikainen, H.: Representation of bi-parameter singular integrals by dyadic operators. Adv. Math. **229**, 1734–1761 (2012)
19. Müller, D., Ricci, F., Stein, E.M.: Marcinkiewicz multipliers and multi-parameter structure on Heisenberg(-type) groups, I. Invent. Math. **119**, 119–233 (1995)
20. Nagel, A., Wainger, S.: L^2 boundedness of Hilbert transforms along surfaces and convolution operators homogeneous with respect to a multiple parameter group. Am. J. Math. **99**, 761–785 (1977)

21. Ricci, F., Stein, E.M.: Multiparameter singular integrals and maximal functions. Ann. Inst. Fourier (Grenoble) **42**, 637–670 (1992)
22. Stein, E.M.: Harmonic Analysis: Real-variable Methods, Orthogonality, and Oscillatory Integrals. Princeton University Press, Princeton (1993)
23. Stein, E.M., Weiss, G.: Interpolation of operators with change of measures. Trans. Am. Math. Soc. **87**, 159–172 (1958)