The Journal of Geometric Analysis (2021) 31:980–1038
https://doi.org/10.1007/s12220-019-00308-x

Two Weight Commutators on Spaces of Homogeneous Type and Applications

Xuan Thinh Duong[1] · Ruming Gong[2] · Marie-Jose S. Kuffner[3] · Ji Li[1] · Brett D. Wick[4] · Dongyong Yang[5]

Received: 25 September 2018 / Published online: 11 November 2019

Abstract
In this paper, we establish the two weight commutator theorem of Calderón–Zygmund operators in the sense of Coifman–Weiss on spaces of homogeneous type, by studying the weighted Hardy and BMO space for A_2 weights and by proving the sparse operator domination of commutators. The main tool here is the Haar basis, the adjacent dyadic systems on spaces of homogeneous type, and the construction of a suitable version of a sparse operator on spaces of homogeneous type. As applications, we provide a two weight commutator theorem (including the high order commutators) for the following Calderón–Zygmund operators: Cauchy integral operator on $\mathbb{R}$, Cauchy–Szegö projection operator on Heisenberg groups, Szegö projection operators on a family of unbounded weakly pseudoconvex domains, the Riesz transform associated with the sub-Laplacian on stratified Lie groups, as well as the Bessel Riesz transforms (in one and several dimensions).

Keywords BMO · Commutator · Two weights · Hardy space · Factorization

Mathematics Subject Classification 42B30 · 42B20 · 42B35

1 Introduction and Statement of Main Results

It is well-known that Coifman et al. [8] characterized the boundedness of the commutator $[b, R_j]$ acting on Lebesgue spaces in terms of BMO, where $R_j = \frac{\partial}{\partial x_j}\Delta^{-1/2}$ is the jth Riesz transform on the Euclidean space $\mathbb{R}^n$. Their result extended the work of Nehari [40] about Hankel operators from complex setting to the real setting $\mathbb{R}^n$. Later, Bloom [3] established the characterisation of weighted BMO in terms of boundedness

✉ Dongyong Yang
dyyang@xmu.edu.cn

Extended author information available on the last page of the article

of commutators $[b, H]$ in the two weight setting, where H is the Hilbert transform on $\mathbb{R}$.

Recent remarkable results were achieved by Holmes–Lacey–Wick [21] giving the characterisation of weighted BMO space on $\mathbb{R}^n$ in terms of boundedness of commutators of Riesz transforms, and by Lerner–Ombrosi–Rivera-Ríos [32,33] in terms of boundedness of commutators of Calderón–Zygmund operators with homogeneous kernels $\Omega(\frac{x}{|x|})\frac{1}{|x|^n}$, and Hytönen [25] in terms of boundedness of commutators of a more general version of Calderón–Zygmund operators and weighted BMO functions on $\mathbb{R}^n$. Meanwhile, the two weight commutator has also been studied extensively in different settings, see for example [13,15,20].

We note that to get the lower bound of the two weight commutator for Riesz transforms (or the Hilbert transform in one dimension) in terms of the weighted BMO space, the first proofs used spherical harmonics to expand the Riesz (Hilbert) kernels, which relies on properties of the Fourier transform of the Riesz (Hilbert) kernels. A similar method of expansion of the Riesz transform associated with Neumann Laplacian was used in [13] for a larger class of A_p weights and for the BMO space associated with Neumann Laplacian which is strictly larger than classical BMO. In [32], concerning the two weight commutator for Calderón–Zygmund operators associated with homogeneous kernel $\Omega(\frac{x}{|x|})\frac{1}{|x|^n}$, the proof of the lower bound was obtained by assuming suitable conditions on the homogeneous function Ω, see also [19,20]. More recently, Hytönen [25] studied the two weight commutator for Calderón–Zygmund operators and proposed a condition denoted by the "non-degenerate Calderón–Zygmund kernel", then he proved the lower bound of the commutator by constructing a factorisation. Also, in [14,15], they established a version of a pointwise kernel lower bound for the Riesz transform associated to the sub-Laplacian on stratified Lie groups, which covers the Heisenberg group, and used this kernel lower bound to obtain the two weight commutator result following the idea in [32].

However, there are other important Calderón–Zygmund operators (not built on the Euclidean space setting) whose kernels do not have a connection to the Fourier transform and are not of homogeneous type such as $\Omega(\frac{x}{|x|})\frac{1}{|x|^n}$. Moreover, whether the kernels fall into Hytönen's "non-degenerate Calderón–Zygmund kernel" has not been studied before and hence the two weight commutator estimates and higher order commutator are unknown.

For example, the Riesz transform from Muckenhoupt–Stein [39]: $R_\lambda := -\frac{d}{dx}(\Delta_\lambda)^{-\frac{1}{2}}$, associated with the Bessel operator on $\mathbb{R}_+$:

$$\Delta_\lambda := -\frac{d^2}{dx^2} - \frac{2\lambda}{x}\frac{d}{dx}, \quad x > 0, \lambda > -\frac{1}{2},$$

and the Riesz transform $R_{\lambda,j} = \frac{d}{dx_j}(\Delta_\lambda^{(n+1)})^{-\frac{1}{2}},\ j = 1, \ldots, n+1$, associated with the Bessel operator $\Delta_\lambda^{(n+1)}$ on $\mathbb{R}_+^{n+1}$ studied in Huber [23]:

$$\Delta_\lambda^{(n+1)} = -\frac{d^2}{dx_1^2} \cdots - \frac{d^2}{dx_n^2} - \frac{d^2}{dx_{n+1}^2} - \frac{2\lambda}{x_{n+1}}\frac{d}{dx_{n+1}}.$$

Another example is the Cauchy–Szegö projection operator $\mathcal{C}$ (for all the notation below we refer to Section 2 in Chapter XII in Stein [43]), which is the orthogonal projection from $L^2(b\mathcal{U}^n)$ to the subspace of functions $\{F^b\}$ that are boundary values of functions $F \in \mathcal{H}^2(\mathcal{U}^n)$. The associated Cauchy–Szegö kernel is as follows:

$$\mathcal{C}(f)(x) = \int_{\mathbb{H}^n} K(y^{-1} \circ x) f(y) dy,$$

where $K(x) = -\frac{\partial}{\partial t}\left(\frac{c}{n}[t + i|\zeta|^2]^{-n}\right)$ for $x = [\zeta, t] \in \mathbb{H}^n = \mathbb{C}^n \times \mathbb{R}$.

Then, it is natural to study the following question: Can one establish the characterisation of boundedness of two weight commutators in terms of the related weighted BMO space for Calderón–Zygmund operators T in a more general setting such that many examples, including the Bessel Riesz transform, the Cauchy–Szegö projection operator on Heisenberg groups and others, can be covered?

To address this question we work in a general setting: spaces of homogeneous type introduced by Coifman and Weiss in the early 1970s, in [9], see also [10]. We say that (X, d, μ) is a space of homogeneous type in the sense of Coifman and Weiss if d is a quasi-metric on X and μ is a nonzero measure satisfying the doubling condition. A *quasi-metric* d on a set X is a function $d : X \times X \longrightarrow [0, \infty)$ satisfying (i) $d(x, y) = d(y, x) \geq 0$ for all x, $y \in X$; (ii) $d(x, y) = 0$ if and only if $x = y$; and (iii) the *quasi-triangle inequality*: there is a constant $A_0 \in [1, \infty)$ such that for all x, $y, z \in X$,

$$d(x, y) \leq A_0[d(x, z) + d(z, y)]. \tag{1.1}$$

We say that a nonzero measure μ satisfies the *doubling condition* if there is a constant C_μ such that for all $x \in X$ and $r > 0$,

$$\mu(B(x, 2r)) \leq C_\mu \mu(B(x, r)) < \infty, \tag{1.2}$$

where $B(x, r)$ is the quasi-metric ball by $B(x, r) := \{y \in X : d(x, y) < r\}$ for $x \in X$ and $r > 0$. We point out that the doubling condition (1.2) implies that there exists a positive constant n (the *upper dimension* of μ) such that for all $x \in X$, $\lambda \geq 1$ and $r > 0$,

$$\mu(B(x, \lambda r)) \leq C_\mu \lambda^n \mu(B(x, r)). \tag{1.3}$$

Throughout this paper we assume that $\mu(X) = \infty$ and that $\mu(\{x_0\}) = 0$ for every $x_0 \in X$.

We now recall the singular integral operator on spaces of homogeneous type in the sense of Coifman and Weiss.

Definition 1.1 We say that T is a Calderón–Zygmund operator on (X, d, μ) if T is bounded on $L^2(X)$ and has the associated kernel $K(x, y)$ such that $T(f)(x) =$

$\int_X K(x, y) f(y) d\mu(y)$ for any $x \notin \operatorname{supp} f$, and $K(x, y)$ satisfies the following estimates: for all $x \neq y$,

$$|K(x, y)| \leq \frac{C}{V(x, y)}, \tag{1.4}$$

and for $d(x, x') \leq (2A_0)^{-1} d(x, y)$,

$$|K(x, y) - K(x', y)| + |K(y, x) - K(y, x')| \leq \frac{C}{V(x, y)} \omega\Big(\frac{d(x, x')}{d(x, y)}\Big), \tag{1.5}$$

where $V(x, y) = \mu(B(x, d(x, y)))$ and by the doubling condition we have that $V(x, y) \approx V(y, x)$, $\omega : [0, 1] \to [0, \infty)$ is continuous, increasing, subadditive, $\omega(0) = 0$.

We say that ω satisfies the Dini condition if $\int_0^1 \omega(t) \frac{dt}{t} < \infty$.

Let T be a Calderón–Zygmund operator on X. Suppose $b \in L^1_{\mathrm{loc}}(X)$ and $f \in L^p(X)$. Let $[b, T]$ be the commutator defined by

$$[b, T] f(x) := b(x) T(f)(x) - T(bf)(x).$$

The iterated commutators T_b^m, $m \in \mathbb{N}$, are defined inductively by

$$T_b^m f(x) := [b, T_b^{m-1}] f(x), \quad T_b^1 f(x) := [b, T] f(x).$$

Next, we use A_p, $1 \leq p \leq \infty$, to denote the Muckenhoupt weighted class on X (see the precise definition of A_p in Sect. 2), and the weighted BMO on X is defined as follows (the Euclidean version of weighted BMO was first introduced by Muckenhoupt and Wheeden [38]).

Definition 1.2 Suppose $w \in A_\infty$. A function $b \in L^1_{\mathrm{loc}}(X)$ belongs to $\mathrm{BMO}_w(X)$ if

$$\|b\|_{\mathrm{BMO}_w(X)} := \sup_B \frac{1}{w(B)} \int_B |b(x) - b_B| \, d\mu(x) < \infty,$$

where $b_B := \frac{1}{\mu(B)} \int_B b(x) d\mu(x)$ and the supremum is taken over all balls $B \subset X$.

Our first main result is the following theorem.

Theorem 1.3 *Suppose* $1 < p < \infty$, $\lambda_1, \lambda_2 \in A_p$, $\nu := \lambda_1^{\frac{1}{p}} \lambda_2^{-\frac{1}{p}}$ *and* $m \in \mathbb{N}$. *Suppose* $b \in \mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)$. *Then for any Calderón–Zygmund operator* T *as in Definition* 1.1 *with* ω *satisfying the Dini condition, there exists a positive constant* C_1 *such that*

$$\|T_b^m : L^p_{\lambda_1}(X) \to L^p_{\lambda_2}(X)\| \leq C_1 \|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \Big([\lambda_1]_{A_p} [\lambda_2]_{A_p}\Big)^{\frac{m+1}{2} \cdot \max\{1, \frac{1}{p-1}\}}.$$

To obtain the upper bound, we characterise the sparse system and then use the idea from [32] to build a suitable version of a sparse operator on a space of homogeneous type. Here, we apply the tool of adjacent dyadic systems from [26], the explicit construction of Haar basis from [28], and we have to allow suitable overlapping for the sparse sets due to the partition and covering of the whole space via quasi-metric balls.

To consider the lower bound of the commutator, we assume that the Calderón–Zygmund operator T, as given in Definition 1.1 with ω satisfying $\omega(t) \to 0$ as $t \to 0$, satisfies the following "non-degenerate" condition:

There exist positive constants c_0 and $\overline{C}$ such that for every $x \in X$ and $r > 0$, there exists $y \in B(x, \overline{C}r)\backslash B(x, r)$ satisfying

$$|K(x, y)| \geq \frac{1}{c_0\mu(B(x, r))}. \tag{1.6}$$

Note that in $\mathbb{R}^n$, this "non-degenerate" condition was first proposed in [24], and a similar assumption on the behaviour of the kernel lower bound was proposed in [32]. On stratified Lie groups, a similar condition of the Riesz transform kernel lower bound was verified in [14].

Then, we have the following lower bound.

Theorem 1.4 *Suppose* $1 < p < \infty$, $\lambda_1, \lambda_2 \in A_p$, $\nu := \lambda_1^{\frac{1}{p}} \lambda_2^{-\frac{1}{p}}$ *and* $m \in \mathbb{N}$. *Suppose* $b \in L^1_{\rm loc}(X)$ *and that* T *is a Calderón–Zygmund operator as in Definition* 1.1 *and satisfies the non-degenerate condition* (1.6). *Also suppose that* T_b^m *is bounded from* $L^p_{\lambda_1}(X)$ *to* $L^p_{\lambda_2}(X)$. *Then* $b \in {\rm BMO}_{\nu^{\frac{1}{m}}}(X)$, *and there exists a positive constant* C_2 *such that*

$$\|b\|^m_{{\rm BMO}_{\nu^{\frac{1}{m}}}(X)} \leq C_2 \|T_b^m : L^p_{\lambda_1}(X) \to L^p_{\lambda_2}(X)\|.$$

Based on the characterisation of ${\rm BMO}_\nu(X)$ via commutators $T_b^1 = [b, T]$, we further have the weak factorisation for the weighted Hardy space $H^1_\nu(X)$ as follows.

Theorem 1.5 *Suppose* $1 < p < \infty$, $\lambda_1, \lambda_2 \in A_p$ *and* $\nu := \lambda_1^{\frac{1}{p}} \lambda_2^{-\frac{1}{p}}$. *Let* p' *be the conjugate of* p *and* $\lambda_2' := \lambda_2^{-\frac{1}{p-1}}$. *For any* $f \in H^1_\nu(X)$, *there exist numbers* $\{\alpha_j^k\}_{k,j}$, *functions* $\{g_j^k\}_{k,j} \subset L^p_{\lambda_1}(X)$ *and* $\{h_j^k\}_{k,j} \subset L^{p'}_{\lambda_2'}(X)$ *such that*

$$f = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \alpha_j^k \, \Pi\big(g_j^k, h_j^k\big) \tag{1.7}$$

in $H^1_\nu(X)$, *where the operator* Π *is defined as follows: for* $g \in L^p_{\lambda_1}(X)$ *and* $h \in L^{p'}_{\lambda_2'}(X)$,

$$\Pi(g, h) := gTh - hT^*g,$$

where T is a Calderón–Zygmund operator as in Definition 1.1 and satisfies the non-degenerate condition (1.6) and T^* is the conjugate of T in the sense that

$$\int_X Tf(x)g(x)d\mu(x) = \int_X f(x)T^*g(x)d\mu(x), \qquad f, g \in L^2(X).$$

Moreover, we have

$$\|f\|_{H^1_\nu(X)} \approx \inf\left\{\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\left|\alpha_j^k\right|\left\|g_j^k\right\|_{L^p_{\lambda_1}(X)}\left\|h_j^k\right\|_{L^{p'}_{\lambda_2'}(X)} : f = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\alpha_j^k\,\Pi\left(g_j^k, h_j^k\right)\right\}, \tag{1.8}$$

where the implicit constants are independent of f.

As applications, besides the classical Hilbert transform, Riesz transform and the Calderón–Zygmund operators with homogeneous kernels $\Omega(\frac{x}{|x|})\frac{1}{|x|^n}$ on $\mathbb{R}^n$ (studied in [21,32]), we use our main theorems to obtain the two weight commutator result of the following operators:

1. the Cauchy integral operator C_A along a Lipschitz curve $z := x + iA(x)$, $x \in (-\infty, \infty)$ and $A' \in L^\infty(\mathbb{R})$;
2. The Cauchy–Szegö projection operator on Heisenberg group $\mathbb{H}^n$;
3. The Szegö projection operator on a family of weakly pseudoconvex domains;
4. Riesz transforms associated to the sub-Laplacian on stratified Lie groups $\mathcal{G}$;
5. Riesz transforms associated to the Bessel operator Δ_λ on $\mathbb{R}_+$ for $\lambda > -1/2$;
6. Riesz transforms associated to the higher order Bessel operator $\Delta_{n,\lambda}$ on $\mathbb{R}^{n+1}_+$ for $\lambda > -1/2$.

The definitions of the above operators will be given in Sect. 7. We have the following result.

Theorem 1.6 *Let T be one of the operators listed above and let (X, d, μ) be the underlying space adapted to T. Suppose $1 < p < \infty$, $\lambda_1, \lambda_2 \in A_p$, $\nu := \lambda_1^{\frac{1}{p}}\lambda_2^{-\frac{1}{p}}$ and $m \in \mathbb{N}$. Suppose that $b \in L^1_{\mathrm{loc}}(X)$. Then, we have*

$$\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \approx \|T_b^m : L^p_{\lambda_1}(X) \to L^p_{\lambda_2}(X)\|.$$

Moreover, based on the result above for $m = 1$ and on the duality, the corresponding weighted Hardy space $H^1_\nu(X)$ has a weak factorisation as in (1.7).

To prove this theorem, the key step is to verify that all these operators listed above satisfy the conditions as in Definition 1.1 and the non-degenerate condition as in (1.6). We point out that such verification for Cauchy integral operator C_A is direct. The verifications of Cauchy–Szegö projection operator on Heisenberg group, the Szegö projection operator on a family of weakly pseudoconvex domains and the Riesz transforms associated with the sub-Laplacian on stratified Lie groups can be derived based on the results in [43, Chapter XII], [18] and [14], respectively. The verification for Riesz transforms associated with the Bessel operator Δ_λ on $\mathbb{R}_+$ for $\lambda > 0$ can be derived from the result in [39], while for $\lambda \in (-1/2, 0)$ is new here. The verification for Riesz transforms associated with higher order Bessel operator is totally new, especially the pointwise kernel lower bound of this Riesz transform.

We now address our result Theorem 1.6 with respect to the 6 examples above, respectively:

1. The unweighted result was obtained in [34] when $m = 1$, and the two weight result is new here for $m \geq 1$;
2. This result is new, even the unweighted version is unknown;
3. This result is new, even the unweighted version is unknown;
4. This result was obtained in [15] when $m = 1$ and is new here when $m > 1$.
5. The unweighted result was obtained in [16] when $\lambda > 0$ and $m = 1$, the two weight result is new here for $m \geq 1$ and for all $\lambda > -1/2$;
6. This result is new, even the unweighted version is unknown;

This paper is organised as follows. In Sect. 2 we recall the necessary preliminaries on spaces of homogeneous type. In Sect. 3, we first characterise the sparse system equivalently via the Λ-Carleson packing condition and the η-sparse condition, and then borrowing the idea from [32], we study the sparse operators and its domination of commutator on spaces of homogeneous type, and using this as a main tool, in Sect. 4 we obtain the upper bound of two weight commutator, i.e., Theorem 1.3. In Sect. 5 we provide the lower bound of two weight commutator, i.e., Theorem 1.4, by combining the ideas in [25,33]. In Sect. 6 we provide a study of weighted Hardy spaces and its duality on spaces of homogeneous type, and provide the proof of Theorem 1.5. In Sect. 7 we provide the applications where we address the new points in this paper. In the last section we also provide a new proof of the lower bound of two weight commutators in the product setting for little bmo space on spaces of homogeneous type. Note that in $\mathbb{R}^n \times \mathbb{R}^m$, this was first studied by [22] by using the Fourier transform for the Riesz transform kernel.

Throughout the paper, we denote by C and $\widetilde{C}$ *positive constants* which are independent of the main parameters, but they may vary from line to line. For every $p \in (1, \infty)$, we denote by p' the conjugate of p, i.e., $\frac{1}{p'} + \frac{1}{p} = 1$. If $f \leq Cg$ or $f \geq Cg$, we then write $f \lesssim g$ or $f \gtrsim g$; and if $f \lesssim g \lesssim f$, we write $f \approx g$.

2 Preliminaries on Spaces of Homogeneous Type

Let (X, d, μ) be a space of homogeneous type as mentioned in Sect. 1.

2.1 A System of Dyadic Cubes

In (X, d, μ), a countable family $\mathscr{D} := \cup_{k\in\mathbb{Z}}\mathscr{D}_k$, $\mathscr{D}_k := \{Q_\alpha^k : \alpha \in \mathscr{A}_k\}$, of Borel sets $Q_\alpha^k \subseteq X$ is called *a system of dyadic cubes with parameters* $\delta \in (0, 1)$ and $0 < a_1 \leq A_1 < \infty$ if it has the following properties:

$$
\begin{aligned}
&X = \bigcup_{\alpha\in\mathscr{A}_k} Q^k_\alpha \quad \text{(disjoint union) for all } k \in \mathbb{Z};\\
&\text{if } \ell \ge k, \text{ then either } Q^\ell_\beta \subseteq Q^k_\alpha \text{ or } Q^k_\alpha \cap Q^\ell_\beta = \emptyset;\\
&\text{for each } (k,\alpha) \text{ and each } \ell \le k, \text{ there exists a unique } \beta \text{ such that } Q^k_\alpha \subseteq Q^\ell_\beta;\\
&\text{for each } (k,\alpha) \text{ there exists at most } M (\text{a fixed geometric constant}) \beta \text{ such that}\\
&Q^{k+1}_\beta \subseteq Q^k_\alpha, \text{ and } Q^k_\alpha = \bigcup_{Q\in\mathscr{D}_{k+1},\, Q\subseteq Q^k_\alpha} Q;\\
&B(x^k_\alpha, a_1\delta^k) \subseteq Q^k_\alpha \subseteq B(x^k_\alpha, A_1\delta^k) =: B(Q^k_\alpha);\\
&\text{if } \ell \ge k \text{ and } Q^\ell_\beta \subseteq Q^k_\alpha, \text{ then } B(Q^\ell_\beta) \subseteq B(Q^k_\alpha).
\end{aligned}
\tag{2.1}
$$

The set Q^k_α is called a *dyadic cube of generation* k with centre point $x^k_\alpha \in Q^k_\alpha$ and sidelength δ^k.

From the properties of the dyadic system above and from the doubling measure, we can deduce that there exists a constant $C_{\mu,0}$ depending only on C_μ as in (1.2) and a_1, A_1 as above, such that for any Q^k_α and Q^{k+1}_β with $Q^{k+1}_\beta \subset Q^k_\alpha$,

$$\mu(Q^{k+1}_\beta) \le \mu(Q^k_\alpha) \le C_{\mu,0}\mu(Q^{k+1}_\beta). \tag{2.2}$$

We recall from [26] the following construction, which is a slight elaboration of seminal work by Christ [5], as well as Sawyer–Wheeden [42].

Theorem 2.1 *On (X, d, μ), there exists a system of dyadic cubes with parameters $0 < \delta \le (12A_0^3)^{-1}$ and $a_1 := (3A_0^2)^{-1}$, $A_1 := 2A_0$. The construction only depends on some fixed set of countably many centre points x^k_α, having the properties that $d(x^k_\alpha, x^k_\beta) \ge \delta^k$ with $\alpha \ne \beta$, $\min_\alpha d(x, x^k_\alpha) < \delta^k$ for all $x \in X$, and a certain partial order "$\le$" among their index pairs (k,α). In fact, this system can be constructed in such a way that*

$$\overline{Q}^k_\alpha = \overline{\{x^\ell_\beta : (\ell,\beta) \le (k,\alpha)\}}, \qquad \widetilde{Q}^k_\alpha := \operatorname{int} \overline{Q}^k_\alpha = \Big(\bigcup_{\gamma\ne\alpha} \overline{Q}^k_\gamma\Big)^c, \qquad \widetilde{Q}^k_\alpha \subseteq Q^k_\alpha \subseteq \overline{Q}^k_\alpha,$$

where Q^k_α are obtained from the closed sets $\overline{Q}^k_\alpha$ and the open sets $\widetilde{Q}^k_\alpha$ by finitely many set operations.

We also recall the following remark from [28, Section 2.3]. The construction of dyadic cubes requires their centre points and an associated partial order be fixed *a priori*. However, if either the centre points or the partial order is not given, their existence already follows from the assumptions; any given system of points and partial order can be used as a starting point. Moreover, if we are allowed to choose the centre points for the cubes, the collection can be chosen to satisfy the additional property that a fixed point becomes a centre point at *all levels*:

$$
\begin{aligned}
&\text{given a fixed point } x_0 \in X, \text{ for every } k \in \mathbb{Z}, \text{ there exists } \alpha \text{ such that}\\
&x_0 = x^k_\alpha, \text{ the centre point of } Q^k_\alpha \in \mathscr{D}_k.
\end{aligned}
\tag{2.3}
$$

2.2 Adjacent Systems of Dyadic Cubes

On (X, d, μ), a finite collection $\{\mathscr{D}^t : t = 1, 2, \ldots, \mathcal{T}\}$ of the dyadic families is called a *collection of adjacent systems of dyadic cubes with parameters* $\delta \in (0, 1)$, $0 < a_1 \leq A_1 < \infty$ and $1 \leq C_{adj} < \infty$ if it has the following properties: individually, each $\mathscr{D}^t$ is a system of dyadic cubes with parameters $\delta \in (0, 1)$ and $0 < a_1 \leq A_1 < \infty$; collectively, for each ball $B(x, r) \subseteq X$ with $\delta^{k+3} < r \leq \delta^{k+2}$, $k \in \mathbb{Z}$, there exist $t \in \{1, 2, \ldots, \mathcal{T}\}$ and $Q \in \mathscr{D}^t$ of generation k and with centre point ${}^t x_\alpha^k$ such that $d(x, {}^t x_\alpha^k) < 2A_0\delta^k$ and

$$B(x, r) \subseteq Q \subseteq B(x, C_{adj} r). \tag{2.4}$$

We recall from [26] the following construction.

Theorem 2.2 *Let (X, d, μ) be a space of homogeneous type. Then, there exists a collection $\{\mathscr{D}^t : t = 1, 2, \ldots, \mathcal{T}\}$ of adjacent systems of dyadic cubes with parameters $\delta \in (0, (96A_0^6)^{-1})$, $a_1 := (12A_0^4)^{-1}$, $A_1 := 4A_0^2$ and $C := 8A_0^3\delta^{-3}$. The centre points ${}^t x_\alpha^k$ of the cubes $Q \in \mathscr{D}_k^t$ have, for each $t \in \{1, 2, \ldots, \mathcal{T}\}$, the two properties*

$$d({}^t x_\alpha^k, {}^t x_\beta^k) \geq (4A_0^2)^{-1}\delta^k \quad (\alpha \neq \beta), \qquad \min_\alpha d(x, {}^t x_\alpha^k) < 2A_0\delta^k \quad \textit{for all } x \in X.$$

Moreover, these adjacent systems can be constructed in such a way that each $\mathscr{D}^t$ satisfies the distinguished centre point property (2.3).

We recall from [28, Remark 2.8] that the number $\mathcal{T}$ of the adjacent systems of dyadic cubes as in the theorem above satisfies the estimate

$$\mathcal{T} = \mathcal{T}(A_0, \widetilde{A}_1, \delta) \leq \widetilde{A}_1^6 (A_0^4/\delta)^{\log_2 \widetilde{A}_1},$$

where $\widetilde{A}_1$ is the geometrically doubling constant, see [28, Section 2].

2.3 An Explicit Haar Basis on Spaces of Homogeneous Type

Next, we recall the explicit construction in [28] of a Haar basis $\{h_Q^\epsilon : Q \in \mathscr{D}, \epsilon = 1, \ldots, M_Q - 1\}$ for $L^p(X, \mu)$, $1 < p < \infty$, associated to the dyadic cubes $Q \in \mathscr{D}$ as follows. Here, $M_Q := \#\mathcal{H}(Q) = \#\{R \in \mathscr{D}_{k+1} : R \subseteq Q\}$ denotes the number of dyadic sub-cubes ("children") the cube $Q \in \mathscr{D}_k$ has; namely $\mathcal{H}(Q)$ is the collection of dyadic children of Q.

Theorem 2.3 ([28]) *Let (X, d, μ) be a space of homogeneous type and suppose μ is a positive Borel measure on X with the property that $\mu(B) < \infty$ for all balls $B \subseteq X$. For $1 < p < \infty$, for each $f \in L^p(X, \mu)$, we have*

$$f(x) = \sum_{Q \in \mathscr{D}} \sum_{\epsilon=1}^{M_Q - 1} \langle f, h_Q^\epsilon \rangle h_Q^\epsilon(x),$$

where the sum converges (unconditionally) both in the $L^p(X,\mu)$-norm and pointwise μ-almost everywhere.

The following theorem collects several basic properties of the functions h_Q^ϵ.

Theorem 2.4 ([28]) *The Haar functions h_Q^ϵ, $Q\in\mathscr{D}$, $\epsilon=1,\dots,M_Q-1$, have the following properties:*

(i) *h_Q^ϵ is a simple Borel-measurable real function on X;*
(ii) *h_Q^ϵ is supported on Q;*
(iii) *h_Q^ϵ is constant on each $R\in\mathcal{H}(Q)$;*
(iv) *$\int h_Q^\epsilon\,d\mu=0$ (cancellation);*
(v) *$\langle h_Q^\epsilon,h_Q^{\epsilon'}\rangle=0$ for $\epsilon\neq\epsilon'$, $\epsilon,\epsilon'\in\{1,\dots,M_Q-1\}$;*
(vi) *the collection $\left\{\mu(Q)^{-1/2}1_Q\right\}\cup\{h_Q^\epsilon:\epsilon=1,\dots,M_Q-1\}$ is an orthogonal basis for the vector space $V(Q)$ of all functions on Q that are constant on each sub-cube $R\in\mathcal{H}(Q)$;*
(vii) *if $h_Q^\epsilon\not\equiv 0$ then $\|h_Q^\epsilon\|_{L^p(X,\mu)}\approx\mu(Q_\epsilon)^{\frac{1}{p}-\frac{1}{2}}$ for $1\le p\le\infty$;*
(viii) *$\|h_Q^\epsilon\|_{L^1(X,\mu)}\cdot\|h_Q^\epsilon\|_{L^\infty(X,\mu)}\approx 1$.*

As stated in [28], we also have $h_Q^0:=\mu(Q)^{-1/2}1_Q$ which is a non-cancellative Haar function. Moreover, the martingale associated with the Haar functions are as follows: for $Q\in\mathscr{D}_k$,

$$\mathbb{E}_Q f=\langle f,h_Q^0\rangle h_Q^0\quad\text{and}\quad\mathbb{D}_Q f=\sum_{\epsilon=1}^{M_Q-1}\mathbb{D}_Q^\epsilon f,$$

where $\mathbb{D}_Q^\epsilon=\langle f,h_Q^\epsilon\rangle h_Q^\epsilon$ is the martingale operator associated with the ϵth subcube of Q. Also we have

$$\mathbb{E}_k f=\sum_{Q\in\mathscr{D}_k}\mathbb{E}_Q f\quad\text{and}\quad\mathbb{D}_k f=\mathbb{E}_{k+1}f-\mathbb{E}_k f.$$

Hence, based on the construction of Haar system $\{h_Q^\epsilon\}$ in [28] we obtain that for each $R\in\mathscr{D}$,

$$\sum_{Q:\,R\subset Q}\sum_{\epsilon=1}^{M_Q-1}\langle f,h_Q^\epsilon\rangle h_Q^\epsilon h_R^\eta=\sum_{Q:\,R\subset Q}\mathbb{D}_Q f\cdot h_R^\eta=\mathbb{E}_R f\cdot h_R^\eta=\langle f,h_R^0\rangle h_R^0 h_R^\eta.$$

2.4 Muckenhoupt A_p Weights

Definition 2.5 Let $w(x)$ be a nonnegative locally integrable function on X. For $1<p<\infty$, we say w is an A_p *weight*, written $w\in A_p$, if

$$[w]_{A_p}:=\sup_B\left(⨍_B w\right)\left(⨍_B\left(\frac{1}{w}\right)^{1/(p-1)}\right)^{p-1}<\infty.$$

Here, the supremum is taken over all balls $B \subset X$ and $⨍_B w := \frac{1}{\mu(B)} \int_B w(x)d\mu(x)$. The quantity $[w]_{A_p}$ is called the A_p *constant of* w. For $p = 1$, we say w is an A_1 *weight*, written $w \in A_1$, if $M(w)(x) \le w(x)$ for μ-almost every $x \in X$, and let $A_\infty := \cup_{1\le p<\infty} A_p$ and we have $[w]_{A_\infty} := \sup_B \left(⨍_B w\right) \exp\left(⨍_B \log\left(\frac{1}{w}\right)\right) < \infty$.

Next, we note that for $w \in A_p$ the measure $w(x)d\mu(x)$ is a doubling measure on X. To be more precise, we have that for all $\lambda > 1$ and all balls $B \subset X$,

$$w(\lambda B) \le \lambda^{np}[w]_{A_p} w(B),$$

where n is the upper dimension of the measure μ, as in (1.3).

We also point out that for $w \in A_\infty$, there exists $\gamma > 0$ such that for every ball B,

$$\mu\left(\left\{x \in B : \ w(x) \ge \gamma ⨍_B w\right\}\right) \ge \frac{1}{2}\mu(B).$$

And this implies that for every ball B and for all $\delta \in (0, 1)$,

$$⨍_B w \le C\left(⨍_B w^\delta\right)^{1/\delta}; \tag{2.5}$$

see also [33].

3 Sparse Operators and Domination of Commutators on Spaces of Homogeneous Type

Let $\mathcal{D}$ be a system of dyadic cubes on X as in Sect. 2.1. As in the Euclidean setting, we have two competing versions of sparsity for a collection of sets, one geometric and the other a Carleson measure condition.

Definition 3.1 Given $0 < \eta < 1$, a collection $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes is said to be η-sparse provided that for every $Q \in \mathcal{S}$, there is a measurable subset $E_Q \subset Q$ such that $\mu(E_Q) \ge \eta\mu(Q)$ and the sets $\{E_Q\}_{Q\in\mathcal{S}}$ have only finite overlap. That is, there exists a constant $c \ge 1$ such that $\sum_Q \chi_{E_Q}(x) \le c$ for all $x \in X$.

The reason for the extra constant c in the above, is that for our arguments in Theorem 3.7, to control the commutator, we need to allow the sets E_Q to have finite overlap. If the sets E_Q were exactly disjoint then one could take $c = 1$ in the above and the statement would be cleaner and more in line with that in [31].

We note that in [29], Karagulyan introduced a more general family of sets, called ball-basis, and then defined the sparse family based on these ball-basis, using the geometric version of sparsity, which is similar to our Definition 3.1. However, the ball-basis in [29] is pairwise disjoint, which does not seem well fit for our proof for upper bound of the commutator.

We now provide the Carleson measure condition for the sparse family, which was not studied in [29].

Definition 3.2 Given $\Lambda > 1$, a collection $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes is said to be Λ-Carleson if for every cube $Q \in \mathcal{D}$,

$$\sum_{P \in \mathcal{S}, P \subseteq Q} \mu(P) \leq \Lambda \mu(Q).$$

We first show that the above two definitions are equivalent in a space of homogeneous type. The proof closely follows the original idea in [31] with modifications, especially on the replacement of using of translation in [31].

Theorem 3.3 *Given $0 < \eta < 1$ and a collection $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes, the following statements hold:*

- *If $\mathcal{S}$ is η-sparse, then $\mathcal{S}$ is $\frac{c}{\eta}$-Carleson, where c is the constant in Definition 3.1;*
- *If $\mathcal{S}$ is $\frac{1}{\eta}$-Carleson, then $\mathcal{S}$ is η-sparse.*

Proof Note that if a collection $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes is η-sparse, that is for every $Q \in \mathcal{S}$, there is a measurable subset $E_Q \subset Q$ such that $\mu(E_Q) \geq \eta\mu(Q)$ and the sets $\{E_Q\}_{Q \in \mathcal{S}}$ have only finite overlap, we will have that $\mathcal{S}$ is $c\eta^{-1}$-Carleson according to Definition 3.2 (following from the standard computation).

Thus, it suffices to show that for $\Lambda > 1$, if a collection $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes is Λ-Carleson, then it is Λ^{-1}-sparse. To see this, we first point out that if the collection $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes $\{Q\}$ has a bottom layer $\mathscr{D}_K$ for some fixed integer K, then it is direct to construct the set E_Q. We begin with considering all dyadic cubes $\{Q\} \subset \mathcal{S} \cap \mathscr{D}_K$ and choose any measurable set $E_Q \subset Q$ of measure $\Lambda^{-1}\mu(Q)$ for them. We now just repeat this choice for each dyadic cube in upper layers one by one. To be more specific, for each $Q \in \mathcal{S} \cap \mathscr{D}_k$ with $k \leq K$, choose a set

$$E_Q \subset Q \Big\backslash \bigcup_{R \in \mathcal{S}, R \subsetneq Q} E_R$$

such that $\mu(E_Q) = \Lambda^{-1}\mu(Q)$. We now show that such choice of E_Q is possible. In fact, note that for every $Q \in \mathcal{S}$, we have

$$\mu\Big(\bigcup_{R \in \mathcal{S}, R \subsetneq Q} E_R\Big) \leq \Lambda^{-1} \sum_{R \in \mathcal{S}, R \subsetneq Q} \mu(R) \leq \Lambda^{-1}(\Lambda - 1)\mu(Q) = (1 - \Lambda^{-1})\mu(Q),$$

where the last inequality follows from the Λ-Carleson condition and from the fact that $R \subsetneq Q$. This shows that

$$\mu\Big(Q \Big\backslash \bigcup_{R \in \mathcal{S}, R \subsetneq Q} E_R\Big) \geq \mu(Q) - (1 - \Lambda^{-1})\mu(Q) = \Lambda^{-1}\mu(Q)$$

and so the choice of the top set E_Q is always possible.

Next, we consider the case that there is no fixed bottom layer. We run the above construction with a particular choice for each $K = 0, 1, 2, \ldots$ and then pass to the limit. To begin with, fix $K \geq 0$. For each $Q \in \mathcal{S} \cap (\cup_{k \leq K} \mathscr{D}_k)$, we define the sets $\widehat{E}_Q^{(K)}$ inductively as follows.

First, for each $Q \in \mathcal{S} \cap \mathscr{D}_k$ with $k \leq K$, we consider the auxiliary set

$$\mathcal{Q}(t, Q) := B(x_Q, t\delta^k) \cap Q, \quad t \in (0, A_1),$$

where x_Q is the centre point of Q and A_1, δ are the constants as introduced in Sect. 2.1. From property (2.1), it is clear that when $0 < t < a_1$, then $B(x_Q, t\delta^k) \subset Q$ and when $t > A_1$, then $Q \subset B(x_Q, t\delta^k)$; moreover, we have $\mu(B(x_Q, t\delta^k)) \to 0$ as $t \to 0^+$.

Now for $Q \in \mathcal{S} \cap \mathscr{D}_K$, from the above observations together with the continuity and monotonicity of the function $t \mapsto \mathcal{Q}(t, Q) = \mu(B(x_Q, t\delta^K)) \cap Q$, we conclude that there must be some $t_{\Lambda,K,K} \in (0, A_1)$ such that $\mu(B(x_Q, t_{\Lambda,K,K}\delta^K) \cap Q) = \Lambda^{-1}\mu(Q)$. Here and in what follows, we use the triple (Λ, k, K) for the subscript of t, where Λ denotes that the value of such t depends on Λ, k denotes that Q is in the layer k and the last K denotes that we start at the layer K. We set

$$\widehat{E}_Q^{(K)} := \mathcal{Q}(t_{\Lambda,K,K}, Q) = B(x_Q, t_{\Lambda,K,K}\delta^K) \cap Q.$$

Suppose now $\widehat{E}_R^{(K)}$ are already defined for every $R \in \mathcal{S} \cap (\cup_{k+1 \leq i \leq K} \mathscr{D}_i)$. We now define $\widehat{E}_Q^{(K)}$ for $Q \in \mathcal{S} \cap \mathscr{D}_k$ in the following manner. We set

$$\widehat{E}_Q^{(K)} := \mathcal{Q}(t_{\Lambda,k,K}, Q) \bigcup F_Q^{(K)},$$

where

$$F_Q^{(K)} := \bigcup_{R \in \mathcal{S} \cap (\cup_{k+1 \leq i \leq K} \mathscr{D}_i), R \subsetneq Q} \widehat{E}_R^{(K)}$$

and $t_{\Lambda,k,K} \in (0, A_1)$ is chosen such that the set

$$E_Q^{(K)} := \mathcal{Q}(t_{\Lambda,k,K}, Q) \backslash F_Q^{(K)}$$

satisfies $\mu(E_Q^{(K)}) = \Lambda^{-1}\mu(Q)$.

Now, we claim that $\widehat{E}_Q^{(K)} \subset \widehat{E}_Q^{(K+1)}$ for every $Q \in \mathcal{S} \cap (\cup_{k \leq K} \mathscr{D}_k)$. To see this, we note that for each $Q \in \mathcal{S} \cap \mathscr{D}_K$, $\widehat{E}_Q^{(K)}$ is just the set $\mathcal{Q}(t_{\Lambda,K,K}, Q)$. On the other hand, $\widehat{E}_Q^{(K+1)}$ contains the set $\mathcal{Q}(t_{\Lambda,K,K+1}, Q)$ which has the same centre point as $\mathcal{Q}(t_{\Lambda,K,K}, Q)$, but with $t_{\Lambda,K,K+1} \geq t_{\Lambda,K,K}$ since

$$\mu\left(\mathcal{Q}(t_{\Lambda,K,K+1}, Q) \backslash \bigcup_{R \in \mathcal{S} \cap \mathscr{D}_{K+1}, R \subsetneq Q} \widehat{E}_R^{(K+1)}\right) = \Lambda^{-1}\mu(Q) = \mu(\mathcal{Q}(t_{\Lambda,K,K}, Q)).$$

Hence, we see that for each $Q \in \mathcal{S} \cap \mathscr{D}_K$, we have $\widehat{E}_Q^{(K)} \subseteq \widehat{E}_Q^{(K+1)}$. Then, we proceed via backward induction. Assume that $\widehat{E}_Q^{(K)} \subseteq \widehat{E}_Q^{(K+1)}$ for every $Q \in \mathcal{S} \cap (\cup_{k<i\le K} \mathscr{D}_i)$. Take any $Q \in \mathcal{S} \cap \mathscr{D}_k$. Then, the inductive hypothesis implies that $F_Q^{(K)} \subseteq F_Q^{(K+1)}$. Let $Q(t_{\Lambda,k,K}, Q)$ be the set added to $F_Q^{(K)}$ when constructing $\widehat{E}_Q^{(K)}$. Then, we have

$$\mu\Big(Q(t_{\Lambda,k,K}, Q)\backslash F_Q^{(K+1)}\Big) \le \mu\Big(Q(t_{\Lambda,k,K}, Q)\backslash F_Q^{(K)}\Big) = \Lambda^{-1}\mu(Q),$$

which implies that $t_{\Lambda,k,K+1} \ge t_{\Lambda,k,K}$. Thus, we have $Q(t_{\Lambda,k,K}, Q) \subset Q(t_{\Lambda,k,K+1}, Q)$, which yields $\widehat{E}_Q^{(K)} \subseteq \widehat{E}_Q^{(K+1)}$, and hence the claim follows.

Now for $Q \in \mathcal{S} \cap \mathscr{D}_k$, we define

$$\widehat{E}_Q := \lim_{K\to\infty} \widehat{E}_Q^{(K)},$$

which, by using the claim above, equals

$$\bigcup_{K=k}^{\infty} \widehat{E}_Q^{(K)} \subset Q.$$

Moreover, for each K we have

$$\mu(E_Q^{(K)}) = \mu\big(\widehat{E}_Q^{(K)}\backslash F_Q^{(K)}\big) = \Lambda^{-1}\mu(Q).$$

Note that the sets $F_Q^{(K)}$ also form an increasing sequence (with respect to K), so for each $Q \in \mathcal{S}$, the limit set

$$E_Q := \lim_{K\to\infty} E_Q^{(K)} = \widehat{E}_Q\backslash\big(\lim_{K\to\infty} F_Q^{(K)}\big) = \widehat{E}_Q\backslash\Big(\bigcup_{R\in\mathcal{S}, R\subsetneq Q} \widehat{E}_R\Big)$$

exists, and is contained in Q and has the required measure. Moreover, all E_Q are disjoint. The proof of Theorem 3.3 is complete. □

We now recall the well-known definition for sparse operators.

Definition 3.4 Given $0 < \eta < 1$ and an η-sparse family $\mathcal{S} \subset \mathcal{D}$ of dyadic cubes. The sparse operator $\mathcal{A}_\mathcal{S}$ is defined by

$$\mathcal{A}_\mathcal{S} f(x) := \sum_{Q\in\mathcal{S}} f_Q \chi_Q(x).$$

Following the proof of [37, Theorem 3.1], we obtain that

$$\|\mathcal{A}_\mathcal{S} f\|_{L^p_w(X)} \le C_{\eta,n,p}[w]_{A_p}^{\max\{1,\frac{1}{p-1}\}} \|f\|_{L^p_w(X)}, \quad 1 < p < \infty.$$

Denote by $\Omega(b, B)$ the standard mean oscillation

$$\Omega(b, B) := \frac{1}{\mu(B)} \int_B |b(x) - b_B| d\mu(x). \tag{3.1}$$

Lemma 3.5 *Given $0 < \gamma < 1$. Let $\mathcal{D}$ be a dyadic system in X and let $\mathcal{S} \subset \mathcal{D}$ be a γ-sparse family. Assume that $b \in L^1_{loc}(X)$. Then there exists a $\frac{\gamma}{2(\gamma+1)}$-sparse family $\tilde{\mathcal{S}} \subset \mathcal{D}$ such that $\mathcal{S} \subset \tilde{\mathcal{S}}$ and for every cube $Q \in \tilde{\mathcal{S}}$,*

$$|b(x) - b_Q| \leq C \sum_{R \in \tilde{\mathcal{S}}, R \subset Q} \Omega(b, R) \chi_R(x) \tag{3.2}$$

for μ-almost every $x \in Q$.

Proof Fix a dyadic cube $Q \in \mathcal{D}$. We now show that there exists a family of pairwise disjoint cubes $\{P_j\} \subset \mathcal{D}(Q)$ such that $\sum_j \mu(P_j) \leq \frac{1}{2}\mu(Q)$ and for μ-almost every $x \in Q$,

$$|b(x) - b_Q| \leq C \cdot C_{\mu,0} \Omega(b, Q) + \sum_j |b(x) - b_{P_j}| \chi_{P_j}(x). \tag{3.3}$$

Let M^d_Q be the standard dyadic local maximal operator restricted to $\mathcal{D}(Q)$ and $C_{M^d_Q}$ be the weak type (1, 1)-norm of M^d_Q. Then one can choose a constant C depending on $C_{M^d_Q}$ such that the set $E := \{x \in Q : M^d_Q(b - b_Q)(x) > 4C_{\mu,0} \cdot C \cdot \Omega(b, Q)\}$ satisfies that $\mu(E) \leq \frac{1}{4C_{\mu,0}}\mu(Q)$, where $C_{\mu,0}$ is the constant as in (2.2).

If $\mu(E) = 0$, then (3.3) holds trivially with the empty family $\{P_j\}_j$. If $\mu(E) > 0$, then we now apply the Calderón–Zygmund decomposition to the function $h(x) := \chi_E(x)$ on Q at height $\lambda := \frac{1}{2C_{\mu,0}}$ as follows: we begin by considering the descendants of Q in $\mathcal{D}(Q)$ since

$$\int_Q |h(x)| d\mu(x) < \lambda\mu(Q).$$

Let $\{Q^{(1)}_j\} \subset \mathcal{D}(Q)$ be the children of Q. If

$$\int_{Q^{(1)}_j} |h(x)| d\mu(x) > \lambda\mu(Q^{(1)}_j) \tag{3.4}$$

then we select it as our candidate cube. If

$$\int_{Q^{(1)}_j} |h(x)| d\mu(x) \leq \lambda\mu(Q^{(1)}_j)$$

then we keep looking at the children of $Q_j^{(1)}$ in $\mathcal{D}(Q)$ and then repeat the above selection criteria and we will stop only when we find some descendant of $Q_j^{(1)}$ in $\mathcal{D}(Q)$ such that it meets the criteria (3.4).

Then, it is direct to see that this produces pairwise disjoint cubes $\{P_j\} \subset \mathcal{D}(Q)$ such that

$$\frac{1}{2C_{\mu,0}}\mu(P_j) < \mu(P_j \cap E) \leq \frac{1}{2}\mu(P_j)$$

and $\mu(E \setminus \cup_j P_j) = 0$. It follows that $\sum_j \mu(P_j) \leq \frac{1}{2}\mu(Q)$ and $P_j \cap E^c \neq \emptyset$.

Therefore, we get

$$|b_{P_j} - b_Q| \leq \frac{1}{\mu(P_j)} \int_{P_j} |b(x) - b_Q| d\mu(x) \leq 4C_{\mu,0} \cdot C \cdot \Omega(b, Q) \tag{3.5}$$

and for μ-almost every $x \in Q$, $|b(x) - b_Q|\chi_{Q\setminus\cup_j P_j} \leq 4C_{\mu,0} \cdot C\Omega(b, Q)$.

Then, we have

$$\begin{aligned} |b(x) - b_Q|\chi_Q(x) &\leq |b(x) - b_Q|\chi_{Q\setminus\cup_j P_j}(x) \\ &\quad + \sum_j |b_{P_j} - b_Q|\chi_{P_j}(x) + \sum_j |b(x) - b_{P_j}|\chi_{P_j}(x) \\ &\leq 4C_{\mu,0} \cdot C\Omega(b, Q) + \sum_j |b(x) - b_{P_j}|\chi_{P_j}(x), \end{aligned}$$

which gives (3.3).

We observe that if $P_j \subset R$, where $R \in \mathcal{D}(Q)$, then $R \cap E^c \neq \emptyset$. Hence P_j in (3.5) can be replaced by R, namely, we have $|b_R - b_Q| \leq 4C_{\mu,0} \cdot C\Omega(b, Q)$. Therefore, if $\cup_j P_j \subset \cup_i R_i$, where $R_i \in \mathcal{D}(Q)$, and the cubes $\{R_i\}$ are pairwise disjoint, then we have

$$|b(x) - b_Q| \leq 4C_{\mu,0} \cdot C\Omega(b, Q) + \sum_i |b(x) - b_{R_i}|\chi_{R_i}(x). \tag{3.6}$$

Iterating (3.3), from the selection of $\{P_j\}$ and from Definition 3.2, we obtain that there exists a $\frac{1}{2}$-sparse family $\mathcal{F}(Q) \subset \mathcal{D}(Q)$ such that for μ-almost every $x \in Q$,

$$|b(x) - b_Q|\chi_Q(x) \leq 4C_{\mu,0} \cdot C \sum_{P\in\mathcal{F}(Q)} \Omega(b, P)\chi_P(x).$$

Now for each $\mathcal{F}(Q)$, let $\tilde{\mathcal{F}}(Q)$ be the family that consists of all cubes $\{P\} \subset \mathcal{F}(Q)$ that are not contained in any cube $R \in \mathcal{S}$ with $R \subsetneq Q$. Then, we define

$$\tilde{\mathcal{S}} := \bigcup_{Q\in\mathcal{S}} \tilde{\mathcal{F}}(Q).$$

It is clear, by construction, that the augmented family $\tilde{\mathcal{S}}$ contains the original family $\mathcal{S}$. Furthermore, if $\mathcal{S}$ and each $\mathcal{F}(Q)$ are sparse families, then the augmented family $\tilde{\mathcal{S}}$ is also sparse.

To be specific, we have that if $\mathcal{S} \subset \mathcal{D}$ is an γ-sparse family then the augmented family $\tilde{\mathcal{S}}$ built upon $\frac{1}{2}$-sparse family $\mathcal{F}(Q)$, $Q \in \mathcal{S}$, is an $\frac{\gamma}{2(\gamma+1)}$-sparse family.

We now show (3.2). Take an arbitrary cube $Q \in \mathcal{S}$. Let P_j be the cubes appearing in (3.3). Denote by $\mathcal{M}(Q)$ the family of the maximal pairwise disjoint cubes from $\tilde{\mathcal{S}}$ which are strictly contained in Q. Then by the augmentation process, $\cup_j P_j \subset \cup_{P\in\mathcal{M}(Q)} P$. Therefore, by (3.6), we have

$$|b(x) - b_Q|\chi_Q(x) \leq 4C_{\mu,0} \cdot C\Omega(b, Q) + \sum_{P\in\mathcal{M}(Q)} |b(x) - b_P|\chi_P(x). \tag{3.7}$$

Now split $\tilde{\mathcal{S}}(Q) := \{P \in \mathcal{S} : \ P \subset Q\}$ into the layers $\tilde{\mathcal{S}}(Q) = \cup_{k=0}^{\infty} \mathcal{M}_k$, where $\mathcal{M}_0 := \{Q\}$, $\mathcal{M}_1 := \mathcal{M}(Q)$ and $\mathcal{M}_k$ is the family of the maximal elements of $\mathcal{M}_{k-1}$. Iterating (3.7) k times, we get that

$$|b(x) - b_Q|\chi_Q(x) \leq 4C_{\mu,0} \cdot C \sum_{P\in\tilde{\mathcal{S}}(Q)} \Omega(b, P)\chi_P(x) + \sum_{P\in\mathcal{M}_k} |b(x) - b_P|\chi_P(x). \tag{3.8}$$

Now, we observe that since $\tilde{\mathcal{S}}$ is $\frac{\gamma}{2(\gamma+1)}$-sparse,

$$\sum_{P\in\mathcal{M}_k} \mu(P) \leq \frac{1}{k+1} \sum_{i=0}^{k} \sum_{P\in\mathcal{M}_i} \mu(P) \leq \frac{1}{k+1} \sum_{P\in\tilde{\mathcal{S}}(Q)} \mu(P) \leq \frac{2(\gamma+1)}{\gamma(k+1)} \mu(Q).$$

By letting $k \to \infty$ in (3.8), we obtain (3.2). □

Let T be a Calderón–Zygmund operator as in Definition 1.1. We now have the maximal truncated operator T_* defined by

$$T_* f(x) := \sup_{\epsilon>0} \left| \int_{d(x,y)>\epsilon} K(x, y) f(y) d\mu(y) \right|.$$

We recall the standard Hardy–Littlewood maximal function $\mathcal{M}f(x)$ on X, defined as

$$\mathcal{M}f(x) := \sup_{B\ni x} \frac{1}{\mu(B)} \int_B |f(y)|\, d\mu(y),$$

where the supremum is taken over all balls $B \subset X$. We now have the grand maximal truncated operator $\mathcal{M}_T$ defined by

$$\mathcal{M}_T f(x) := \sup_{B\ni x} \operatorname*{ess\,sup}_{\xi\in B} \left| T\big(f\chi_{X\setminus C_{\tilde{j}_0} B}\big)(\xi) \right|,$$

where the supremum is taken over all balls $B \subset X$ containing x, $\widetilde{j}_0$ is the smallest integer such that

$$2^{\widetilde{j}_0} > \max\{3A_0, 2A_0 \cdot C_{adj}\} \tag{3.9}$$

and $C_{\widetilde{j}_0} := 2^{\widetilde{j}_0+2}A_0$, where C_{adj} is an absolute constant as mentioned in Sect. 2.2. Given a ball $B_0 \subset X$, for $x \in B_0$ we define a local grand maximal truncated operator $\mathcal{M}_{T,B_0}$ as follows:

$$\mathcal{M}_{T,B_0} f(x) := \sup_{B \ni x,\, B \subset B_0} \operatorname*{ess\,sup}_{\xi \in B} \left| T\left(f \chi_{C_{\widetilde{j}_0} B_0 \setminus C_{\widetilde{j}_0} B}\right)(\xi)\right|.$$

Then, we first claim that the following lemma holds.

Lemma 3.6 *The following pointwise estimates hold:*

(i) *for μ-almost every $x \in B_0$,*

$$|T(f\chi_{C_{\widetilde{j}_0} B_0})(x)| \le C\|T\|_{L^1 \to L^{1,\infty}} |f(x)| + \mathcal{M}_{T,B_0} f(x).$$

(ii) *for all $x \in X$, $\mathcal{M}_T f(x) \le C\mathcal{M} f(x) + T_* f(x)$.*

Proof The result in the Euclidean setting is from [30, Lemma 3.2]. Here, we can adapt the proof in [30] to our setting of spaces of homogeneous type. □

Next, we have the sparse domination for the higher order commutator.

Theorem 3.7 *Let T be the Calderón–Zygmund operator as in Definition 1.1 and let $b \in L^1_{loc}(X)$. For every $f \in L^\infty(X)$ with bounded support, there exist $\mathcal{T}$ dyadic systems $\mathcal{D}^t$, $t = 1, 2, \ldots, \mathcal{T}$ and η-sparse families $\mathcal{S}_t \subset \mathcal{D}^t$ such that for μ-almost every $x \in X$,*

$$\begin{aligned} |T_b^m(f)(x)| \ &\le C \sum_{t=1}^{\mathcal{T}} \sum_{k=0}^{m} C_m^k \sum_{Q \in \mathcal{S}_t} |b(x) - b_Q|^{m-k} \\ &\quad \times \left(\frac{1}{\mu(Q)} \int_Q |b(z) - b_Q|^k |f(z)| d\mu(z) \right) \chi_Q(x), \end{aligned} \tag{3.10}$$

where $C_m^k := \frac{m!}{(m-k)! \cdot k!}$.

Proof We follow the idea as in [33] for the domination, and adapt it to our setting of space of homogeneous type.

Suppose f is supported in a ball $B_0 := B(x_0, r) \subset X$. We now consider a decomposition of X with respect to this ball B_0. We define the annuli $U_j := 2^{j+1}B_0 \backslash 2^j B_0$, $j \ge 0$ and we choose j_0 to be the smallest integer such that

$$j_0 > \widetilde{j}_0 \quad \text{and} \quad 2^{j_0} > 4A_0. \tag{3.11}$$

Next, for each U_j, we choose the balls

$$\{\widetilde{B}_{j,\ell}\}_{\ell=1}^{L_j} \tag{3.12}$$

centred in U_j and with radius $2^{j-\widetilde{j}_0}r$ to cover U_j. From the geometric doubling property [9, p. 67], it is direct to see that

$$\sup_j L_j \le C_{A_0,\mu,\widetilde{j}_0}, \tag{3.13}$$

where $C_{A_0,\mu,\widetilde{j}_0}$ is an absolute constant depending only on A_0, $\widetilde{j}_0$ and C_μ.

We now first study the properties of these $\widetilde{B}_{j,\ell}$. Denote $\widetilde{B}_{j,\ell} := B(x_{j,\ell}, 2^{j-\widetilde{j}_0}r)$, where $\widetilde{j}_0$ is defined as in (3.9). Then we have $C_{adj}\widetilde{B}_{j,\ell} := B(x_{j,\ell}, C_{adj}2^{j-\widetilde{j}_0}r)$, where C_{adj} is an absolute constant as mentioned in Sect. 2.2. We claim that

$$C_{adj}\widetilde{B}_{j,\ell} \cap U_{j+j_0} = \emptyset, \quad \forall j \ge 0 \quad \text{and} \quad \forall \ell = 1, 2, \ldots, L_j; \tag{3.14}$$

and that

$$C_{adj}\widetilde{B}_{j,\ell} \cap U_{j-j_0} = \emptyset, \quad \forall j \ge j_0 \quad \text{and} \quad \forall \ell = 1, 2, \ldots, L_j. \tag{3.15}$$

Assume (3.14) and (3.15) at the moment. Now combining the properties as in (3.14) and (3.15), we see that each $C_{adj}\widetilde{B}_{j,\ell}$ only intersects with at most $2j_0 + 1$ annuli U_js. Moreover, for every j and ℓ, $C_{\widetilde{j}_0}\widetilde{B}_{j,\ell}$ covers B_0.

Now, for the given ball B_0 as above, we point out that from (2.4) we have that there exist an integer $t_0 \in \{1, 2, \ldots, \mathcal{T}\}$ and $Q_0 \in \mathscr{D}^{t_0}$ such that $B_0 \subseteq Q_0 \subseteq C_{adj}B_0$. Moreover, for this Q_0, as in (2.1) we use $B(Q_0)$ to denote the ball that contains Q_0 and has measure comparable to Q_0. Then it is easy to see that $B(Q_0)$ covers B_0 and $\mu(B(Q_0)) \lesssim \mu(B_0)$, where the implicit constant depends only on C_{adj}, C_μ and A_1 as in (2.1).

We show that there exists a $\frac{1}{2}$-sparse family $\mathcal{F}^{t_0} \subset \mathcal{D}^{t_0}(Q_0)$, the set of all dyadic cubes in t_0th dyadic system that are contained in Q_0, such that for μ-almost every $x \in B_0$,

$$\begin{aligned}&|T_b^m(f\chi_{C_{\widetilde{j}_0}B(Q_0)})(x)| \\ &\quad \le C\sum_{k=0}^m C_m^k \sum_{Q\in\mathcal{F}^{t_0}} \Big(|b(x) - b_{R_Q}|^{m-k}\Big||f|\,|b - b_{R_Q}|^k\Big|_{C_{\widetilde{j}_0}B(Q)}\Big)\chi_Q(x).\end{aligned} \tag{3.16}$$

Here, R_Q is the dyadic cube in $\mathscr{D}^t$ for some $t \in \{1, 2, \ldots, \mathcal{T}\}$ such that $C_{\widetilde{j}_0}B(Q) \subset R_Q \subset C_{adj} \cdot C_{\widetilde{j}_0}B(Q)$, where $B(Q)$ is defined as in (2.1), j_0 defined as in (3.11) and $\widetilde{j}_0$ defined as in (3.9).

To prove the claim it suffices to prove the following recursive estimate: there exist pairwise disjoint cubes $P_j \in \mathscr{D}^{t_0}(Q_0)$ such that $\sum_j \mu(P_j) \leq \frac{1}{2}\mu(Q_0)$ and

$$
\begin{aligned}
&|T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{Q_0}(x) \\
&\quad \leq C\sum_{k=0}^{m} C_m^k\Big(|b(x)-b_{R_{Q_0}}|^{m-k}\Big\||f|\,|b-b_{R_{Q_0}}|^k\Big\|_{C_{\tilde{j}_0}B_0}\Big)\chi_{Q_0}(x) \\
&\qquad + \sum_j \big|T_b^m(f\chi_{C_{\tilde{j}_0}B(P_j)})(x)\big|\chi_{P_j}(x) \qquad (3.17)
\end{aligned}
$$

for μ-almost every $x \in B_0$.

Iterating this estimate we obtain (3.16) with $\mathcal{F}^{t_0}$ being the union of all the families $\{P_j^k\}$ where $\{P_j^0\} = \{Q_0\}$, $\{P_j^1\} = \{Q_j\}$ as mentioned above, and $\{P_j^k\}$ are the cubes obtained at the kth stage of the iterative process. It is also clear that $\mathcal{F}^{t_0}$ is a $1/2$-sparse family.

Let us prove then the recursive estimate. We observe that for any arbitrary family of disjoint cubes $\{P_j\} \subset \mathcal{D}^{t_0}(Q_0)$, we have that

$$
\begin{aligned}
&|T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{Q_0}(x) \\
&\quad \leq |T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{Q_0\setminus\cup_j P_j}(x) + \sum_j |T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{P_j}(x) \\
&\quad \leq |T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{Q_0\setminus\cup_j P_j}(x) + \sum_j |T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)\setminus C_{\tilde{j}_0}B(P_j)})(x)|\chi_{P_j}(x) \\
&\qquad + \sum_j |T_b^m(f\chi_{C_{\tilde{j}_0}B(P_j)})(x)|\chi_{P_j}(x).
\end{aligned}
$$

So it suffices to show that we can choose a family of pairwise disjoint cubes $\{P_j\} \subset \mathcal{D}^{t_0}(Q_0)$ with $\sum_j \mu(P_j) \leq \frac{1}{2}\mu(Q_0)$ and such that for μ-almost every $x \in B_0$,

$$
\begin{aligned}
&|T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{Q_0\setminus\cup_j P_j}(x) + \sum_j |T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)\setminus C_{\tilde{j}_0}B(P_j)})(x)|\chi_{P_j}(x) \\
&\quad \leq C\sum_{k=0}^{m} C_m^k |b(x)-b_{R_{Q_0}}|^{m-k}\Big\||f|\,|b-b_{R_{Q_0}}|^k\Big\|_{C_{\tilde{j}_0}B(Q_0)}.
\end{aligned}
$$

To see this, using the fact that

$$
T_b^m f = T_{b-b_{R_{Q_0}}}^m f = \sum_{k=0}^{m}(-1)^k C_m^k T\big((b-b_{R_{Q_0}})^k f\big)(b-b_{R_{Q_0}})^{m-k},
$$

we obtain that

$$
\begin{aligned}
&|T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)})(x)|\chi_{Q_0\setminus\cup_j P_j}(x)+\sum_j |T_b^m(f\chi_{C_{\tilde{j}_0}B(Q_0)\setminus C_{\tilde{j}_0}B(P_j)})(x)|\chi_{P_j}(x)\\
&\quad\le \sum_{k=0}^m C_m^k |T\big((b-b_{R_{Q_0}})^k f\chi_{C_{\tilde{j}_0}B(Q_0)}\big)(x)||b(x)-b_{R_{Q_0}}|^{m-k}\chi_{Q_0\setminus\cup_j P_j}(x)\\
&\qquad+\sum_{k=0}^m C_m^k |T\big((b-b_{R_{Q_0}})^k f\chi_{C_{\tilde{j}_0}B(Q_0)\setminus C_{\tilde{j}_0}B(P_j)}\big)(x)||b(x)-b_{R_{Q_0}}|^{m-k}\chi_{P_j}(x)\\
&\quad=: I_1+I_2.
\end{aligned}
$$

Now, for $k=0,1,\dots,m$, we define the set E_k as

$$
\begin{aligned}
E_k &:= \Big\{x\in B_0 : |b(x)-b_{R_{Q_0}}|^k|f(x)|>\alpha\Big||b-b_{R_{Q_0}}|^k\,|f|\Big|_{C_{\tilde{j}_0}B(Q_0)}\Big\}\\
&\quad\bigcup\Big\{x\in B_0 : \mathcal{M}_{T,B_0}\big((b-b_{R_{Q_0}})^k f\big)(x)>\alpha C_T\Big||b-b_{R_{Q_0}}|^k\,|f|\Big|_{C_{\tilde{j}_0}B(Q_0)}\Big\}
\end{aligned}
$$

and $E:=\cup_{k=0}^m E_k$. Then, choosing α big enough (depending on $C_{\tilde{j}_0}$, C_{adj}, C_μ and A_1 as in (2.1)), we have that

$$\mu(E)\le \frac{1}{4C_{\mu,0}}\mu(B_0),$$

where $C_{\mu,0}$ is the constant in (2.2). We now apply the Calderón–Zygmund decomposition to the function χ_E on B_0 at the height $\lambda:=\frac{1}{2C_{\mu,0}}$, to obtain pairwise disjoint cubes $\{P_j\}\subset\mathcal{D}^{t_0}(Q_0)$ such that

$$\frac{1}{2C_{\mu,0}}\mu(P_j)\le\mu(P_j\cap E)\le\frac{1}{2}\mu(P_j)$$

and $\mu(E\setminus\cup_j P_j)=0$. It follows that

$$\sum_j \mu(P_j)\le\frac{1}{2}\mu(B_0)\quad\text{and}\quad P_j\cap E^c\ne\emptyset.$$

Then, we have

$$\operatorname*{ess\,sup}_{\xi\in P_j}\Big|T\Big(|b-b_{R_{Q_0}}|^k|f|\chi_{C_{\tilde{j}_0}B(Q_0)\setminus C_{\tilde{j}_0}B(P_j)}\Big)(\xi)\Big|\le C\Big||f|\,|b-b_{R_{Q_0}}|^k\Big|_{C_{\tilde{j}_0}B(Q_0)},$$

which allows us to control the summation in the term I_2 above.

Now, from (i) in Lemma 3.6, we obtain that for μ-almost every $x \in B_0$,

$$\begin{aligned}&\left|T\big((b-b_{R_{Q_0}})^k f\chi_{C_{\tilde{j}_0}B(Q_0)}\big)(x)\right|\\&\quad\le C|b(x)-b_{R_{Q_0}}|^k|f(x)|+\mathcal{M}_{T,B_0}\big((b-b_{R_{Q_0}})^k f\chi_{C_{\tilde{j}_0}B(Q_0)}\big)(x).\end{aligned}$$

Since $\mu(E\backslash\cup_j P_j)=0$, we have that from the definition of the set E, the following estimate

$$|b(x)-b_{R_{Q_0}}|^k|f(x)|\le\alpha\Big||f|\,|b-b_{R_{Q_0}}|^k\Big|_{C_{\tilde{j}_0}B(Q_0)}$$

holds for μ-almost every $x\in B_0\backslash\cup_j P_j$, and also

$$\mathcal{M}_{T,B_0}\big((b-b_{R_{Q_0}})^k f\chi_{C_{\tilde{j}_0}B_0}\big)(x)\le\alpha C_T\Big||f|\,|b-b_{R_{Q_0}}|^k\Big|_{C_{\tilde{j}_0}B(Q_0)}$$

holds for μ-almost every $x\in B_0\backslash\cup_j P_j$. These estimates allow us to control the summation in the term I_1 above. Thus, we obtain that (3.17) holds, which yields that (3.16) holds.

We now consider the partition of the space as follows. Suppose f is supported in a ball $B_0\subset X$. We have

$$X=\bigcup_{j=0}^{\infty}2^jB_0.$$

We now consider the annuli $U_j:=2^{j+1}B_0\backslash 2^jB_0$ for $j\ge 0$ and the covering $\{\widetilde{B}_{j,\ell}\}_{\ell=1}^{L_j}$ of U_j as in (3.12). We note that for each $\tilde{B}_{j,\ell}$, there exist $t_{j,\ell}\in\{1,2,\ldots,\mathcal{T}\}$ and $\tilde{Q}_{j,\ell}\in\mathscr{D}^{t_{j,\ell}}$ such that $\tilde{B}_{j,\ell}\subseteq\tilde{Q}_{j,\ell}\subseteq C_{adj}\tilde{B}_{j,\ell}$. Moreover, we note that for each such $\tilde{B}_{j,\ell}$, the enlargement $C_{\tilde{j}_0}B(\tilde{Q}_{j,\ell})$ covers B_0 since $C_{\tilde{j}_0}\widetilde{B}_{j,\ell}$ covers B_0.

We now apply (3.16) to each $\widetilde{B}_{j,\ell}$, then we obtain a $\frac{1}{2}$-sparse family $\tilde{\mathcal{F}}_{j,\ell}\subset\mathscr{D}^{t_{j,\ell}}(\tilde{Q}_{j,\ell})$ such that (3.16) holds for μ-almost every $x\in\widetilde{B}_{j,\ell}$.

Now we set $\mathcal{F}:=\cup_{j,\ell}\tilde{\mathcal{F}}_{j,\ell}$. Note that the balls $C_{adj}\widetilde{B}_{j,\ell}$ are overlapping at most $C_{A_0,\mu,\tilde{j}_0}(2j_0+1)$ times, where $C_{A_0,\mu,\tilde{j}_0}$ is the constant in (3.13). Then, we obtain that $\mathcal{F}$ is a $\frac{1}{2C_{A_0,\mu,\tilde{j}_0}(2j_0+1)}$-sparse family and for μ-almost every $x\in X$,

$$|T_b^m(f)(x)|\le C\sum_{k=0}^{m}C_m^k\sum_{Q\in\mathcal{F}}\Big(|b(x)-b_{R_Q}|^{m-k}\Big||f|\,|b-b_{R_Q}|^k\Big|_{C_{\tilde{j}_0}B(Q)}\Big)\chi_Q(x).$$

Since $C_{\tilde{j}_0}B(Q)\subset R_Q$, and it is clear that $\mu(R_Q)\le\overline{C}\mu(C_{\tilde{j}_0}B(Q))$ ($\overline{C}$ depends only on C_μ and C_{adj}), we obtain that $|f|_{C_{\tilde{j}_0}B(Q)}\le\overline{C}|f|_{R_Q}$. Next, we further set $\mathcal{S}_t:=\{R_Q\in\mathcal{D}^t:\ Q\in\mathcal{F}\},\ \ t\in\{1,2,\ldots,\mathcal{T}\}$, and from the fact that $\mathcal{F}$ is

$\frac{1}{2C_{A_0,\mu,\widetilde{j}_0}(2j_0+1)}$-sparse, we can obtain that each family $\mathcal{S}_t$ is $\frac{1}{2C_{A_0,\mu,\widetilde{j}_0}(2j_0+1)\overline{c}}$-sparse. Now, we let

$$\eta := \frac{1}{2C_{A_0,\mu,\widetilde{j}_0}(2j_0+1)\overline{c}},$$

where $\overline{c}$ is a constant depending only on $\overline{C}$, $C_{\widetilde{j}_0}$ and the doubling constant C_μ. Then it follows that (3.10) holds, which finishes the proof.

In the end, we show (3.14) and (3.15).

We first show (3.14) by contradiction. Suppose there exists some $\widetilde{B}_{j,\ell} = B(x_{j,\ell}, 2^{j-\widetilde{j}_0}r)$ such that $C_{adj}\widetilde{B}_{j,\ell} \cap U_{j+j_0} \neq \emptyset$. Then there exists at least one $y_0 \in C_{adj}\widetilde{B}_{j,\ell} \cap U_{j+j_0}$. Then from the definition of U_{j+j_0} we see that

$$d(x_0, y_0) \geq 2^{j+j_0}r.$$

Moreover, from the definition of $x_{j,\ell}$ and the quasi triangular inequality (1.1) we get that

$$d(x_0, y_0) \leq A_0\big(d(x_0, x_{j,\ell}) + d(x_{j,\ell}, y_0)\big) < A_0\big(2^{j+1}r + C_{adj}2^{j-\widetilde{j}_0}r\big),$$

which, together with the previous inequality, shows that $2^{j+j_0}r \leq A_0\big(2^{j+1}r + C_{adj}2^{j-\widetilde{j}_0}r\big)$. And hence we have

$$2^{j_0} \leq A_0\big(2 + C_{adj}2^{-\widetilde{j}_0}\big) < 3A_0,$$

which contradicts to (3.11). Hence, we see that (3.14) holds.

We now show (3.15), and again we will prove it by contradiction. Suppose there exists some $\widetilde{B}_{j,\ell} = B(x_{j,\ell}, 2^{j-\widetilde{j}_0}r)$ such that $C_{adj}\widetilde{B}_{j,\ell} \cap U_{j-j_0} \neq \emptyset$, where $j \geq j_0$. Then there exists at least one $y_0 \in C_{adj}\widetilde{B}_{j,\ell} \cap U_{j-j_0}$. From the definition of $x_{j,\ell}$ and the quasi triangular inequality (1.1), we see that

$$2^j r \leq d(x_0, x_{j,\ell}) \leq A_0\big(d(x_0, y_0) + d(y_0, x_{j,\ell})\big) < A_0\big(2^{j-j_0+1}r + C_{adj}2^{j-\widetilde{j}_0}r\big),$$

which implies that

$$1 \leq A_0(2^{-j_0+1} + C_{adj}2^{-\widetilde{j}_0}).$$

This contradicts to (3.11) and (3.9). Hence, we see that (3.15) holds. □

4 Upper Bound of the Commutator T_b^m: Proof of Theorem 1.3

In this section we provide the proof of Theorem 1.3 following the idea in [33].

Let $\mathcal{D}$ be a dyadic system in (X, d, μ) and let $\mathcal{S}$ be a sparse family from $\mathcal{D}$. We now define

$$A_b^{m,k} f(x) := \sum_{Q\in\mathcal{S}} |b(x) - b_Q|^{m-k}\left(\frac{1}{\mu(Q)}\int_Q |b(z) - b_Q|^k |f(z)| d\mu(z)\right)\chi_Q(x).$$

By duality, we have that

$$\begin{aligned} \|A_b^{m,k} f\|_{L^p_{\lambda_2}(X)} \le \sup_{g:\|g\|_{L^{p'}_{\lambda_2}(X)}=1} \sum_{Q\in\mathcal{S}} \left(\int_Q |g(x)\lambda_2(x)||b(x) - b_Q|^{m-k} d\mu(x)\right) \\ \times \left(\frac{1}{\mu(Q)}\int_Q |b(z) - b_Q|^k |f(z)| d\mu(z)\right). \end{aligned} \tag{4.1}$$

Now, by Lemma 3.5, there exists a sparse family $\tilde{\mathcal{S}} \subset \mathcal{D}$ such that $\mathcal{S} \subset \tilde{\mathcal{S}}$ and for every cube $Q \in \tilde{\mathcal{S}}$, for μ-almost every $x \in Q$,

$$|b(x) - b_Q| \le C \sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \Omega(b, P)\chi_P(x).$$

Since b is in $\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)$, then we have for μ-almost every $x \in Q$

$$|b(x) - b_Q| \le C\|b\|_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \frac{\nu^{\frac{1}{m}}(P)}{\mu(P)}\chi_P(x).$$

Then, combining this estimate and inequality (4.1), we further have

$$\begin{aligned} &\|A_b^{m,k} f\|_{L^p_{\lambda_2}(X)} \\ &\quad\le C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \sup_{g:\|g\|_{L^{p'}_{\lambda_2}(X)}=1} \sum_{Q\in\mathcal{S}} \left(\frac{1}{\mu(Q)}\int_Q |g(x)\lambda_2(x)| \right.\\ &\quad\quad\left.\times \left(\sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \frac{\nu^{\frac{1}{m}}(P)}{\mu(P)}\chi_P(x)\right)^{m-k} d\mu(x)\right) \\ &\quad\quad\times \left(\frac{1}{\mu(Q)}\int_Q \left(\sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \frac{\nu^{\frac{1}{m}}(P)}{\mu(P)}\chi_P(z)\right)^k |f(z)| d\mu(z)\right)\mu(Q). \end{aligned}$$

Next, note that for each $\ell \in \mathbb{N}$, we have

$$\begin{aligned}
&\Big(\sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \frac{\nu^{\frac{1}{m}}(P)}{\mu(P)}\chi_P(x)\Big)^{\ell} \\
&= \sum_{P_1,P_2,\ldots,P_\ell\in\tilde{\mathcal{S}}, P_1,P_2,\ldots,P_\ell\subset Q} \frac{\nu^{\frac{1}{m}}(P_1)}{\mu(P_1)}\cdots\frac{\nu^{\frac{1}{m}}(P_\ell)}{\mu(P_\ell)}\chi_{\{P_1\cap\cdots\cap P_\ell\}}(x) \\
&\le \ell! \sum_{P_1,\ldots,P_\ell\in\tilde{\mathcal{S}}, P_\ell\subset P_{\ell-1}\cdots\subset P_1\subset Q} \frac{\nu^{\frac{1}{m}}(P_1)}{\mu(P_1)}\cdots\frac{\nu^{\frac{1}{m}}(P_\ell)}{\mu(P_\ell)}\chi_{P_\ell}(x).
\end{aligned}$$

Therefore, for an arbitrary function h, we have

$$\begin{aligned}
&\int_Q |h(x)|\Big(\sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \frac{\nu^{\frac{1}{m}}(P)}{\mu(P)}\chi_P(x)\Big)^{\ell} d\mu(x) \\
&\le \ell! \sum_{P_1,\ldots,P_\ell\in\tilde{\mathcal{S}}, P_\ell\subset P_{\ell-1}\cdots\subset P_1\subset Q} \frac{\nu^{\frac{1}{m}}(P_1)}{\mu(P_1)}\cdots\frac{\nu^{\frac{1}{m}}(P_\ell)}{\mu(P_\ell)}|h|_{P_\ell}\mu(P_\ell) \\
&\le C \sum_{P_1,\ldots,P_{\ell-1}\in\tilde{\mathcal{S}}, P_{\ell-1}\subset P_{\ell-2}\cdots\subset P_1\subset Q} \frac{\nu^{\frac{1}{m}}(P_1)}{\mu(P_1)}\cdots\frac{\nu^{\frac{1}{m}}(P_{\ell-1})}{\mu(P_{\ell-1})} \\
&\quad\times \sum_{P_\ell\subset P_{\ell-1}, P_\ell\in\tilde{\mathcal{S}}} |h|_{P_\ell}\nu^{\frac{1}{m}}(P_\ell) \\
&\le C \sum_{P_1,\ldots,P_{\ell-1}\in\tilde{\mathcal{S}}, P_{\ell-1}\subset P_{\ell-2}\cdots\subset P_1\subset Q} \frac{\nu^{\frac{1}{m}}(P_1)}{\mu(P_1)}\cdots\frac{\nu^{\frac{1}{m}}(P_{\ell-1})}{\mu(P_{\ell-1})} \int_{P_{\ell-1}} \\
&\quad\times A_{\tilde{\mathcal{S}}}(|h|)(x)\nu^{\frac{1}{m}}(x)d\mu(x) \\
&= C \sum_{P_1,\ldots,P_{\ell-1}\in\tilde{\mathcal{S}}, P_{\ell-1}\subset P_{\ell-2}\cdots\subset P_1\subset Q} \frac{\nu^{\frac{1}{m}}(P_1)}{\mu(P_1)}\cdots\frac{\nu^{\frac{1}{m}}(P_{\ell-1})}{\mu(P_{\ell-1})} \\
&\quad\times \Big(A_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|h|)\Big)_{P_{\ell-1}}\mu(P_{\ell-1}),
\end{aligned}$$

where $A_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|h|)(x) := A_{\tilde{\mathcal{S}}}(|h|)(x)\nu^{\frac{1}{m}}(x)$ and $A_{\tilde{\mathcal{S}}}(h) := \sum_{Q\in\tilde{\mathcal{S}}} h_Q\chi_Q$.

By iteration, we obtain that

$$\int_Q |h(x)|\Big(\sum_{P\in\tilde{\mathcal{S}}, P\subset Q} \frac{\nu^{\frac{1}{m}}(P)}{\mu(P)}\chi_P(x)\Big)^{\ell} d\mu(x) \le C\int_Q A^{\ell}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|h|)(x)\,d\mu(x),$$

where $A^{\ell}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}$ denotes the ℓ-fold iteration of $A_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}$. Then we have

$$
\begin{aligned}
\|A_b^{m,k} f\|_{L^p_{\lambda_2}(X)} &\le C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \sup_{g:\|g\|_{L^{p'}_{\lambda_2}(X)}=1} \\
&\qquad \times \sum_{Q\in\mathcal{S}} \left(\frac{1}{\mu(Q)}\int_Q A^{m-k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|g|\,\lambda_2)(x)d\mu(x)\right) \\
&\qquad \times \left(\frac{1}{\mu(Q)}\int_Q A^{k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)(z)d\mu(z)\right)\mu(Q) \\
&\le C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \sup_{g:\|g\|_{L^{p'}_{\lambda_2}(X)}=1} \\
&\qquad \times \left(\int_X A_{\tilde{\mathcal{S}}}\Big(A^{k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\Big)(x) A^{m-k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|g|\,\lambda_2)(x)d\mu(x)\right).
\end{aligned}
$$

Observe that $A_{\tilde{\mathcal{S}}}$ is self-adjoint. We have

$$
\begin{aligned}
&\int_X A_{\tilde{\mathcal{S}}}\Big(A^{k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\Big)(x) A^{m-k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|g|\,\lambda_2)(x)d\mu(x) \\
&= \int_X A_{\tilde{\mathcal{S}}}\left(A_{\tilde{\mathcal{S}}}\Big(A^{k}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\Big)\right)(x) A^{m-k-1}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|g|\,\lambda_2)(x)d\mu(x) \\
&= \cdots \\
&= \int_X A_{\tilde{\mathcal{S}}}\Big(A^{m}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\Big)(x)|g(x)|\,\lambda_2(x)d\mu(x).
\end{aligned}
$$

‘Then from Hölder’s inequality, we further have

$$
\begin{aligned}
&\|A_b^{m,k} f\|_{L^p_{\lambda_2}(X)} \\
&\quad \le C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \sup_{g:\|g\|_{L^{p'}_{\lambda_2}(X)}=1} \\
&\qquad \times \left(\int_X \left[A_{\tilde{\mathcal{S}}}\Big(A^{m}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\Big)(x)\right]^p \lambda_2(x)d\mu(x)\right)^{\frac{1}{p}} \|g\|_{L^{p'}_{\lambda_2}(X)} \\
&\quad \le C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} [\lambda_2]_{A_p}^{\max\{1,\frac{1}{p-1}\}} \left\|A^{m}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\right\|_{L^p_{\lambda_2}(X)} \\
&\quad = C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} [\lambda_2]_{A_p}^{\max\{1,\frac{1}{p-1}\}} \left\|A_{\tilde{\mathcal{S}}}\Big(A^{m-1}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\Big)\right\|_{L^p_{\lambda_2\cdot\nu^{\frac{p}{m}}}(X)} \\
&\quad \le C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)} \Big([\lambda_2]_{A_p}[\lambda_2\cdot\nu^{\frac{p}{m}}]_{A_p}\Big)^{\max\{1,\frac{1}{p-1}\}} \left\|A^{m-1}_{\tilde{\mathcal{S}},\nu^{\frac{1}{m}}}(|f|)\right\|_{L^p_{\lambda_2\cdot\nu^{\frac{p}{m}}}(X)}.
\end{aligned}
$$

Then, by iteration we have that

$$\begin{aligned}\|A_b^{m,k} f\|_{L^p_{\lambda_2}(X)} &\leq C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)}\\ &\quad\times\Big([\lambda_2]_{A_p}[\lambda_2\cdot\nu^{\frac{p}{m}}]_{A_p}\cdots[\lambda_2\cdot\nu^{\frac{mp}{m}}]_{A_p}\Big)^{\max\{1,\frac{1}{p-1}\}}\|f\|_{L^p_{\lambda_2\cdot\nu^{\frac{mp}{m}}}(X)}\\ &\leq C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)}\\ &\quad\times\Big([\lambda_2]_{A_p}[\lambda_1]_{A_p}\prod_{i=1}^{m-1}[\lambda_2^{1-\frac{i}{m}}\cdot\nu^{\frac{i}{m}}]_{A_p}\Big)^{\max\{1,\frac{1}{p-1}\}}\|f\|_{L^p_\nu(X)}.\end{aligned}$$

By Hölder's inequality, we have

$$\prod_{i=1}^{m-1}[{\lambda_2}^{1-\frac{i}{m}}\cdot\nu^{\frac{i}{m}}]_{A_p}\leq\Big([\lambda_2]_{A_p}[\lambda_1]_{A_p}\Big)^{\frac{m-1}{2}}.$$

As a consequence, we have that

$$\|A_b^{m,k} f\|_{L^p_{\lambda_2}(X)}\leq C\|b\|^m_{\mathrm{BMO}_{\nu^{\frac{1}{m}}}(X)}\Big([\lambda_2]_{A_p}[\lambda_1]_{A_p}\Big)^{\frac{m+1}{2}\cdot\max\{1,\frac{1}{p-1}\}}\|f\|_{L^p_\nu(X)}.$$

5 Lower Bound of the Commutator T_b^m: Proof of Theorem 1.4

In this section, we use some ideas from [25,32,33] and adapt them to our general setting with the aim to prove Theorem 1.4. To begin with, let T be the Calderón–Zygmund operator as in Definition 1.1 with the kernel K and ω satisfying $\omega(t)\to 0$ as $t\to 0$, and satisfying the homogeneous condition as in (1.6).

We first introduce another version of the homogeneous condition: There exist positive constants $3\leq A_1\leq A_2$ such that for any ball $B:=B(x_0,r)\subset X$, there exist balls $\widetilde{B}:=B(y_0,r)$ such that $A_1r\leq d(x_0,y_0)\leq A_2r$, and for all $(x,y)\in(B\times\widetilde{B})$, $K(x,y)$ does not change sign and

$$|K(x,y)|\gtrsim\frac{1}{\mu(B)}. \tag{5.1}$$

If the kernel $K(x,y):=K_1(x,y)+iK_2(x,y)$ is complex-valued, where $i^2=-1$, then at least one of K_i satisfies (5.1).

Then, we first point out that the homogeneous condition (1.6) implies (5.1).

Proposition 5.1 *Let T be the Calderón–Zygmund operator as in Definition 1.1 with the kernel K and ω satisfying $\omega(t)\to 0$ as $t\to 0$, and satisfy the homogeneous condition as in (1.6). Then T satisfies (5.1).*

Proof Since $\omega(t) \to 0$ as $t \to 0$, there exists $\delta \in (0, 1)$ such that when $0 < t < \delta$,

$$\omega(t) < \frac{1}{20 \cdot 3^n \cdot C \cdot C_\mu \cdot c_0},$$

where c_0 is from (1.6), C is from Definition 1.1 and C_μ is from (1.3).

For all numbers A with

$$A > \max\Big\{3, 2 + \frac{1}{\delta}, 2A_0\Big\}, \tag{5.2}$$

and for any ball $B := B(x_0, r) \subset X$, according to the homogeneous condition (1.6), there exists a point $y_0 \in B(x_0, \overline{C}Ar)\backslash B(x_0, Ar)$ such that

$$|K(x_0, y_0)| \geq \frac{1}{c_0\mu(B(x_0, Ar))}. \tag{5.3}$$

Next, from the smoothness condition (1.5), we have that for every $x \in B(x_0, r)$ and $y \in B(y_0, r)$,

$$\begin{aligned}
|K(x, y) - K(x_0, y_0)| &\leq |K(x, y) - K(x, y_0)| + |K(x, y_0) - K(x_0, y_0)| \\
&\leq \frac{C}{V(x, y)}\omega\Big(\frac{d(y, y_0)}{d(x, y)}\Big) + \frac{C}{V(x_0, y_0)}\omega\Big(\frac{d(x, x_0)}{d(x_0, y_0)}\Big) \\
&\leq \frac{C}{\mu(B(x_0, (A-2)r))}\omega\Big(\frac{r}{(A-2)r}\Big) \\
&\quad + \frac{C}{\mu(B(x_0, Ar))}\omega\Big(\frac{r}{Ar}\Big) \\
&\leq \frac{2C}{\mu(B(x_0, (A-2)r))}\omega\Big(\frac{1}{A-2}\Big),
\end{aligned}$$

where we use the fact that $\omega(t)$ is increasing. Next, by (1.3), we obtain that

$$\begin{aligned}
|K(x, y) - K(x_0, y_0)| &\leq 2CC_\mu\Big(\frac{A}{A-2}\Big)^n\omega\Big(\frac{1}{A-2}\Big)\,\frac{1}{\mu(B(x_0, Ar))} \\
&\leq \frac{1}{10c_0\mu(B(x_0, Ar))},
\end{aligned}$$

where the last inequality follows from the choice of A as in (5.2).

We now fix a positive number A_1 satisfying (5.2) and set $A_2 := \overline{C}A_1$.

We first consider the kernel $K(x, y)$ to be a real-valued function. If $K(x_0, y_0) > 0$, then for every $x \in B(x_0, r)$ and $y \in B(y_0, r)$ we have that

$$\begin{aligned}
K(x, y) &= K(x_0, y_0) - (K(x_0, y_0) - K(x, y)) \\
&\geq K(x_0, y_0) - |K(x, y) - K(x_0, y_0)| \\
&\geq \frac{1}{c_0\mu(B(x_0, Ar))} - \frac{1}{10c_0\mu(B(x_0, Ar))} > \frac{1}{2c_0\mu(B(x_0, Ar))}.
\end{aligned}$$

Similarly, if $K(x_0, y_0) < 0$, then every $x \in B(x_0, r)$ and $y \in B(y_0, r)$ we have that

$$K(x, y) < -\frac{1}{2c_0\mu(B(x_0, Ar))}.$$

Thus, combining these two cases we obtain that (5.1) holds.

Next, we consider the kernel $K(x, y)$ to be a complex function. We write $K(x, y) = K_1(x, y) + iK_2(x, y)$, with $i^2 = -1$. Then (5.3) implies that

$$\text{either} \quad |K_1(x_0, y_0)| \geq \frac{\sqrt{2}}{2c_0\mu(B(x_0, Ar))} \quad \text{or} \quad |K_2(x_0, y_0)| \geq \frac{\sqrt{2}}{2c_0\mu(B(x_0, Ar))}.$$

Suppose $|K_j(x_0, y_0)| \geq \frac{\sqrt{2}}{2c_0\mu(B(x_0, Ar))}$ for some $j \in \{1, 2\}$. If $K_j(x_0, y_0) > 0$, then every $x \in B(x_0, r)$ and $y \in B(y_0, r)$ we have that

$$\begin{aligned} K_j(x, y) &= K_j(x_0, y_0) - (K_j(x_0, y_0) - K_j(x, y)) \\ &\geq K_j(x_0, y_0) - |K(x, y) - K(x_0, y_0)| \\ &\geq \frac{\sqrt{2}}{2c_0\mu(B(x_0, Ar))} - \frac{1}{10c_0\mu(B(x_0, Ar))} > \frac{1}{2c_0\mu(B(x_0, Ar))}. \end{aligned}$$

Similarly, if $K_j(x_0, y_0) < 0$ for some $j \in \{1, 2\}$, then for every $x \in B(x_0, r)$ and $y \in B(y_0, r)$ we have that

$$K_j(x, y) < -\frac{1}{2c_0\mu(B(x_0, Ar))}.$$

Thus, (5.1) holds for $K_j(x, y)$.

The proof of Proposition 5.1 is complete. □

Definition 5.2 By a median value of a real-valued measurable function f over a ball B we mean a possibly non-unique, real number $\alpha_B(f)$ such that

$$\mu(\{x \in B : f(x) > \alpha_B(f)\}) \leq \frac{1}{2}\mu(B) \text{ and } \mu(\{x \in B : f(x) < \alpha_B(f)\}) \leq \frac{1}{2}\mu(B).$$

It is known that for a given function f and ball B, the median value exists and may not be unique; see, for example, [27].

Lemma 5.3 *Let b be a real-valued measurable function. For any ball B, let $\widetilde{B}$ be as in (5.1). Then there exist measurable sets $E_1, E_2 \subset B$, and $F_1, F_2 \subset \widetilde{B}$, such that*

(i) $B = E_1 \cup E_2$, $\widetilde{B} = F_1 \cup F_2$ *and* $\mu(F_i) \geq \frac{1}{2}\mu(\widetilde{B})$, $i = 1, 2$;
(ii) $b(x) - b(y)$ *does not change sign for all* (x, y) *in* $E_i \times F_i$, $i = 1, 2$;
(iii) $|b(x) - \alpha_{\widetilde{B}}(b)| \leq |b(x) - b(y)|$ *for all* (x, y) *in* $E_i \times F_i$, $i = 1, 2$.

Proof For the given balls B and $\widetilde{B}$, following the idea in [33, Proposition 3.1] we set

$$F_1 := \{y \in \widetilde{B} : b(y) \le \alpha_{\widetilde{B}}(b)\} \quad \text{and} \quad F_2 := \{y \in \widetilde{B} : b(y) \ge \alpha_{\widetilde{B}}(b)\}.$$

Moreover, we define

$$E_1 := \{x \in B : b(x) \ge \alpha_{\widetilde{B}}(b)\} \quad \text{and} \quad E_2 := \{x \in B : b(x) \le \alpha_{\widetilde{B}}(b)\}.$$

Then, by Definition 5.2, we see that $\mu(F_i) \ge \frac{1}{2}\mu(\widetilde{B})$, $i = 1, 2$. Moreover, for $(x, y) \in E_i \times F_i$, $i = 1, 2$,

$$\begin{aligned}|b(x) - b(y)| &= \left|b(x) - \alpha_{\widetilde{B}}(b) + \alpha_{\widetilde{B}}(b) - b(y)\right| \\ &= \left|b(x) - \alpha_{\widetilde{B}}(b)\right| + \left|\alpha_{\widetilde{B}}(b) - b(y)\right| \ge \left|b(x) - \alpha_{\widetilde{B}}(b)\right|.\end{aligned}$$

This finishes the proof of Lemma 5.3. □

We now return to the proof of Theorem 1.4, following the approach and method in [33].

Proof of Theorem 1.4 For given $b \in L^1_{\text{loc}}(X)$ and for any ball B, let $\Omega(b, B)$ be the oscillation as in (3.1). Under the assumptions of Theorem 1.4, we will show that for any ball B,

$$\Omega(b, B) \lesssim \frac{\nu^{\frac{1}{m}}(B)}{\mu(B)}. \tag{5.4}$$

Without loss of generality, we assume that $K(x, y)$ is real-valued. Let B be a ball. We apply the assumption (5.1) and Lemma 5.3 to get sets E_i, F_i, $i = 1, 2$.

On the one hand, by Lemma 5.3 and (5.1), we have that for $f_i := \chi_{F_i}$, $i = 1, 2$,

$$\begin{aligned}&\frac{1}{\mu(B)}\sum_{i=1}^{2}\int_B \left|T_b^m f_i(x)\right| d\mu(x) \\ &\ge \frac{1}{\mu(B)}\sum_{i=1}^{2}\int_{E_i} \left|T_b^m f_i(x)\right| d\mu(x) \\ &= \frac{1}{\mu(B)}\sum_{i=1}^{2}\int_{E_i}\int_{F_i} |b(x) - b(y)|^m |K(x, y)|\, d\mu(y)\, d\mu(x) \\ &\gtrsim \frac{1}{\mu(B)}\sum_{i=1}^{2}\int_{E_i}\int_{F_i} \frac{|b(x) - \alpha_{\widetilde{B}}(b)|^m}{\mu(B)}\, d\mu(y)\, d\mu(x) \\ &\gtrsim \frac{1}{\mu(B)}\int_B \left|b(x) - \alpha_{\widetilde{B}}(b)\right|^m d\mu(x) \\ &\gtrsim \Omega(b; B)^m.\end{aligned}$$

On the other hand, from Hölder's inequality and the boundedness of T_b^m, we deduce that

$$\begin{aligned}
&\frac{1}{\mu(B)}\sum_{i=1}^{2}\int_B \left|T_b^m f_i(x)\right| d\mu(x)\\
&\leq \frac{1}{\mu(B)}\sum_{i=1}^{2}\left[\int_B \left|T_b^m f_i(x)\right|^p \lambda_2(x)\,d\mu(x)\right]^{1/p}\left(\int_B \lambda_2(x)^{-\frac{1}{p-1}}d\mu(x)\right)^{1/p'}\\
&\lesssim \frac{1}{\mu(B)}\sum_{i=1}^{2}[\lambda_1(F_i)]^{1/p}\left(\int_B \lambda_2(x)^{-\frac{1}{p-1}}d\mu(x)\right)^{1/p'}\\
&\lesssim \frac{1}{\mu(B)}[\lambda_1(\widetilde{B})]^{1/p}\left(\int_B \lambda_2(x)^{-\frac{1}{p-1}}d\mu(x)\right)^{1/p'}\\
&\lesssim \frac{1}{\mu(B)}[\lambda_1(B)]^{1/p}\left(\int_B \lambda_2(x)^{-\frac{1}{p-1}}d\mu(x)\right)^{1/p'},
\end{aligned}$$

where in the last inequality, we use the fact that $K_1 r_B \leq d(x_B, x_{\widetilde{B}}) \leq K_2 r_B$ and $\lambda_1 \in A_p$.

Combining the two inequalities above and invoking $\lambda_i \in A_p$, we conclude that

$$\Omega(b, B)^m \lesssim \frac{1}{\mu(B)}[\lambda_1(B)]^{1/p}\left(\int_B \lambda_2(x)^{-\frac{1}{p-1}}d\mu(x)\right)^{1/p'} \lesssim \left(\frac{\nu^{\frac{1}{m}}(B)}{\mu(B)}\right)^m,$$

where the last inequality follows from the argument as in the proof of Theorem 1.1 in [33], by using (2.5). Thus, (5.4) holds and hence, the proof of Theorem 1.4 is complete. □

6 Weighted Hardy Space, Duality and Weak Factorisation: Proof of Theorem 1.5

In this section, we study the weighted Hardy, BMO spaces and duality, as well as their dyadic versions on spaces of homogeneous type.

6.1 Dyadic Littlewood–Paley Square Function

Following the form in [21], we now introduce the dyadic Littlewood–Paley square function on spaces of homogeneous type.

Definition 6.1 Given a dyadic grid $\mathscr{D}$ on X, the dyadic square function $S_{\mathscr{D}}$ is defined by

$$S_{\mathscr{D}} f := \left[\sum_{Q \in \mathscr{D}} \sum_{\epsilon=1}^{M_Q - 1} |\langle f, h_Q^\epsilon \rangle|^2 \frac{\chi_Q}{\mu(Q)} \right]^{\frac{1}{2}}.$$

Our main result in this subsection is:

Theorem 6.2 *Suppose $1 < p < \infty$ and $w \in A_p$. Then we have*

$$\|S_{\mathscr{D}} f\|_{L^p_w(X)} \le C_p [w]_{A_p}^{\max\{1, \frac{1}{p-1}\}} \|f\|_{L^p_w(X)}.$$

We prove this theorem by following the idea in [41, Theorem 3.1 and Corollary 3.2]. We would like to remark that there are additional Euclidean proofs that one could adapt to this setting to achieve this result (e.g. a sparse domination). We don't pursue those additional proofs, but instead give one that utilizes Carleson Embedding and extrapolation. To begin with, we first introduce an auxiliary lemma.

Lemma 6.3 *Let w be an A_2 weight in (X, d, μ). Then*

$$\sum_{Q \in \mathscr{D}} \sum_{\epsilon=1}^{M_Q - 1} |\langle f, h_Q^\epsilon \rangle|^2 \frac{1}{\langle w \rangle_Q} \lesssim [w]_{A_2} \|f\|^2_{L^2_{w^{-1}}(X)} \quad \textit{for all } f \in L^2_{w^{-1}}(X),$$

where $\langle w \rangle_Q := \frac{1}{\mu(Q)} \int_Q w(x) d\mu(x)$.

Proof Recall from [28], we have $h_Q^\epsilon = a_\epsilon \chi_{Q_\epsilon} - b_\epsilon \chi_{E_{\epsilon+1}}$, where

$$a_\epsilon := \sqrt{\frac{\mu(E_{\epsilon+1})}{\mu(Q_\epsilon)\mu(E_\epsilon)}}, \quad b_\epsilon := \sqrt{\frac{\mu(Q_\epsilon)}{\mu(E_\epsilon)\mu(E_{\epsilon+1})}} \quad \text{and} \quad E_\epsilon = Q_\epsilon \cup E_{\epsilon+1},$$

where Q_ϵ and $E_{\epsilon+1}$ are disjoint. Now we introduce the weighted Haar system $\{h_Q^{w,\epsilon}\}_{1 \le \epsilon \le M_Q - 1, Q \in \mathscr{D}}$ in $L^2_w(X)$, where

$$h_Q^{w,\epsilon} := \frac{1}{\sqrt{w(E_\epsilon)}} \left(\frac{\sqrt{w(E_{\epsilon+1})}}{\sqrt{w(Q_\epsilon)}} \chi_{Q_\epsilon} - \frac{\sqrt{w(Q_\epsilon)}}{\sqrt{w(E_{\epsilon+1})}} \chi_{E_{\epsilon+1}} \right).$$

Note that when $w = 1$, we have

$$h_Q^{1,\epsilon} := h_Q^\epsilon = \frac{1}{\sqrt{\mu(E_\epsilon)}} \left(\frac{\sqrt{\mu(E_{\epsilon+1})}}{\sqrt{\mu(Q_\epsilon)}} \chi_{Q_\epsilon} - \frac{\sqrt{\mu(Q_\epsilon)}}{\sqrt{\mu(E_{\epsilon+1})}} \chi_{E_{\epsilon+1}} \right).$$

We set

$$h^1_{E_\epsilon} := \frac{\chi_{E_\epsilon}}{\mu(E_\epsilon)},$$

and write $h_Q^\epsilon = C_Q(w, \epsilon) h_Q^{w,\epsilon} + D_Q(w, \epsilon) h^1_{E_\epsilon}$.

It is easy to see that $\int_Q h_Q^{w,\epsilon} dw = 0$ and $\int_Q (h_Q^{w,\epsilon})^2 dw = 1$. This implies

$$D_Q(w,\epsilon) = \frac{\hat{w}(Q,\epsilon)}{\langle w\rangle_{E_\epsilon}}, \quad \text{where } \hat{w}(Q,\epsilon) := \langle w, h_Q^\epsilon\rangle$$

and, after some computation,

$$\begin{aligned} C_Q(w,\epsilon)^2 = &\frac{\mu(E_{\epsilon+1})}{\mu(E_\epsilon)}\langle w\rangle_{Q_\epsilon} + \frac{\mu(Q_\epsilon)}{\mu(E_\epsilon)}\langle w\rangle_{E_{\epsilon+1}} - \frac{\mu(E_{\epsilon+1})}{\mu(E_\epsilon)}\frac{w(Q_\epsilon)}{w(E_\epsilon)}\langle w\rangle_{Q_\epsilon} \\ &- \frac{\mu(Q_\epsilon)}{\mu(E_\epsilon)}\frac{w(E_{\epsilon+1})}{w(E_\epsilon)}\langle w\rangle_{E_{\epsilon+1}} + 2\frac{w(E_{\epsilon+1})}{w(E_\epsilon)}\frac{w(Q_\epsilon)}{\mu(E_\epsilon)}. \end{aligned}$$

Note that it does not really matter what $C_Q(w,\epsilon)$ really is as long as we have some nice bound for it. In fact, from Lemma 4.6 in [28], we have that

$$\begin{aligned} C_Q(w,\epsilon)^2 &\le \frac{\mu(E_{\epsilon+1})}{\mu(E_\epsilon)}\langle w\rangle_{Q_\epsilon} + \frac{\mu(Q_\epsilon)}{\mu(E_\epsilon)}\langle w\rangle_{E_{\epsilon+1}} \\ &\lesssim \frac{w(Q_\epsilon) + w(E_{\epsilon+1})}{\mu(E_\epsilon)} = \langle w\rangle_{E_\epsilon} \lesssim \langle w\rangle_Q, \end{aligned}$$

which implies that $C_Q(w,\epsilon)^2\langle w\rangle_Q^{-1} \lesssim 1$.

Now,

$$\begin{aligned} &\sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1}\langle w\rangle_Q^{-1}|\langle f, h_Q^\epsilon\rangle|^2 \\ &= \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1}\langle w\rangle_Q^{-1}|\langle f, C_Q(w,\epsilon)h_Q^{w,\epsilon} + D_Q(w,\epsilon)h_{E_\epsilon}^1\rangle|^2 \\ &= \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1}\langle w\rangle_Q^{-1}|C_Q(w,\epsilon)\langle f, h_Q^{w,\epsilon}\rangle + D_Q(w,\epsilon)\langle f, h_{E_\epsilon}^1\rangle|^2 \\ &= \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1}\langle w\rangle_Q^{-1}C_Q(w,\epsilon)^2|\langle f, h_Q^{w,\epsilon}\rangle|^2 \\ &\quad + 2\sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1}\langle w\rangle_Q^{-1}C_Q(w,\epsilon)\,D_Q(w,\epsilon)\,\langle f, h_Q^{w,\epsilon}\rangle\,\langle f, h_{E_\epsilon}^1\rangle \\ &\quad + \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1}\langle w\rangle_Q^{-1}D_Q(w,\epsilon)^2|\langle f, h_{E_\epsilon}^1\rangle|^2 \\ &=: S_1 + S_2 + S_3. \end{aligned}$$

S_2 can be bounded by $\sqrt{S_1}\sqrt{S_3}$, so it suffices to bound S_1 and S_3. By using the bound on $C_Q(w,\epsilon)$, we have

$$S_1 \lesssim \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1} |\langle f, h_Q^{w,\epsilon}\rangle|^2 = \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1} |\langle w^{-1} f, h_Q^{w,\epsilon}\rangle_{L^2_w(X)}|^2 \leq \|f\|^2_{L^2_{w^{-1}}(X)}.$$

On the other hand,

$$S_3 = \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1} \langle w\rangle_Q^{-1} D_Q(w,\epsilon)^2 |\langle f, h^1_{E_\epsilon}\rangle|^2 = \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1} D_Q(w,\epsilon)^2 \langle f\rangle^2_{E_\epsilon}\langle w\rangle_Q^{-1}.$$

Now,

$$\begin{aligned}
&\frac{1}{\mu(E_\epsilon)}\sum_{R\subseteq Q}\sum_{\eta:E_\eta\subseteq E_\epsilon} D_R(w,\eta)^2\langle w\rangle_R^{-1}\langle w\rangle^2_{E_\eta}\\
&= \frac{1}{\mu(E_\epsilon)}\sum_{R\subseteq Q}\sum_{\eta:E_\eta\subseteq E_\epsilon} \frac{\hat{w}(R,\eta)^2}{\langle w\rangle^2_{E_\eta}}\langle w\rangle_R^{-1}\langle w\rangle^2_{E_\eta}\\
&\lesssim \frac{1}{\mu(E_\epsilon)}\sum_{R\subseteq Q}\sum_{\eta:E_\eta\subseteq E_\epsilon} \hat{w}(R,\eta)^2\langle w\rangle_R^{-1}\\
&\lesssim [w]_{A_2}\langle w\rangle_{E_\epsilon},
\end{aligned}$$

where the last inequality follows from a Bellman function technique that can be found in [6]. Thus, by adopting the remark of Treil in [45, Section 5] on the dyadic Carleson Embedding Theorem on a general space of homogeneous type, we get:

$$S_3 = \sum_{Q\in\mathscr{D}}\sum_{\epsilon=1}^{M_Q-1} D_Q(w,\epsilon)^2\langle f\rangle^2_{E_\epsilon}\langle w\rangle_Q^{-1} \lesssim [w]_{A_2}\|f w^{-\frac{1}{2}}\|^2_{L^2(X)}.$$

The proof of Lemma 6.3 is complete. □

Proof of Theorem 6.2 Suppose $w \in A_2$. Following the argument in the proof of Theorem 3.1 in [41], we obtain that the Lemma 6.3 above implies that

$$\|f\|_{L^2_w(X)} \lesssim [w]_{A_2}^{\frac{1}{2}}\|S_{\mathscr{D}} f\|_{L^2_w(X)},$$

where the implicit constant is independent of f and w.

Then following the argument in the proof of Corollary 3.2 in [41], we obtain that

$$\|S_{\mathscr{D}} f\|_{L^2_w(X)} \lesssim [w]_{A_2}\|f\|_{L^2_w(X)}.$$

Next, by the sharp form of Rubio de Francia's extrapolation theorem (due to Dragiĉević, Grafakos, Pereyra and Petermichl [12] in the Euclidean space and due to Anderson and Damián [1] on spaces of homogeneous type), this implies the corresponding weighted L^p bound

$$\|S_{\mathscr{D}} f\|_{L^p_w(X)} \le C_p [w]_{A_p}^{\max\{1,\frac{1}{p-1}\}} \|f\|_{L^p_w(X)}.$$

The proof of Theorem 6.2 is complete. □

6.2 Weighted Hardy Spaces, Duality and Weak Factorisation

We now introduce the atoms for the weighted Hardy space.

Definition 6.4 Suppose $w \in A_2$. A function a is called a $(1, 2)$-atom, if there exists a ball $B \subset X$ such that

$$(1)\ \operatorname{supp}(a) \subset B; \quad (2)\ \int_B a(x)\, d\mu(x) = 0; \quad (3)\ \|a\|_{L^2_w(B)} \le [w(B)]^{-\frac{1}{2}}.$$

Definition 6.5 Suppose $w \in A_2$. A function f is said to belong to the Hardy space $H^1_{w,2}(X)$, if $f = \sum_{j=1}^{\infty} \lambda_j a_j$ with $\sum_{j=1}^{\infty} |\lambda_j| < \infty$ and a_j is a $(1, 2)$-atom for each j. Moreover, the norm of f on $H^1_{w,2}(X)$ is defined by $\|f\|_{H^1_{w,2}(X)} = \inf \big\{ \sum_{j=1}^{\infty} |\lambda_j| \big\}$, where the infimum is taken over all possible decompositions of f as above.

We then have the duality between the weighted Hardy space and weighted BMO. We point out that for the sake of simplicity, we obtain this result for $p = 2$. For the Euclidean version of duality of weighted Hardy and BMO spaces, we refer to [13, Section 4] for the full range of $p \in [1, 2]$.

Theorem 6.6 *Suppose $w \in A_2$. Then we have $\big(H^1_{w,2}(X)\big)' = \mathrm{BMO}_w(X)$.*

Proof To prove $\mathrm{BMO}_w(X) \subset \big(H^1_{w,2}(X)\big)'$, for any $g \in \mathrm{BMO}_w(X)$, let

$$\ell_g(a) = \int_X a(x) g(x) d\mu(x),$$

where a is an atom as in Definition 6.4.

Assume that a is supported in a ball $B \subset X$. Then by Hölder's inequality and $w \in A_2$, we see that

$$
\begin{aligned}
\left|\int_X g(x)a(x)\,d\mu(x)\right| &= \left|\int_B [g(x)-g_B]a(x)\,d\mu(x)\right| \\
&\le \left[\int_B |g(x)-g_B|^2 w^{-1}(x)\,d\mu(x)\right]^{\frac12}\left[\int_B |a(x)|^2 w(x)\,d\mu(x)\right]^{\frac12} \\
&\le \left[\frac{1}{w(B)}\int_B |g(x)-g_B|^2 w^{-1}(x)\,d\mu(x)\right]^{\frac12} \\
&\le C\|g\|_{\mathrm{BMO}_w(X)}.
\end{aligned}
$$

Thus ℓ_g can be extended to a bounded linear functional on $H^1_{w,2}(X)$.

Conversely, assume that $\ell \in \left(H^1_{w,2}(X)\right)'$. For any ball $B \subset X$, let

$$L^2_{0,\,w}(B) = \left\{f \in L^2_w(B) : \mathrm{supp}(f) \subset B,\ \int_B f(x)\,d\mu(x) = 0\right\}.$$

Then we see that for any $f \in L^2_{0,\,w}(B)$, $a := \frac{1}{[w(B)]^{\frac12}\|f\|_{L^2_w(B)}} f$ is an atom as in Definition 6.4. This implies that

$$|\ell(a)| \le \|\ell\|\|a\|_{H^1_{w,2}(X)} \le \|\ell\|.$$

Moreover, we see that

$$|\ell(f)| \le \|\ell\|[w(B)]^{\frac12}\|f\|_{L^2_w(B)}.$$

From the Riesz Representation theorem, there exists $[\varphi] \in [L^2_{0,\,w}(B)]^* = L^2_{w^{-1}}(B)/\mathbb{C}$, and $\varphi \in [\varphi]$, such that for any $f \in L^2_{0,\,w}(B)$,

$$\ell(f) = \int_B f(x)\varphi(x)\,d\mu(x)$$

and

$$\|[\varphi]\| = \inf_c \|\varphi + c\|_{L^2_{w^{-1}}(B)} \le \|\ell\|[w(B)]^{\frac12}.$$

Now, for a fixed ball B, we define $B_j = 2^j B$, $j \in \mathbb{N}$. And for B_0, we mean the ball B itself. Then, we have that for all $f \in L^2_{0,\,w}(B)$ and $j \in \mathbb{N}$,

$$\int_B f(x)\varphi_B(x)\,d\mu(x) = \int_B f(x)\varphi_{B_j}(x)\,d\mu(x).$$

It follows that for μ-almost every $x \in B$, $\varphi_{B_j}(x) - \varphi_{B_0}(x) = C_j$ for some constant C_j. From this we further deduce that for all j, $l \in \mathbb{N}$, $j \le l$ and μ-almost every $x \in B_j$,

$$\varphi_{B_j}(x) - C_j = \varphi_{B_0}(x) = \varphi_{B_l}(x) - C_l.$$

Define $\varphi(x) = \varphi_j(x) - C_j$ on B_j for $j \in \mathbb{N}$. Then, φ is well defined. Moreover, since $X = \cup_j B_j$, by Hölder's inequality and $w \in A_2$, we see that for any c and any ball $B \subset X$,

$$\begin{aligned}
\left[\int_B |\varphi(x) - \varphi_B|^2 w^{-1}(x)\, d\mu(x)\right]^{\frac{1}{2}} &= \sup_{\|f\|_{L^2_w(B)} \le 1} |\langle f, \varphi - \varphi_B\rangle| \\
&= \sup_{\|f\|_{L^2_w(B)} \le 1} \left|\int_B f(x)[\varphi(x) - \varphi_B]\, d\mu(x)\right| \\
&= \sup_{\|f\|_{L^2_w(B)} \le 1} \left|\int_B [f(x) - f_B][\varphi(x) + c]\, d\mu(x)\right| \\
&\le \sup_{\|f\|_{L^2_w(B)} \le 1} \left[\|f\|_{L^2_w(B)} + |f_B|[w(B)]^{\frac{1}{2}}\right] \|[\varphi(x) + c]\chi_B\|_{L^2_{w^{-1}}(B)} \\
&\le \|[\varphi(x) + c]\chi_B\|_{L^2_{w^{-1}}(B)}.
\end{aligned}$$

Taking the infimum over c, we have that $\varphi \in \mathrm{BMO}_w(X)$ and $\|\varphi\|_{\mathrm{BMO}_w(X)} \le C\|\ell\|$. □

We now provide a sketch of the proof of Theorem 1.5, the details are similar to the weak factorisation result obtained in [8].

Proof of Theorem 1.5 Similar to [8] (see also Corollary 1.4 in [21] and its proof), as a consequence of the duality of weighted Hardy space $H^1_\nu(X)$ and $\mathrm{BMO}_\nu(X)$ (Theorem 6.6 above), we see that if f is of the form (1.7), then f is in $H^1_\nu(X)$ with the $H^1_\nu(X)$-norm controlled by the right-hand side of (1.8). Conversely, based the characterisation of $\mathrm{BMO}_\nu(X)$ via the commutator $[b, T]$ as in Theorem 1.4 (the case of $m = 1$) and the linear functional analysis argument as in [7], we get that every $f \in H^1_\nu(X)$ admits an factorisation as in (1.7). Hence, we obtain that Theorem 1.5 holds. □

7 Applications

The aim of this section is to show that Theorem 1.6 holds for each of the six operators listed in the introduction.

7.1 Cauchy's Integral Operator

Let $A(x)$ be a Lipschitz function on $\mathbb{R}$. Consider the Lipschitz curve as $z = x + iA(x)$, $x \in (-\infty, \infty)$. Recall that the Cauchy integral adapted to this Lipschitz curve is:

$$C_A(f)(x) = p.v.\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(y)dy}{(x-y) + i(A(x) - A(y))}.$$

The (unweighted version) commutator result was obtained in [34], see also [44]. Here we point out that the two weight commutator and high-order commutator results also hold for Cauchy's Integral Operator.

Proposition 7.1 *Theorem 1.6 holds for the Cauchy integral operator C_A with the underlying setting* $(\mathbb{R}, |\cdot|, dx)$.

Proof To see this, we point out that this operator has the associated kernel

$$C_A(x, y) = \frac{1}{\pi} \frac{1}{(x-y) + i(A(x) - A(y))},$$

which satisfies the size condition

$$|C_A(x, y)| \le \frac{1}{|x-y|}$$

and the smoothness condition

$$|C_A(x, y) - C_A(x, y')| + |C_A(y, x) - C_A(y', x)| \le 2(\|A'\|_\infty + 1) \frac{|y-y'|}{|x-y|^2}$$

for every x, y, y' such that $|y - y'| \le |x - y|/2$. Moreover, for any interval $I := I(x_0, r)$, we take $y_0 = x_0 + 4r$. Then we see that $\mathrm{Re}\, C_A(x_0, y_0)$, the real part of $C_A(x_0, y_0)$, satisfies that $\mathrm{Re}\, C_A(x_0, y_0) < 0$ and

$$\begin{aligned} |\mathrm{Re}\, C_A(x_0, y_0)| &= \frac{1}{\pi} \frac{y_0 - x_0}{(x_0 - y_0)^2 + (A(x_0) - A(y_0))^2} \\ &\gtrsim \frac{y_0 - x_0}{(\|A'\|_\infty^2 + 1)(x_0 - y_0)^2} \gtrsim \frac{1}{|I|}. \end{aligned}$$

Therefore, (1.6) holds. As a consequence of this fact and Theorems 1.3 and 1.4, we see that Theorem 1.6 holds. □

7.2 The Cauchy–Szegö Projection Operator on the Heisenberg Group $\mathbb{H}^n$

We recall all the related definitions for the Heisenberg group in [43, Chapter XII]. Recall that $\mathbb{H}^n$ is the Lie group with underlying manifold $\mathbb{C}^n \times \mathbb{R} = \{[z, t] : z = (z_1, \cdots, z_n) \in \mathbb{C}^n, t \in \mathbb{R}\}$ and multiplication law

$$[z, t] \circ [z', t'] := \Big[z_1 + z_1', \cdots, z_n + z_n', t + t' + 2\mathrm{Im}\big(\sum_{j=1}^n z_j \bar{z}_j'\big)\Big].$$

The identity of $\mathbb{H}^n$ is the origin and the inverse is given by $[z, t]^{-1} = [-z, -t]$. Hereafter, we identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ and use the following notation to denote the points of $\mathbb{C}^n \times \mathbb{R} = \mathbb{R}^{2n+1}$: $g = [z, t] = [x, y, t] = [x_1, \cdots, x_n, y_1, \cdots, y_n, t]$ with

$z = [z_1, \cdots, z_n]$, $z_j = x_j + iy_j$ and $x_j, y_j, t \in \mathbb{R}$ for $j = 1, \ldots, n$. Then, the composition law $\circ$ can be explicitly written as

$$g \circ g' = [x, y, t] \circ [x', y', t'] = [x + x', y + y', t + t' + 2\langle y, x'\rangle - 2\langle x, y'\rangle],$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^n$.

We recall the upper half-space $\mathcal{U}^n$ and its boundary $b\mathcal{U}^n$ as follows:

$$\mathcal{U}^n = \Big\{z \in \mathbb{C}^{n+1} : \operatorname{Im}(z_{n+1}) > \sum_{j=1}^{n} |z_j|^2\Big\},$$

$$b\mathcal{U}^n = \Big\{z \in \mathbb{C}^{n+1} : \operatorname{Im}(z_{n+1}) = \sum_{j=1}^{n} |z_j|^2\Big\}.$$

For any function F defined on $\mathcal{U}^n$, we write F_ϵ for its vertical translate: $F_\epsilon(z) = F(z + \epsilon\mathbf{i})$ with $\mathbf{i} = (0, \ldots, 0, i)$. We also recall the Hardy space $\mathcal{H}^2(\mathcal{U}^n)$, which consists of all functions F holomorphic on $\mathcal{U}^n$ for which

$$\|F\|_{\mathcal{H}^2(\mathcal{U}^n)} = \Big(\sup_{\epsilon>0} \int_{b\mathcal{U}^n} |F_\epsilon(z)|^2 d\beta(z)\Big)^{\frac{1}{2}} < \infty,$$

where $d\beta(z)$ is the surface measure on $b\mathcal{U}^n$.

The Cauchy–Szegö projection operator $\mathcal{C}$ is the orthogonal projection from $L^2(b\mathcal{U}^n)$ to the subspace of functions $\{F^b\}$ that are boundary values of functions $F \in \mathcal{H}^2(\mathcal{U}^n)$. According to [43, Section 2.3, Section 2.4, Chapter XII], we get that for $f \in L^2(\mathbb{H}^n)$,

$$\mathcal{C}(f)(x) = \int_{\mathbb{H}^n} K(x, y) f(y) dy,$$

where $K(x, y) = K(y^{-1} \circ x)$ for $x \neq y$ and

$$K(x) = -\frac{\partial}{\partial t}\Big(\frac{c}{n}[t + i|\zeta|^2]^{-n}\Big) \quad \text{for} \quad x = [\zeta, t] \in \mathbb{H}^n = \mathbb{C}^n \times \mathbb{R},$$

and $c = 2^{n-1} i^{n+1} n! \pi^{-n-1}$.

Proposition 7.2 *Theorem 1.6 holds for the Cauchy–Szegö projection operator $\mathcal{C}$ with the underlying setting $(\mathbb{H}^n, \rho, dx)$, where dx is the usual Lebesgue measure on $\mathbb{C}^n \times \mathbb{R}$ and ρ is the norm on $\mathbb{H}^n$ defined by $\rho(x) := \max\{|\zeta|, |t|^{\frac{1}{2}}\}$ for $x = [\zeta, t] \in \mathbb{H}^n = \mathbb{C}^n \times \mathbb{R}$.*

Proof We begin by recalling that with this norm $\rho(x)$ as above, and we set $\rho(x, y) := \rho(y^{-1} \circ x)$. From [43, Section 2.5, Chapter XII] we obtain that the Cauchy–Szegö kernel $K(x, y)$ satisfies the following conditions:

$$|K(x,y)| \approx \rho(x,y)^{-2n-2},$$
$$|K(x,y)-K(x,y_0)| \lesssim \frac{\rho(y,y_0)}{\rho(x,y)}\frac{1}{\rho(x,y)^{2n+2}} \quad \text{whenever} \quad \rho(x,y)\geq c\rho(y,y_0),$$
$$|K(x,y)-K(x_0,y)| \lesssim \frac{\rho(x,x_0)}{\rho(x,y)}\frac{1}{\rho(x,y)^{2n+2}} \quad \text{whenever} \quad \rho(x,y)\geq c\rho(x,x_0).$$

Thus, it is straightforward to see that $K(x,y)$ satisfies (1.6). Hence, we see that Theorem 1.6 holds for the Cauchy–Szegö projection operator $\mathcal{C}$. □

7.3 The Szegö Projection Operator on a Family of Unbounded Weakly Pseudoconvex Domains

We now recall the weakly pseudoconvex domains Ω_k and their boundary $\partial\Omega_k$, $k\in\mathbb{Z}_+$, from Greiner and Stein [18]:

$$\Omega_k = \left\{(z_1,z_2)\in\mathbb{C}^2 : \mathrm{Im}(z_2) > |z_1|^{2k}\right\}, \quad \partial\Omega_k = \left\{(z_1,z_2)\in\mathbb{C}^2 : \mathrm{Im}(z_2) = |z_1|^{2k}\right\}.$$

Recall that $\partial\Omega_k$ is naturally parameterized by z_1 and $\mathrm{Re}z_2$. We use the following notation. Points in $\partial\Omega_k$ are denoted by ζ,ω,ν etc.

$$\zeta=(z_1,z_2)\sim(z,t),\quad z=z_1\in\mathbb{C},\ t=\mathrm{Re}(z_2)\in\mathbb{R};$$
$$\omega=(w_1,w_2)\sim(w,s),\quad w=w_1\in\mathbb{C},\ s=\mathrm{Re}(w_2)\in\mathbb{R};$$
$$\nu=(u_1,u_2)\sim(u,r),\quad u=u_1\in\mathbb{C}, r=\mathrm{Re}(u_2)\in\mathbb{R}.$$

The Szegö projection S on Ω_k is the orthogonal projection from $L^2(\partial\Omega_k)$ to the Hardy space $H^2(\Omega_k)$ of holomorphic functions on Ω_k with L^2 boundary values. The Szegö kernel $S(\zeta,\omega)$ is the kernel for which

$$S(f)(\zeta)=\int_{\partial\Omega_k} f(\omega)S(\zeta,\omega)dV(\omega),$$

where $dV(\omega)=dV(x,y,s)=dxdyds$ with $\omega=(w_1,s)=(x+iy,s)$, which is Lebesgue measure on the parameter space $\mathbb{R}^3$. Greiner and Stein [18] have computed the Szegö kernel with Lebesgue measure on the parameter space with the formula

$$S(\zeta,\omega)=\frac{1}{4\pi^2}\Bigg[\left(\left(\frac{i}{2}[s-t]+\frac{|z_1|^{2k}+|w_1|^{2k}}{2}+\frac{\mu+\eta}{2}\right)^{\frac{1}{k}}-z_1\bar{w}_1\right)^2$$
$$\times\left(\frac{i}{2}[s-t]+\frac{|z_1|^{2k}+|w_1|^{2k}}{2}+\frac{\mu+\eta}{2}\right)^{\frac{k-1}{k}}\Bigg]^{-1},$$

where $\mu=\mathrm{Im}(z_2)-|z_1|^{2k}$ and $\eta=\mathrm{Im}(w_2)-|w_1|^{2k}$.

In [11], Diaz defined and analyzed a pseudometric $d(\zeta,\omega)$ globally suited to the complex geometry of $\partial\Omega_k$, which was arrived at by the study of the Szegö kernel. This allows the treatment of the Szegö kernel as a singular integral kernel:

$$d(\zeta,\omega)=\left|\left(\frac{i}{2}[s-t]+\frac{|z_1|^{2k}+|w_1|^{2k}}{2}\right)^{\frac{1}{k}}-z\bar{w}\right|^{\frac{1}{2}}.$$

Then, the pseudometric balls are defined as

$$B_\zeta(\delta)=B_\zeta^d(\delta)=\{\omega\in\partial\Omega_k:\ d(\zeta,\omega)<\delta\}$$

and the volume of the balls is

$$V(B_\zeta(\delta))=4\pi\delta^2\left(\frac{(\sin(\pi/k))^{2k-2}}{4}|z|^{2k-2}\delta^2+\frac{1}{2}\delta^{2k}\right),$$

and it is shown that this measure is doubling.

Proposition 7.3 *Theorem 1.6 holds for the Szegö projection S with the underlying space of homogenous type $(\partial\Omega_k, d, dV)$, where d and dV are as introduced above.*

Proof We point out that it is proved in [11] that $S(\zeta,\omega)$ satisfies the following size and smoothness conditions:

$$\begin{aligned}|S(\zeta,\omega)|&\approx\frac{1}{V(B_\zeta(d(\zeta,\omega)))},\\ |S(\zeta,\omega)-S(\zeta',\omega)|&\lesssim\left(\frac{d(\zeta,\zeta')}{d(\zeta,\omega)}\right)\frac{1}{V(B_\zeta(d(\zeta,\omega)))},\quad\text{for } cd(\zeta,\zeta')\le d(\zeta,\omega),\\ |S(\zeta,\omega)-S(\zeta,\omega')|&\lesssim\left(\frac{d(\omega,\omega')}{d(\zeta,\omega)}\right)\frac{1}{V(B_\zeta(d(\zeta,\omega)))},\quad\text{for } cd(\omega,\omega')\le d(\zeta,\omega).\end{aligned}\tag{7.1}$$

Thus, from (7.1) it is direct to see that $S(\zeta,\omega)$ satisfies (1.6). Hence, we see that Theorem 1.6 holds for the Szegö projection operator S on Ω_k for $k\in\mathbb{Z}_+$. □

7.4 Riesz Transforms Associated with Sub-Laplacian on Stratified Nilpotent Lie Groups

Recall that a connected, simply connected nilpotent Lie group $\mathcal{G}$ is said to be stratified if its left-invariant Lie algebra $\mathfrak{g}$ (assumed real and of finite dimension) admits a direct sum decomposition

$$\mathfrak{g}=\bigoplus_{i=1}^{k}V_i \text{ where } [V_1,V_i]=V_{i+1} \text{ for } i\le k-1.$$

One identifies $\mathfrak{g}$ and $\mathcal{G}$ via the exponential map $\exp:\mathfrak{g}\longrightarrow\mathcal{G}$, which is a diffeomorphism. We fix once and for all a (bi-invariant) Haar measure dg on $\mathcal{G}$ (which is just the lift of Lebesgue measure on $\mathfrak{g}$ via exp). There is a natural family of dilations on $\mathfrak{g}$ defined for $r > 0$ as follows:

$$\delta_r\left(\sum_{i=1}^{k} v_i\right) = \sum_{i=1}^{k} r^i v_i, \quad \text{with } v_i \in V_i.$$

This allows the definition of dilation on $\mathcal{G}$, which we still denote by δ_r. We choose once and for all a basis $\{\mathrm{X}_1, \cdots, \mathrm{X}_n\}$ for V_1 and consider the sub-Laplacian $\Delta = \sum_{j=1}^{n} \mathrm{X}_j^2$. Observe that X_j $(1 \leq j \leq n)$ is homogeneous of degree 1 and Δ of degree 2 with respect to the dilations in the sense that: $\mathrm{X}_j\,(f \circ \delta_r) = r\,\left(\mathrm{X}_j f\right) \circ \delta_r,\ 1 \leq j \leq n,\ r > 0,\ f \in C^1$ and that $\delta_{\frac{1}{r}} \circ \Delta \circ \delta_r = r^2\,\Delta, \quad \forall r > 0.$

Let Q denote the homogeneous dimension of $\mathcal{G}$, namely, $Q = \sum_{i=1}^{k} i \dim V_i$. And let p_h $(h > 0)$ be the heat kernel (that is, the integral kernel of $e^{h\Delta}$) on $\mathcal{G}$. For convenience, we set $p_h(g) = p_h(g, o)$ (that is, in this article, for a convolution operator, we will identify the integral kernel with the convolution kernel) and $p(g) = p_1(g)$.

Recall that (c.f. for example [17]) $p_h(g) = h^{-\frac{Q}{2}} p(\delta_{\frac{1}{\sqrt{h}}}(g)),\ \forall h > 0,\ g \in \mathcal{G}$. The kernel of the jth Riesz transform $\mathrm{X}_j(-\Delta)^{-\frac{1}{2}}$ $(1 \leq j \leq n)$ is written simply as $K_j(g, g') = K_j(g'^{-1} \circ g)$, where

$$K_j(g) = \frac{1}{\sqrt{\pi}} \int_0^{+\infty} h^{-\frac{1}{2}} \mathrm{X}_j p_h(g)\, dh = \frac{1}{\sqrt{\pi}} \int_0^{+\infty} h^{-\frac{Q}{2}-1} \left(\mathrm{X}_j p\right) (\delta_{\frac{1}{\sqrt{h}}}(g))\, dh.$$

Proposition 7.4 *Theorem 1.6 holds for the Riesz transform* $\mathrm{X}_j(-\Delta)^{-\frac{1}{2}}$ $(1 \leq j \leq n)$ *with the underlying setting* $(\mathcal{G}, \rho, dg)$*, where* ρ *is the homogeneous norm on* $\mathcal{G}$ *(see* [14]).

Proof For the Riesz transform kernel, we have the following lower bound estimate, obtained in [14]:

Fix $j = 1, \ldots, n$. *There exist* $0 < \varepsilon_o \ll 1$ *and* $C > 0$ *such that for any* $0 < \eta < \varepsilon_o$ *and for all* $g \in \mathcal{G}$ *and* $r > 0$, *we can find* $g_* = g_*(j, g, r) \in \mathcal{G}$ *satisfying*

$$\rho(g, g_*) = r, \quad |K_j(g_1, g_2)| \geq C r^{-Q}, \quad \forall g_1 \in B(g, \eta r),\ g_2 \in B(g_*, \eta r)$$

and all $K_j(g_1, g_2)$ *have the same sign.*

From this kernel lower bound estimate, it is direct to see that for each j, $K_j(g_1, g_2)$ satisfies (1.6). Hence, Theorem 1.6 holds for $\mathrm{X}_j(-\Delta)^{-\frac{1}{2}}$ $(1 \leq j \leq n)$. ◻

7.5 Riesz Transform Associated with the Bessel Operator on $\mathbb{R}_+$

Consider $\mathbb{R}_+ = (0, \infty)$. For $\lambda > -\frac{1}{2}$, the Bessel operator Δ_λ on $\mathbb{R}_+$ [39] is defined by

$$\Delta_\lambda = -\frac{d^2}{dx^2} - \frac{2\lambda}{x}\frac{d}{dx}.$$

It is a formally self-adjoint operator in $L^2(\mathbb{R}_+, dm_\lambda)$, where $dm_\lambda(x) = x^{2\lambda}dx$. For any $x \in \mathbb{R}_+$ and $r > 0$, let $I(x,r) = (x-r, x+r) \cap \mathbb{R}_+$. Moreover, we assume that $r \le x$ without loss of generality. Observe that for any $x \in \mathbb{R}_+$ and $r \in (0, x]$, $m_\lambda(I(x,r)) \sim x^{2\lambda} r$. Thus, $(\mathbb{R}_+, |\cdot|, dm_\lambda)$ is a space of homogeneous type.

The Bessel Riesz transform is defined as $R_\lambda = \frac{d}{dx}(\Delta_\lambda)^{-\frac{1}{2}}$. In [39], Muckenhoupt–Stein introduced and obtained the $L^p(\mathbb{R}_+, dm_\lambda)$-boundedness of R_λ for $\lambda \in (0, \infty)$. Under this condition, the unweighted version of commutator theorem for R_λ was obtained in [16] via weak factorisation. However, the two weight commutator and high-order commutator are unknown, and the case when $\lambda \in (-1/2, 0)$ is totally unknown. Here, we will establish the two weight commutator and high order commutator for R_λ for all $\lambda \in (-1/2, \infty)$.

Proposition 7.5 *Theorem 1.6 holds for the Bessel Riesz transform R_λ with the underlying setting $(\mathbb{R}_+, |\cdot|, dm_\lambda)$.*

Proof In [2], Betancor et al. further considered R_λ for the range $\lambda \in (-1/2, \infty)$. They showed that for $f \in C_c^\infty(\mathbb{R}_+)$ and $x \in (0, \infty)$,

$$R_\lambda f(x) = \text{p.v.} \int_0^\infty R_\lambda(x,y) f(y) dm_\lambda(y)$$

with the kernel

$$R_\lambda(x,y) = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{\partial}{\partial x} W_t^\lambda(x,y) \frac{dt}{\sqrt{t}}$$

for $x, y \in (0, \infty)$ with $x \ne y$. Here $W_t^\lambda(x,y)$ is the heat kernel associated to Δ_λ

$$W_t^\lambda(x,y) = \frac{(xy)^{-\lambda+1/2}}{2t} e^{-(x^2+y^2)/4t} I_{\lambda-1/2}\left(\frac{xy}{2t}\right) \tag{7.2}$$

with I_ν being the modified Bessel function of the first kind and order $\nu > -1$. They also showed that R_λ is bounded on the space $L^p(\mathbb{R}_+, x^\delta dx)$ if and only if $p > 1$ and $-1 - p < \delta < (2\lambda + 1)p - 1$. Moreover, the kernel $R_\lambda(x,y)$ has the following estimates (see [2, Lemmas 4.3 and 4.4]):

(i) for $x/2 < y < 2x$ and $x \ne y$,

$$R_\lambda(x,y) = \frac{1}{\pi}\frac{(xy)^{-\lambda}}{y-x} + \mathcal{O}\left(y^{-2\lambda-1}\left(1 + \log\frac{xy}{(y-x)^2}\right)\right);$$

(ii) in the off-diagonal region,

$$|R_\lambda(x,y)| \lesssim \begin{cases} x^{-2\lambda-1}, & y \le x/2; \\ xy^{-2\lambda-2}, & 2x \le y. \end{cases}$$

From this fact, one deduces that $R_\lambda(x, y)$ satisfies (1.4) and (1.5) (see [4, Theorem 2.2]). Moreover, there exist $K_1 \in (0, 1/2)$ small enough, $K_2 > 1$ and $C_\lambda > 0$ such that

(i) for any $x,\ y \in \mathbb{R}_+$ with $0 < y/x - 1 < K_1$,

$$R_\lambda(x,y) \ge C_\lambda \frac{1}{x^\lambda y^\lambda} \frac{1}{y-x}; \tag{7.3}$$

(ii) for any $x,\ y \in \mathbb{R}_+$ with $0 < K_2 x \le y$,

$$R_\lambda(x,y) \ge C_\lambda x y^{-2\lambda-2}. \tag{7.4}$$

Then, an argument involving (7.3) and (7.4) shows that assumption (5.1) holds (see also [36, Lemma 2.3]). In fact, let $I := I(x_0, r)$ with $x_0 \ge r$ and $K_0 := (K_1+K_2+2)/2K_1$. We consider the following two cases.

Case (a): $x_0 \le 2K_0 r$. In this case, $m_\lambda(I) \sim x_0^{2\lambda} r \sim K_0 x_0^{2\lambda+1}$. Let $y_0 := x_0 + 4K_0^2 r$. Then $(2K_0+1)x_0 \le y_0 \le (4K_0^2+1)x_0$. This via (7.4) implies that

$$R_\lambda(x_0, y_0) \gtrsim \frac{x_0}{y_0^{2\lambda+2}} \sim \frac{1}{m_\lambda(I)}.$$

Case (b): $x_0 > 2K_0 r$. In this case, $m_\lambda(I) \sim x_0^{2\lambda} r$. Let $y_0 := x_0 + K_2 r$. Then $0 < y_0/x_0 - 1 < K_1$ and

$$R_\lambda(x_0, y_0) \gtrsim \frac{1}{x_0^{2\lambda}(y_0 - x_0)} \sim \frac{1}{m_\lambda(I)},$$

which implies that Theorem 1.6 holds for R_λ. □

7.6 Riesz Transforms Associated with Bessel Operators on $\mathbb{R}_+^{n+1}$

We now recall the Bessel operator and the Bessel Riesz transform in high dimension from Huber [23]. Consider $\mathbb{R}_+^{n+1} = \mathbb{R}^n \times (0, \infty)$. For $\lambda > -\frac{1}{2}$,

$$\Delta_\lambda^{(n+1)} = -\frac{d^2}{dx_1^2} \cdots - \frac{d^2}{dx_n^2} - \frac{d^2}{dx_{n+1}^2} - \frac{2\lambda}{x_{n+1}} \frac{d}{dx_{n+1}}. \tag{7.5}$$

The operator $\Delta_\lambda^{(n+1)}$ is symmetric and non-negative in $C_c^\infty(\mathbb{R}_+^{n+1}) \subset L^2(\mathbb{R}^{n+1}, d\mu_\lambda)$, where

$$d\mu_\lambda(x) := \prod_{j=1}^{n} dx_j x_{n+1}^{2\lambda} dx_{n+1}.$$

The jth Riesz transform is defined as

$$R_{\lambda,j} = \frac{d}{dx_j}(\Delta_\lambda^{(n+1)})^{-\frac{1}{2}}, \quad j = 1, \ldots, n+1.$$

We point out that there is no known results for the commutator of $R_{\lambda,j}$. Here we provide an intensive study of the kernel of $R_{\lambda,j}$, especially for the lower bound, and then we obtain the two weight commutator and higher order commutator for $R_{\lambda,j}$.

Proposition 7.6 *Theorem 1.6 holds for the Bessel Riesz transform $R_{\lambda,j}$, $j = 1, \ldots, n+1$, with the underlying setting $(\mathbb{R}_+^{n+1}, |\cdot|, d\mu_\lambda)$.*

Proof To begin with, note that (7.5) can be written as $\Delta_\lambda^{(n+1)} = \Delta^{(n)} + \Delta_\lambda$, where $\Delta^{(n)}$ denotes the standard Laplacian on $\mathbb{R}^n$, and Δ_λ denotes the Bessel operator on $\mathbb{R}_+$, which is one-dimensional as shown in Sect. 7.5. Then, it is clear that $e^{-t\Delta_\lambda^{(n+1)}} = e^{-t(\Delta^{(n)}+\Delta_\lambda)}$ and hence the heat kernel

$$p_{t,\Delta_\lambda^{(n+1)}}(x, y) = p_{t,\Delta_\lambda^{(n)}}(x', y') W_t^\lambda(x_{n+1}, y_{n+1})$$

for $x = (x', x_{n+1})$, $y = (y', y_{n+1}) \in \mathbb{R}^n \times (0, \infty)$, where W_t^λ is the heat kernel of Δ_λ as in (7.2).

Then, it is direct that for $1 \le j \le n$:

$$\begin{aligned} R_{\lambda,j}(x, y) &= c_{n,\lambda} \frac{\partial}{\partial x_j} \int_0^\infty p_{t,\Delta_\lambda^{(n+1)}}(x, y) \frac{dt}{\sqrt{t}} \\ &= c_{n,\lambda} \frac{\partial}{\partial x_j} \int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x'-y'|^2}{4t}} W_t^\lambda(x_{n+1}, y_{n+1}) \frac{dt}{\sqrt{t}} \end{aligned}$$

and for $j = n+1$,

$$\begin{aligned} R_{\lambda,n+1}(x, y) &= c_{n,\lambda} \frac{\partial}{\partial x_{n+1}} \int_0^\infty p_{t,\Delta_\lambda^{(n+1)}}(x, y) \frac{dt}{\sqrt{t}} \\ &= c_{n,\lambda} \frac{\partial}{\partial x_{n+1}} \int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x'-y'|^2}{4t}} W_t^\lambda(x_{n+1}, y_{n+1}) \frac{dt}{\sqrt{t}}. \end{aligned}$$

By [4, Theorem 2.2], $\{R_{\lambda,j}\}_{j=1}^{n+1}$ are Calderón–Zygumnd operators with kernel satisfying (1.4) and (1.5).

Moreover, we have the following estimates on $\{R_{\lambda,j}(x, y)\}_{j=1}^{n+1}$.

Lemma 7.7 *Let $j \in \{1, \dots, n\}$. The following statements hold:*

(i) *There exist positive constants $\widetilde{C}, \widetilde{c} > 1$ such that for any (x, y) with $0 < x_{n+1} \leq y_{n+1}/\widetilde{c}$, $R_{\lambda, j}(x, y)$ does not change sign and*

$$|R_{\lambda, j}(x, y)| \geq \widetilde{C} \frac{|x_j - y_j|}{(y_{n+1}^2 + |x' - y'|^2)^{\frac{n}{2}+\lambda+1}}.$$

(ii) *There exists a positive constant C_n such that for any (x, y),*

$$R_{\lambda, j}(x, y) = C_n \frac{y_j - x_j}{x_{n+1}^{\lambda} y_{n+1}^{\lambda} |x - y|^{n+2}} + \mathcal{O}\left(\frac{1}{x_{n+1}^{\lambda+1} y_{n+1}^{\lambda+1}} \frac{|x_j - y_j|}{|x - y|^n} \right).$$

Proof By (7.2) and letting $x_{n+1} = z y_{n+1}$ and $u = \frac{2t}{y_{n+1}^2}$, we see that

$$\begin{aligned}
&R_{\lambda,j}(x, y) \\
&= c_{n,\lambda} \int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x'-y'|^2}{4t}} \frac{x_j - y_j}{-2t} \frac{(x_{n+1} y_{n+1})^{-\lambda+\frac{1}{2}}}{2t} I_{\lambda-\frac{1}{2}} \\
&\times \left(\frac{x_{n+1} y_{n+1}}{2t} \right) e^{-\frac{x_{n+1}^2 + y_{n+1}^2}{4t}} \frac{dt}{\sqrt{t}} \\
&= c_{n,\lambda} \int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x'-y'|^2}{4t}} \frac{x_j - y_j}{-2t} \left(\frac{y_{n+1}^2 z}{2t} \right)^{-\lambda+\frac{1}{2}} \\
&\times I_{\lambda-\frac{1}{2}} \left(\frac{y_{n+1}^2 z}{2t} \right) (2t)^{-\lambda} e^{-\frac{y_{n+1}^2 (1+z^2)}{4t}} \frac{dt}{\sqrt{t}} \\
&= c_{n,\lambda} (y_j - x_j) \int_0^\infty (u y_{n+1}^2)^{-\frac{n}{2}-\lambda-1} e^{-\frac{|x'-y'|^2}{2u y_{n+1}^2}} e^{-\frac{1+z^2}{2u}} \left(\frac{z}{u} \right)^{-\lambda+\frac{1}{2}} I_{\lambda-\frac{1}{2}} \left(\frac{z}{u} \right) \frac{du}{u}.
\end{aligned}$$

Recall that for any $z > 0$ and $\nu > -1$,

$$\lim_{z \to 0^+} z^{-\nu} I_\nu(z) = \frac{1}{2^\nu \Gamma(\nu + 1)}, \tag{7.6}$$

and for any $z > 0$, $\nu > -1$ and $n = 0, 1, 2, \dots$,

$$I_\nu(z) = \frac{e^z}{\sqrt{2\pi z}} \left(\sum_{k=0}^n (-1)^k [\nu, k] (2z)^{-k} + \mathcal{O}(z^{-n-1}) \right), \tag{7.7}$$

where $[\nu, 0] := 1$ and for $k \in \mathbb{N}$,

$$[\nu, k] := \frac{(4\nu^2 - 1)(4\nu^2 - 3^2) \cdots (4\nu^2 - (2k - 1)^2)}{2^{2k} \Gamma(k + 1)};$$

see [2, p. 109].

Then, letting $z \to 0$ and applying the Lebesgue Dominated Convergence Theorem, we have

$$R_{\lambda,j}(x,y) \to c_{n,\lambda}(x_j - y_j)y_{n+1}^{-n-2\lambda-2}\int_0^\infty u^{-\frac{n}{2}-\lambda-2}e^{-\frac{1}{2u}(\frac{|x'-y'|^2}{y_{n+1}^2}+1)}du$$
$$= c\frac{x_j - y_j}{(y_{n+1}^2 + |x'-y'|^2)^{\frac{n}{2}+1+\lambda}}.$$

This shows (i).

Now, for $j = 1, 2, \ldots, n$, let

$$R_j(x,y) := c_{n,\lambda}\int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x'-y'|^2}{4t}}\frac{x_j - y_j}{-2t}\frac{1}{x_{n+1}^\lambda y_{n+1}^\lambda}W_t(x_{n+1}, y_{n+1})\frac{dt}{\sqrt{t}},$$

where

$$W_t(x_{n+1}, y_{n+1}) := \frac{1}{\sqrt{4\pi t}}e^{-\frac{(x_{n+1}-y_{n+1})^2}{4t}}.$$

Then, we see that for any $x, y \in \mathbb{R}^n$ such that $x_j \neq y_j$,

$$R_j(x,y) = C_n\frac{y_j - x_j}{x_{n+1}^\lambda y_{n+1}^\lambda |x-y|^{n+2}}.$$

On the other hand, using (7.7) for $n = 0$, we conclude that

$$\begin{aligned}
&|R_{\lambda,j}(x,y) - R_j(x,y)|\\
&= c_{n,\lambda}\left|\int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x'-y'|^2}{4t}}\frac{x_j - y_j}{-2t}\right.\\
&\quad\times\left[\frac{(x_{n+1}y_{n+1})^{-\lambda+\frac{1}{2}}}{2t}\frac{e^{\frac{x_{n+1}y_{n+1}}{2t}}}{\sqrt{2\pi\cdot\frac{x_{n+1}y_{n+1}}{2t}}}\left(1 + C\frac{2t}{x_{n+1}y_{n+1}}\right)e^{-\frac{x_{n+1}^2+y_{n+1}^2}{4t}}\right.\\
&\quad\left.\left.-\frac{1}{x_{n+1}^\lambda y_{n+1}^\lambda}\frac{1}{\sqrt{4\pi t}}e^{-\frac{(x_{n+1}-y_{n+1})^2}{4t}}\right]\frac{dt}{\sqrt{t}}\right|\\
&\lesssim \int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x-y|^2}{4t}}\frac{|x_j - y_j|}{x_{n+1}^{\lambda+1}y_{n+1}^{\lambda+1}}\frac{dt}{t}\\
&\lesssim \frac{|x_j - y_j|}{x_{n+1}^{\lambda+1}y_{n+1}^{\lambda+1}}\int_0^\infty \frac{1}{t^{\frac{n}{2}+1}}e^{-\frac{|x-y|^2}{4t}}dt\\
&\lesssim \frac{|x_j - y_j|}{x_{n+1}^{\lambda+1}y_{n+1}^{\lambda+1}}\frac{1}{|x-y|^n}.
\end{aligned}$$

The proof of Lemma 7.7 is complete. □

Similarly, for $R_{\lambda,\,n+1}(x,y)$, we have show the following lemma.

Lemma 7.8 *The following statements hold:*

(i) *There exist positive constants* $\widetilde{C}, \widetilde{c} \ge 1$ *such that for any* (x, y) *with* $0 < x_{n+1} \le y_{n+1}/\widetilde{c}$ *and* $\frac{\lambda+\frac{1}{2}}{\lambda+1} < \frac{y_{n+1}^2}{y_{n+1}^2+|x'-y'|^2} < 1$, $R_{\lambda,\,n+1}(x,y)$ *does not change sign and*

$$|R_{\lambda,\,n+1}(x,y)| \ge \widetilde{C}\frac{x_{n+1}}{(y_{n+1}^2+|x'-y'|^2)^{\frac{n}{2}+\lambda+1}}.$$

(ii) *There exists a positive constant* C_n *such that for any* (x, y),

$$R_{\lambda,\,n+1}(x,y) = C_n\frac{y_{n+1}-x_{n+1}}{(x_{n+1}y_{n+1})^{\lambda}}\frac{1}{|x-y|^{n+2}} + \mathcal{O}\left(\frac{1}{x_{n+1}^{\lambda}y_{n+1}^{\lambda+1}}\frac{1}{|x-y|^{n}}\right).$$

Proof Observe that

$$\begin{aligned}
&\frac{\partial}{\partial x_{n+1}} W_t^{\lambda}(x_{n+1}, y_{n+1})\\
&= \frac{1}{(2t)^{\lambda+\frac{1}{2}}}\left[x_{n+1}\left(\frac{y_{n+1}}{2t}\right)^2\left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda-\frac{1}{2}} I_{\lambda+\frac{1}{2}}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)\right.\\
&\quad\left.-\frac{x_{n+1}}{2t}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda+\frac{1}{2}} I_{\lambda-\frac{1}{2}}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)\right] e^{-\frac{x_{n+1}^2+y_{n+1}^2}{4t}};
\end{aligned}$$

see, [2]. Then we have

$$\begin{aligned}
R_{\lambda,\,n+1}(x,y) &= c_{n,\lambda}\int_0^{\infty}\frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x'-y'|^2}{4t}}\frac{1}{(2t)^{\lambda+\frac{1}{2}}}x_{n+1}\left(\frac{y_{n+1}}{2t}\right)^2\left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda-\frac{1}{2}}\\
&\quad\times I_{\lambda+\frac{1}{2}}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)e^{-\frac{x_{n+1}^2+y_{n+1}^2}{4t}}\frac{dt}{\sqrt{t}}\\
&\quad- c_{n,\lambda}\int_0^{\infty}\frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x'-y'|^2}{4t}}\frac{1}{(2t)^{\lambda+\frac{1}{2}}}\frac{x_{n+1}}{2t}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda+\frac{1}{2}}\\
&\quad\times I_{\lambda-\frac{1}{2}}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)e^{-\frac{x_{n+1}^2+y_{n+1}^2}{4t}}\frac{dt}{\sqrt{t}}.
\end{aligned}$$

By change of variables, we have

$$\begin{aligned}
R_{\lambda,\,n+1}(x,y) &= c_{n,\lambda}\left[x_{n+1}\int_0^{\infty}\frac{1}{(4\pi t)^{\frac{n}{2}}}e^{-\frac{|x'-y'|^2}{4t}}\frac{1}{(2t)^{\lambda+\frac{1}{2}}}\left(\frac{y_{n+1}}{2t}\right)^2\left(\frac{y_{n+1}^2 z}{2t}\right)^{-\lambda-\frac{1}{2}}\right.\\
&\quad\times I_{\lambda+\frac{1}{2}}\left(\frac{y_{n+1}^2 z}{2t}\right)e^{-\frac{y_{n+1}^2(1+z^2)}{4t}}\frac{dt}{\sqrt{t}}
\end{aligned}$$

$$- x_{n+1}\int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x'-y'|^2}{4t}} \frac{1}{(2t)^{\lambda+\frac{1}{2}}}\frac{1}{2t}\left(\frac{y_{n+1}^2 z}{2t}\right)^{-\lambda+\frac{1}{2}}$$

$$\times I_{\lambda-\frac{1}{2}}\left(\frac{y_{n+1}^2 z}{2t}\right) e^{-\frac{y_{n+1}^2(1+z^2)}{4t}}\frac{dt}{\sqrt{t}}\Bigg]$$

$$= c_{n,\lambda}\left[\frac{x_{n+1}}{y_{n+1}^{n+2+2\lambda}}\int_0^\infty \frac{1}{u^{\frac{n}{2}+\lambda+2}} e^{-\frac{|x'-y'|^2}{2uy_{n+1}^2}}\left(\frac{z}{u}\right)^{-\lambda-\frac{1}{2}} I_{\lambda+\frac{1}{2}}\left(\frac{z}{u}\right) e^{-\frac{1+z^2}{2u}}\frac{du}{u}\right.$$

$$\left. - \frac{x_{n+1}}{y_{n+1}^{n+2\lambda+2}}\int_0^\infty \frac{1}{u^{\frac{n}{2}+\lambda+1}} e^{-\frac{|x'-y'|^2}{2uy_{n+1}^2}}\left(\frac{z}{u}\right)^{-\lambda+\frac{1}{2}} I_{\lambda-\frac{1}{2}}\left(\frac{z}{u}\right) e^{-\frac{1+z^2}{2u}}\frac{du}{u}\right].$$

By letting $z \to 0$ and applying (7.6), we see that

$$R_{\lambda,n+1}(x,y)\frac{y_{n+1}^{n+2\lambda+2}}{x_{n+1}} \to C_{n,\lambda}\left[\frac{1}{2^{\lambda+\frac{1}{2}}\Gamma(\lambda+\frac{3}{2})}\int_0^\infty \frac{1}{u^{\frac{n}{2}+\lambda+2}} e^{-\frac{1}{2u}(1+\frac{|x'-y'|^2}{y_{n+1}^2})}\frac{du}{u}\right.$$

$$\left. - \frac{1}{2^{\lambda-\frac{1}{2}}\Gamma(\lambda+\frac{1}{2})}\int_0^\infty \frac{1}{u^{\frac{n}{2}+\lambda+1}} e^{-\frac{1}{2u}(1+\frac{|x'-y'|^2}{y_{n+1}^2})}\frac{du}{u}\right]$$

$$= C_{n,\lambda} 2^{\frac{n+3}{2}}\frac{y_{n+1}^{n+2\lambda+2}}{(y_{n+1}^2+|x'-y'|^2)^{\frac{n}{2}+\lambda+1}}\frac{\Gamma(\lambda+\frac{n}{2}-1)}{\Gamma(\lambda+\frac{1}{2})}$$

$$\times\left(\frac{y_{n+1}^2}{y_{n+1}^2+|x'-y'|^2}\frac{\lambda+\frac{n}{2}-1}{\lambda+\frac{1}{2}}-1\right)$$

$$= C_{n,\lambda}\frac{y_{n+1}^{n+2\lambda+2}}{(y_{n+1}^2+|x'-y'|^2)^{\frac{n}{2}+\lambda+1}}$$

$$\times\left(\frac{y_{n+1}^2}{y_{n+1}^2+|x'-y'|^2}\frac{\lambda+\frac{n}{2}+1}{\lambda+\frac{1}{2}}-1\right).$$

Since

$$\frac{y_{n+1}^2}{y_{n+1}^2+|x'-y'|^2}\frac{\lambda+\frac{n}{2}+1}{\lambda+\frac{1}{2}}-1>0,$$

the conclusion (i) holds.

Let

$$\mathrm{H}(x,y) := c_{n,\lambda}\int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x-y|^2}{4t}}\frac{1}{(2t)^{\lambda+\frac{1}{2}}}\left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda}$$

$$\times\left[x_{n+1}\left(\frac{y_{n+1}}{2t}\right)^2\left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-1} - \frac{x_{n+1}}{2t}\right]\frac{dt}{\sqrt{t}}.$$

Then, observe that

$$\int_0^\infty \frac{1}{(4\pi t)^{\frac{n}{2}}} e^{-\frac{|x-y|^2}{4t}} \frac{1}{(2t)^{\lambda+\frac{1}{2}}} \left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda}$$
$$\times \left[x_{n+1} \left(\frac{y_{n+1}}{2t}\right)^2 \left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-1} - \frac{x_{n+1}}{2t} \right] \frac{dt}{\sqrt{t}}$$
$$= C_n \frac{y_{n+1}-x_{n+1}}{(x_{n+1}y_{n+1})^\lambda} \int_0^\infty \frac{1}{t^{\frac{n}{2}+2}} e^{-\frac{|x-y|^2}{4t}} dt$$
$$= C_n \frac{y_{n+1}-x_{n+1}}{(x_{n+1}y_{n+1})^\lambda} \frac{1}{|x-y|^{n+2}}.$$

By using (7.7) for $n = 0$,

$$\left|R_{\lambda,\, n+1}(x, y) - \mathrm{H}(x, y)\right| \lesssim \int_0^\infty \frac{1}{t^{\frac{n}{2}+\lambda+1}} e^{-\frac{|x-y|^2}{4t}} \left(\frac{x_{n+1}y_{n+1}}{2t}\right)^{-\lambda} \frac{1}{y_{n+1}} dt$$
$$\lesssim \frac{1}{x_{n+1}^\lambda y_{n+1}^{\lambda+1}} \frac{1}{|x-y|^n}.$$

We then see that (ii) holds. The proof of Lemma 7.8 is complete. □

We now return to the proof of Proposition 7.6. Based on Lemmas 7.7 and 7.8, we see that (5.1) holds. Indeed, For any $x := (x_1, \ldots, x_{n+1}) \in \mathbb{R}_+^{n+1}$ and $r \in (0, \infty)$, let

$$Q(x, r) := \{y := (y_1, \ldots, y_{n+1}) \in \mathbb{R}_+^{n+1} : |x_j - y_j| \le r/2,\ j \in \{1, \ldots, n+1\}\}.$$

Then, we have $\mu_\lambda(Q(x, r)) \sim r^{n+1} x_{n+1}^{2\lambda}$. Let $C_0 \gg \tilde{c}$. For any $x := (x_1, \ldots, x_{n+1}) \in \mathbb{R}_+^{n+1}$ and $r \in (0, \infty)$, if $r > \frac{\tilde{c}-1}{C_0} x_{n+1}$, take $y := (y_1, \ldots, y_{n+1})$ such that $y_i = x_i + C_0 r$ for $i = j$ or $i = n+1$ and $y_i := x_i$ otherwise. Then by Lemma 7.7(i), we see that $y_{n+1} \ge \tilde{c} x_{n+1}$ and

$$|R_{\lambda,\, j}(x, y)| \gtrsim \frac{C_0 r}{[y_{n+1}^2 + (C_0 r)^2]^{\frac{n}{2}+\lambda+1}} \gtrsim \frac{1}{\mu_\lambda(Q(x, r))}.$$

If $r \le \frac{\tilde{c}-1}{C_0} x_{n+1}$, then there exists $y \in \mathbb{R}_+^{n+1}$ such that $|y_{n+1} - x_{n+1}| = |y_j - x_j| \sim |y - x|$ and $|y_{n+1} - x_{n+1}| \ll x_{n+1}$. Then by Lemma 7.7(ii), we also have

$$|R_{\lambda,\, j}(x, y)| \gtrsim \frac{|y_j - x_j|}{x_{n+1}^\lambda y_{n+1}^\lambda |x-y|^{n+2}} \sim \frac{1}{\mu_\lambda(Q(x, r))}.$$

Therefore, (1.6) holds for $R_{\lambda,\, j}(x, y)$, $j \in \{1, \ldots, n\}$. The argument for $R_{\lambda,\, n+1}(x, y)$ is similar and omitted. Then Theorem 1.6 holds for the Bessel Riesz transform $R_{\lambda, j}$, $j = 1, \ldots, n+1$.

The proof of Proposition 7.6 is complete. □

8 A Digression to Product Setting: Little bmo Space

In this section, we consider the weighted little bmo space on product spaces of homogeneous type. To begin with, let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be two copies of spaces of homogeneous type as stated in Sect. 2, and denote $\vec{X} := X_1 \times X_2$, $\vec{\mu} := \mu_1 \times \mu_2$. Moreover, for the points in $\vec{X}$, we denote $\vec{x} := (x_1, x_2) \in \vec{X}$.

Macías and Segovia [35] proved the following fundamental result on spaces of homogeneous type. Suppose that (X, d) is a space endowed with a quasi-metric d that may have no regularity. Then there exists a quasi-metric d' that is pointwise equivalent to d such that $d(x, y) \sim d'(x, y)$ for all $x, y \in X$ and there exist constants $\theta \in (0, 1)$ and $C > 0$ so that d' has the following regularity:

$$|d'(x, y) - d'(x', y)| \leq C\, d'(x, x')^{\theta}\, [d'(x, y) + d'(x', y)]^{1-\theta}$$

for all $x, x', y \in X$. Moreover, if the quasi-metric balls are defined by this new quasi-metric d', that is, $B'(x, r) := \{y \in X : d'(x, y) < r\}$ for $r > 0$, then these balls are open in the topology induced by d'. See [35, Theorem 2, p.259]. So, without lost of generality, we assume that in our product setting, the quasi-metrics d_1 and d_2 have regularity with constants θ_1 and θ_2, respectively.

We now recall the product $A_p(\vec{X})$ weights on product spaces of homogeneous type.

Definition 8.1 Let $w(x_1, x_2)$ be a nonnegative locally integrable function on $\vec{X}$. For $1 < p < \infty$, we say w is a product A_p *weight*, written as $w \in A_p(\vec{X})$, if

$$[w]_{A_p(\vec{X})} := \sup_R \left(\fint_R w \right) \left(\fint_R \left(\frac{1}{w} \right)^{1/(p-1)} \right)^{p-1} < \infty.$$

Here, the supremum is taken over all "rectangles" $R := B_1 \times B_2 \subset \vec{X}$, where B_i are balls in X_i for $i = 1, 2$. The quantity $[w]_{A_p(\vec{X})}$ is called the A_p *constant of* w.

Next, we recall the weighted little bmo space on product spaces of homogeneous type.

Definition 8.2 For $1 < p < \infty$ and $w \in A_p(\vec{X})$, the weighted little bmo space $\mathrm{bmo}_w(\vec{X})$ is the space of all locally integrable functions b on $\vec{X}$ such that

$$\|b\|_{\mathrm{bmo}_w(\vec{X})} = \sup_R \frac{1}{w(R)} \int_R |b(\vec{x}) - b_R| d\vec{\mu}(\vec{x}) < \infty,$$

where the supremum is taken over all "rectangles" $R = B_1 \times B_2 \subset \vec{X}$, where B_i are balls in X_i for $i = 1, 2$.

Similar to [22, Section 7.1], we introduce the bi-parameter Journé operator on $\vec{X}$ as follows. Let $C_0^{\eta_1}(X_1)$, $\eta_1 \in (0, \theta_1]$, denote the space of continuous functions f with bounded support such that

$$\|f\|_{C_0^{\eta_1}(X_1)} := \sup_{x,y\in X_1, x\neq y} \frac{|f(x)-f(y)|}{d_1(x,y)^{\eta_1}} < \infty.$$

Let $C_0^{\eta_2}(X_2)$, $\eta_2 \in (0, \theta_2]$ be defined similarly.

I Structural Assumptions: Given $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$, where $f_i, g_i : X_i \to \mathbb{C}$, $f_i, g_i \in C_0^{\eta_i}(X_i)$ satisfy $\operatorname{supp} f_i \cap \operatorname{supp} g_i = \emptyset$ for $i = 1, 2$, we assume the kernel representation

$$\langle Tf, g\rangle = \int_{\vec{X}} \int_{\vec{X}} K(\vec{x}, \vec{y}) f(\vec{y}) g(\vec{x}) d\vec{\mu}(\vec{y}) d\vec{\mu}(\vec{x}).$$

The kernel $K : \vec{X} \times \vec{X} \setminus \{(\vec{x}, \vec{y}) \in \vec{X} \times \vec{X} : \ x_1 = y_1, \ \text{or} \ x_2 = y_2\} \to \mathbb{C}$ is assumed to satisfy:

1. Size condition:

$$|K(\vec{x}, \vec{y})| \leq C \frac{1}{\mu_1(B(x_1, d_1(x_1, y_1)))\mu_2(B(x_2, d_2(x_2, y_2)))}.$$

2. Hölder conditions:

2a. if $d_1(y_1, y_1') \leq \frac{1}{2A_0} d_1(x_1, y_1)$ and $d_2(y_2, y_2') \leq \frac{1}{2A_0} d_2(x_2, y_2)$:

$$\begin{aligned}&|K(\vec{x}, \vec{y}) - K(\vec{x}, (y_1, y_2')) - K(\vec{x}, (y_1', y_2)) + K(\vec{x}, \vec{y}')|\\ &\leq C \frac{d_1(y_1, y_1')^\delta d_2(y_2, y_2')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))d_1(x_1, y_1)^\delta \mu_2(B(x_2, d_2(x_2, y_2)))d_2(x_2, y_2)^\delta}.\end{aligned}$$

2b. if $d_1(x_1, x_1') \leq \frac{1}{2A_0} d_1(x_1, y_1)$ and $d_2(x_2, x_2') \leq \frac{1}{2A_0} d_2(x_2, y_2)$:

$$\begin{aligned}&|K(\vec{x}, \vec{y}) - K((x_1, x_2'), \vec{y}) - K((x_1', x_2), \vec{y}) + K(\vec{x}', \vec{y})|\\ &\leq C \frac{d_1(x_1, x_1')^\delta d_2(x_2, x_2')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))d_1(x_1, y_1)^\delta \mu_2(B(x_2, d_2(x_2, y_2)))d_2(x_2, y_2)^\delta}.\end{aligned}$$

2c. if $d_1(y_1, y_1') \leq \frac{1}{2A_0} d_1(x_1, y_1)$ and $d_2(x_2, x_2') \leq \frac{1}{2A_0} d_2(x_2, y_2)$:

$$\begin{aligned}&|K(\vec{x}, \vec{y}) - K((x_1, x_2'), \vec{y}) - K(\vec{x}, (y_1', y_2)) + K((x_1, x_2'), (y_1', y_2))|\\ &\leq C \frac{d_1(y_1, y_1')^\delta d_2(x_2, x_2')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))d_1(x_1, y_1)^\delta \mu_2(B(x_2, d_2(x_2, y_2)))d_2(x_2, y_2)^\delta}.\end{aligned}$$

2d. if $d_1(x_1, x_1') \leq \frac{1}{2A_0} d_1(x_1, y_1)$ and $d_2(y_2, y_2') \leq \frac{1}{2A_0} d_2(x_2, y_2)$:

$$\begin{aligned}&|K(\vec{x}, \vec{y}) - K(\vec{x}, (y_1, y_2')) - K((x_1', x_2), \vec{y}) + K((x_1', x_2), (y_1, y_2'))|\\ &\leq C \frac{d_1(x_1, x_1')^\delta d_2(y_2, y_2')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))d_1(x_1, y_1)^\delta \mu_2(B(x_2, d_2(x_2, y_2)))d_2(x_2, y_2)^\delta}.\end{aligned}$$

3. Mixed size and Hölder conditions:

3a. if $d_1(x_1, x_1') \le \frac{1}{2A_0} d_1(x_1, y_1)$:

$$|K(\vec{x}, \vec{y}) - K((x_1', x_2), \vec{y})| \le C \frac{d_1(x_1, x_1')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))d_1(x_1, y_1)^\delta \mu_2(B(x_2, d_2(x_2, y_2)))}.$$

3b. if $d_1(y_1, y_1') \le \frac{1}{2A_0} d_1(x_1, y_1)$:

$$|K(\vec{x}, \vec{y}) - K(\vec{x}, (y_1', y_2))| \le C \frac{d_1(y_1, y_1')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))d_1(x_1, y_1)^\delta \mu_2(B(x_2, d_2(x_2, y_2)))}.$$

3c. if $d_2(x_2, x_2') \le \frac{1}{2A_0} d_2(x_2, y_2)$:

$$|K(\vec{x}, \vec{y}) - K((x_1, x_2'), \vec{y})| \le C \frac{d_2(x_2, x_2')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))\mu_2(B(x_2, d_2(x_2, y_2)))d_2(x_2, y_2)^\delta}.$$

3d. if $d_2(y_2, y_2') \le \frac{1}{2A_0} d_2(x_2, y_2)$:

$$|K(\vec{x}, \vec{y}) - K(\vec{x}, (y_1, y_2'))| \le C \frac{d_2(y_2, y_2')^\delta}{\mu_1(B(x_1, d_1(x_1, y_1)))\mu_2(B(x_2, d_2(x_2, y_2)))d_2(x_2, y_2)^\delta}.$$

4. Calderón–Zygmund structure in X_1 and X_2 separately: If $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ with $\mathrm{supp} f_1 \cap \mathrm{supp} g_1 = \emptyset$, we assume the kernel representation:

$$\langle Tf, g\rangle = \int_{X_1}\int_{X_1} K_{f_2,g_2}(x_1, y_1) f_1(y_1) g_1(x_1) d\mu_1(x_1) d\mu_1(y_1),$$

where the kernel $K_{f_2,g_2} : X_1 \times X_1 \backslash \{(x_1, y_1) \in X_1 \times X_1 : x_1 = y_1\}$ satisfies the following size condition:

$$|K_{f_2,g_2}(x_1, y_1)| \le C(f_2, g_2) \frac{1}{\mu_1(B(x_1, d_1(x_1, y_1)))}$$

and Hölder conditions:

$$\begin{aligned}
&|K_{f_2,g_2}(x_1,y_1)-K_{f_2,g_2}(x_1',y_1)| \\
&\quad \le \frac{C(f_2,g_2)\, d_1(x_1,x_1')^\delta}{\mu_1(B(x_1,d_1(x_1,y_1)))d_1(x_1,y_1)^\delta},\ d_1(x_1,x_1')\le \frac{1}{2A_0}d_1(x_1,y_1),\\
&|K_{f_2,g_2}(x_1,y_1)-K_{f_2,g_2}(x_1,y_1')| \\
&\quad \le \frac{C(f_2,g_2) d_1(y_1,y_1')^\delta}{\mu_1(B(x_1,d_1(x_1,y_1)))d_1(x_1,y_1)^\delta},\ d_1(x_1,x_1')\le \frac{1}{2A_0}d_1(x_1,y_1).
\end{aligned}$$

We only assume the above representation and a certain control over $C(f_2, g_2)$ on the diagonal, that is:

$$C(\chi_{Q_2},\chi_{Q_2})+C(\chi_{Q_2},u_{Q_2})+C(u_{Q_2},\chi_{Q_2})\le C\mu_2(Q_2)$$

for all cubes $Q_2 \subset X_2$ and all "Q_2-adapted zero-mean" functions $u_{Q_2}-$ that is, $\operatorname{supp} u_{Q_2} \subset Q_2$, $|u_{Q_2}| \le 1$ and $\int_{X_2} u_{Q_2}(x_2)d\mu_2(x_2) = 0$. We assume the symmetrical representation with kernel K_{f_1,g_1} in the case $\operatorname{supp} f_2 \cap \operatorname{supp} g_2 = \emptyset$.

II Boundedness and Cancellation Assumptions:

1. Assume $T1$, T^*1, $T_1 1$, and $T_1^* 1$ are in product $\mathrm{BMO}(\vec{X})$, where T_1 is the partial adjoint of T defined by $\langle T_1(f_1\otimes f_2), g_1\otimes g_2\rangle = \langle T_1(g_1\otimes f_2), (f_1\otimes g_2)\rangle$.
2. Assume $|\langle T_1(\chi_{Q_1}\otimes\chi_{Q_2}), \chi_{Q_1}\otimes\chi_{Q_2}\rangle| \le C\mu_1(Q_1)\mu_2(Q_2)$ for all cubes $Q_i \in X_i$ (weak boundedness).
3. Diagonal BMO conditions: for all cubes $Q_i \subset X_i$ and all non-zero functions a_{Q_1} and b_{Q_2} that are Q_1- and Q_2- adapted, respectively, assume:

$$\begin{aligned}
|\langle T_1(a_{Q_1}\otimes\chi_{Q_2}),\chi_{Q_1}\otimes\chi_{Q_2}\rangle| &\le C\mu_1(Q_1)\mu_2(Q_2),\\
|\langle T_1(\chi_{Q_1}\otimes\chi_{Q_2}),a_{Q_1}\otimes\chi_{Q_2}\rangle| &\le C\mu_1(Q_1)\mu_2(Q_2),\\
|\langle T_1(\chi_{Q_1}\otimes b_{Q_2}),\chi_{Q_1}\otimes\chi_{Q_2}\rangle| &\le C\mu_1(Q_1)\mu_2(Q_2),\\
|\langle T_1(\chi_{Q_1}\otimes\chi_{Q_2}),\chi_{Q_1}\otimes b_{Q_2}\rangle| &\le C\mu_1(Q_1)\mu_2(Q_2).
\end{aligned}$$

For the upper bound of the commutator of such operators T and $b \in \mathrm{bmo}_w(\vec{X})$, following the same approach as that in [22], and combining all necessary tools as recalled in Sect. 2 on spaces of homogeneous type (such as the adjacent dyadic system, Haar basis, et al), we obtain that

Theorem 8.3 *Let $1 < p < \infty$ and $\lambda_1, \lambda_2 \in A_p(\vec{X})$, and define $\nu = \lambda_1^{\frac{1}{p}}\lambda_2^{-\frac{1}{p}}$. Let T be a bi-parameter Journé operator on $\vec{X}$ and $b \in \mathrm{bmo}_\nu(\vec{X})$. Then we obtain that*

$$\|[b,T]:\ L^p_{\lambda_1}(\vec{X})\to L^p_{\lambda_2}(\vec{X})\| \lesssim \|b\|_{\mathrm{bmo}_\nu(\vec{X})}.$$

We now provide a broader version of the lower bound. Note that in [22] the authors only considered the lower bound of commutator with respect to double Riesz transforms, and their proof relies on Fourier transform and hence can not be adapted to spaces of homogeneous type.

We assume that the bi-parameter Journé operator T satisfies the following "homogeneous" condition:
there exist positive constants c_0 and $\overline{C}$ such that for every $x_1 \in X_1$, $x_2 \in X_2$ and $r_1, r_2 > 0$, there exist $y_1 \in B_1(x_1, \overline{C}r_1)\backslash B_1(x_1, r_1)$ and $y_2 \in B_2(x_2, \overline{C}r_2)\backslash B_2(x_2, r_2)$ satisfying

$$|K(x_1, y_1; x_2, y_2)| \geq \frac{1}{c_0\mu_1(B_1(x_1, r_1))\mu_2(B_2(x_2, r_2))}. \tag{8.1}$$

Then, we have the following lower bound.

Theorem 8.4 *Let T be a bi-parameter Journé operator on $\vec{X}$ and T satisfies the following "homogeneous" condition as above. Let $1 < p < \infty$ and $\lambda_1, \lambda_2 \in A_p(\vec{X})$, and define $\nu := \lambda_1^{\frac{1}{p}}\lambda_2^{-\frac{1}{p}}$. Suppose that $b \in L^1_{loc}(\vec{X})$ and that $\|[b, T] : L^p_{\lambda_1}(\vec{X}) \to L^p_{\lambda_2}(\vec{X})\| < \infty$. Then we obtain that $b \in \mathrm{bmo}_\nu(\vec{X})$ with*

$$\|b\|_{\mathrm{bmo}_\nu(\vec{X})} \lesssim \|[b, T] : L^p_{\lambda_1}(\vec{X}) \to L^p_{\lambda_2}(\vec{X})\|.$$

To see this, we first point out that the homogeneous condition (8.1) implies the following condition: there exist positive constants $3 \leq A_1 \leq A_2$ such that for any ball $B_i := B_i(x_0^{(i)}, r_i) \subset X_i$, there exist balls $\widetilde{B}_i := B_i(y_0^{(i)}, r_i)$ such that $A_1 r_i \leq d_i(x_0^{(i)}, y_0^{(i)}) \leq A_2 r_i$. Moreover, for all $(x_1, y_1; x_2, y_2) \in (B_1 \times \widetilde{B}_1) \times (B_2 \times \widetilde{B}_2)$, $K(x_1, y_1; x_2, y_2)$ does not change sign and

$$|K(x_1, y_1; x_2, y_2)| \gtrsim \frac{1}{\mu_1(B_1)}\frac{1}{\mu_2(B_2)}.$$

If $K(x_1, y_1; x_2, y_2) := K_1(x_1, y_1; x_2, y_2) + iK_2(x_1, y_1; x_2, y_2)$ is complex-valued, where $i^2 = -1$, then at least one of K_i satisfies the assumption above.

We next consider the median value on "rectangles" $R = B_1 \times B_2 \subset \vec{X}$. By a median value of a real-valued measurable function f over R we mean a possibly non-unique, real number $\alpha_R(f)$ such that $\vec{\mu}(\{(x_1, x_2) \in R : f(x_1, x_2) > \alpha_R(f)\}) \leq \frac{1}{2}\mu_1(B_1)\mu_2(B_2)$ and $\vec{\mu}(\{(x_1, x_2) \in R : f(x_1, x_2) < \alpha_R(f)\}) \leq \frac{1}{2}\mu_1(B_1)\mu_2(B_2)$.

Now, following the idea in Lemma 5.3, for the given rectangle $R = B_1 \times B_2$, $\widetilde{B}_1$ and $\widetilde{B}_2$, set

$$E_1 := \{(x_1, x_2) \in B_1 \times B_2 : b(x_1, x_2) \geq \alpha_{\widetilde{B}_1\times\widetilde{B}_2}(b)\},$$
$$E_2 := \{(x_1, x_2) \in B_1 \times B_2 : b(x_1, x_2) \leq \alpha_{\widetilde{B}_1\times\widetilde{B}_2}(b)\},$$

and

$$F_1 := \{(y_1, y_2) \in \widetilde{B}_1 \times \widetilde{B}_2 : b(y_1, y_2) \leq \alpha_{\widetilde{B}_1\times\widetilde{B}_2}(b)\},$$
$$F_2 := \{(y_1, y_2) \in \widetilde{B}_1 \times \widetilde{B}_2 : b(y_1, y_2) \geq \alpha_{\widetilde{B}_1\times\widetilde{B}_2}(b)\}.$$

Then, by the definition of $\alpha_R(f)$, we see that $\vec{\mu}(F_i) \geq \frac{1}{4}\mu_1(\widetilde{B}_1)\mu_2(\widetilde{B}_2)$ for $i = 1, 2$. Moreover, for $(x_1, x_2) \times (y_1, y_2) \in (E_1 \times F_1) \cup (E_2 \times F_2)$,

$$\begin{aligned}|b(x_1, x_2) - b(y_1, y_2)| &= \left|b(x_1, x_2) - \alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b) + \alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b) - b(y_1, y_2)\right| \\ &= \left|b(x_1, x_2) - \alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b)\right| + \left|\alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b) - b(y_1, y_2)\right| \\ &\geq \left|b(x_1, x_2) - \alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b)\right|.\end{aligned}$$

Proof of Theorem 8.4 For given $b \in L^1_{\rm loc}(\vec{X})$ and for any rectangle $R = B_1 \times B_2$, let

$$\mathcal{O}(b; R) := \frac{1}{\vec{\mu}(R)} \int_R |b(x_1, x_2) - b_R| \, d\mu_1(x_1) d\mu_2(x_2).$$

Under the assumptions of Theorem 8.4, we will show that for any ball B,

$$\mathcal{O}(b; R) \lesssim \frac{\nu(R)}{\vec{\mu}(R)}. \tag{8.2}$$

Without loss of generality, we assume that $K(x_1, y_1; x_2, y_2)$ is real-valued. Let $R = B_1 \times B_2$ be a rectangle. Then we have two rectangles $B_1 \times B_2$, $\widetilde{B}_1 \times \widetilde{B}_2$ and sets E_i, F_i, $i = 1, 2$, as above.

On the one hand, we have that for $f_i := \chi_{F_i}$, $i = 1, 2$,

$$\begin{aligned}&\frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} \int_{B_1 \times B_2} |[b, T] f_i(x_1, x_2)| \, d\mu_1(x_1) d\mu_2(x_2) \\ &\geq \frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} \int_{E_i} |[b, T] f_i(x_1, x_2)| \, d\mu_1(x_1) d\mu_2(x_2) \\ &= \frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} \int_{E_i} \int_{F_i} |b(x_1, x_2) - b(y_1, y_2)| \\ &\quad |K(x_1, y_1; x_2, y_2)| \, d\mu_1(y_1) d\mu_2(y_2) \, d\mu_1(x_1) d\mu_2(x_2) \\ &\gtrsim \frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} \int_{E_i} \int_{F_i} \frac{|b(x_1, x_2) - \alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b)|}{\mu_1(B_1)\mu_2(B_2)} \, d\mu_1(y_1) d\mu_2(y_2) \, d\mu_1(x_1) d\mu_2(x_2) \\ &\gtrsim \frac{1}{\vec{\mu}(R)} \int_{B_1 \times B_2} \left|b(x_1, x_2) - \alpha_{\widetilde{B}_1 \times \widetilde{B}_2}(b)\right| \, d\mu_1(x_1) d\mu_2(x_2) \\ &\gtrsim \mathcal{O}(b; R).\end{aligned}$$

On the other hand, from Hölder's inequality and the boundedness of $[b, T]$, we deduce that

$$\frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} \int_{B_1 \times B_2} |[b, T] f_i(x_1, x_2)| \, d\mu_1(x_1) d\mu_2(x_2)$$

$$
\begin{aligned}
&\le \frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} \left[\int_{B_1\times B_2} |[b,T]f_i(x_1,x_2)|^p \, \lambda_2(x_1,x_2)\, d\mu_1(x_1)d\mu_2(x_2) \right]^{1/p} \\
&\quad \times \left(\int_{B_1\times B_2} \lambda_2(x_1,x_2)^{-\frac{1}{p-1}} d\mu_1(x_1)d\mu_2(x_2) \right)^{1/p'} \\
&\lesssim \frac{1}{\vec{\mu}(R)} \sum_{i=1}^{2} [\lambda_1(F_i)]^{1/p} \left(\int_{B_1\times B_2} \lambda_2(x_1,x_2)^{-\frac{1}{p-1}} d\mu_1(x_1)d\mu_2(x_2) \right)^{1/p'} \\
&\lesssim \frac{1}{\vec{\mu}(R)} [\lambda_1(\widetilde{B}_1 \times \widetilde{B}_2)]^{1/p} \left(\int_{B_1\times B_2} \lambda_2(x_1,x_2)^{-\frac{1}{p-1}} d\mu_1(x_1)d\mu_2(x_2) \right)^{1/p'} \\
&\lesssim \frac{1}{\vec{\mu}(R)} [\lambda_1(R)]^{1/p} \left(\int_{R} \lambda_2(x_1,x_2)^{-\frac{1}{p-1}} d\mu_1(x_1)d\mu_2(x_2) \right)^{1/p'},
\end{aligned}
$$

where in the last inequality, we use the facts that $K_1 r_{B_1} \le d(x_{B_1}, x_{\widetilde{B}_1}) \le K_2 r_{B_1}$ and $\lambda_1(x_1,x_2) \in A_p(\vec{X})$.

Combining the two inequalities above and invoking $\lambda_i \in A_p(\vec{X})$, we conclude that

$$
\mathcal{O}(b;R) \lesssim \frac{1}{\vec{\mu}(R)} [\lambda_1(R)]^{1/p} \left(\int_{R} \lambda_2(x_1,x_2)^{-\frac{1}{p-1}} d\mu_1(x_1)d\mu_2(x_2) \right)^{1/p'} \lesssim \frac{\nu(R)}{\vec{\mu}(R)}.
$$

Thus, (8.2) holds and hence, the proof of Theorem 8.4 is complete. □

Acknowledgements The authors would like to thank the referees for careful reading and helpful suggestions, which helped to make this paper more accurate and readable. X. T. Duong and J. Li are supported by the Australian Research Council (ARC) through the research grants DP 190100970 and DP 170101060, respectively, and also supported by Macquarie University Research Seeding Grant. B. D. Wick's research supported in part by National Science Foundation DMS grant #1560995 and # 1800057. R. M. Gong is supported by NNSF of China (Grant No. 11401120) and the Foundation for Distinguished Young Teachers in Higher Education of Guangdong Province (Grant No. YQ2015126). D. Yang is supported by the NNSF of China (Grant Nos. 11971402 and 11871254).

References

1. Anderson, T.C., Damián, W.: Calderón–Zygmund operators and commutators in spaces of homogeneous type: weighted inequalities. arXiv:1401.2061
2. Betancor, J.J., Harboure, E., Nowak, A., Viviani, B.: Mapping properties of fundamental operators in harmonic analysis related to Bessel operators. Stud. Math. **197**, 101–140 (2010)
3. Bloom, S.: A commutator theorem and weighted BMO. Trans. Am. Math. Soc. **292**, 103–122 (1985)
4. Castro, A., Szarek, T.Z.: Calderón-Zygmund operators in the Bessel setting for all possible type indices. Acta Math. Sin. (Engl. Ser.), **30**, 637–648 (2014)
5. Christ, M.: A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral. Colloq. Math. **60**(61), 601–628 (1990)
6. Chung, D.: Weighted inequalities for multivariable dyadic paraproducts. Publ. Mat. **55**, 475–499 (2011)
7. Coifman, R., Lions, P.L., Meyer, Y., Semmes, S.: Compensated compactness and Hardy sapces. J. Math. Pures Appl. **72**, 247–286 (1993)
8. Coifman, R.R., Rochberg, R., Weiss, G.: Factorization theorems for Hardy spaces in several variables. Ann. Math. (2) **103**, 611–635 (1976)

9. Coifman, R.R., Weiss, G.: *Analyse harmonique non-commutative sur certains espaces homogènes. Étude de certaines intégrales singulières*, Lecture Notes in Math. vol. 242, Springer, Berlin (1971)
10. Coifman, R.R., Weiss, G.: Extensions of Hardy spaces and their use in analysis. Bull. Am. Math. Soc. **83**, 569–645 (1977)
11. Diaz, K.P.: The Szegö kernel as a singular integral kernel on a family of weakly pseudoconvex domains. Trans. Am. Math. Soc. **304**, 141–170 (1987)
12. Dragiĉević, O., Grafakos, L., Pereyra, M.C., Petermichl, S.: Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces. Publ. Mat. **49**, 73–91 (2005)
13. Duong, X., Holmes, I., Li, J., Wick, B.D., Yang, D.: Two weight Commutators in the Dirichlet and Neumann Laplacian settings. J. Funct. Anal. **276**, 1007–1060 (2019)
14. Duong, X.T., Li, H.-Q., Li, J., Wick, B.D.: Lower bound for Riesz transform kernels and commutator theorems on stratified nilpotent Lie groups. J. Math. Pures Appl. (9) **124**, 273–299 (2019)
15. Duong, X.T., Li, H.-Q., Li, J., Wick, B.D., Wu, Q.Y.: Lower bound of Riesz transform kernels revisited and commutators on stratified Lie groups. arXiv:1803.01301
16. Duong, X.T., Li, J., Wick, B.D., Yang, D.: Factorization for Hardy spaces and characterization for BMO spaces via commutators in the Bessel setting. Indiana Univ. Math. J. **66**, 1081–1106 (2017)
17. Folland, G.B., Stein, E.M.: Hardy Spaces on Homogeneous Groups. Princeton University Press, Princeton, NJ (1982)
18. Greiner, P.C., Stein, E.M.: On the solvability of some differential operators of type $\Box_b$. In: Proc. Internat. Conf., (Cortona, Italy, 1976–1977), Scuola Norm. Sup. Pisa, Pisa, pp. 106–165 (1978)
19. Guo, W., He, J., Wu, H., Yang, D.: Characterizations of the compactness of commutators associated with Lipschitz functions, arXiv:1801.06064v1
20. Guo, W., Lian, J., Wu, H.: The unified theory for the necessity of bounded commutators and applications. J. Geom. Anal. https://doi.org/10.1007/s12220-019-00226-y
21. Holmes, I., Lacey, M., Wick, B.D.: Commutators in the two-weight setting. Math. Ann. **367**, 51–80 (2017)
22. Holmes, I., Petermichl, S., Wick, B.D.: Weighted little bmo and two-weight inequalities for Journé commutators. Anal. PDE **11**, 1693–1740 (2018)
23. Huber, A.: On the uniqueness of generalized axially symmetric potentials. Ann. Math. **60**, 351–358 (1954)
24. Hytönen, T.: The sharp weighted bound for general Calderón-Zygmund operators. Ann. Math. (2) **175**, 1473–1506 (2012)
25. Hytönen, T.: The $L^p \to L^q$ boundedness of commutators with applications to the Jacobian operator. arXiv:1804.11167
26. Hytönen, T., Kairema, A.: Systems of dyadic cubes in a doubling metric space. Colloq. Math. **126**, 1–33 (2012)
27. Journé, J.L.: Calderón-Zygmund Operators, Pseudodifferential Operators and the Cauchy Integral of Calderón. Lecture Notes in Mathematics, vol. 994. Springer, Berlin (1983)
28. Kairema, A., Li, J., Pereyra, C., Ward, L.A.: Haar bases on quasi-metric measure spaces, and dyadic structure theorems for function spaces on product spaces of homogeneous type. J. Funct. Anal. **271**, 1793–1843 (2016)
29. Karagulyan, G.A.: An abstract theory of singular operators. Trans. Am. Math. Soc. **372**, 4761–4803 (2019)
30. Lerner, A.K.: On pointwise estimates involving sparse operators. N. Y. J. Math. **22**, 341–349 (2016)
31. Lerner, A.K., Nazarov, F.: Intuitive dyadic calculus: the basics. Expo. Math. **37**, 225–265 (2019)
32. Lerner, A.K., Ombrosi, S., Rivera-Ríos, I.P.: On pointwise and weighted estimates for commutators of Calderón-Zygmund operators. Adv. Math. **319**, 153–181 (2017)
33. Lerner, A.K., Ombrosi, S., Rivera-Ríos, I.P.: Commutators of singular integrals revisited. Bull. Lond. Math. Soc. **51**, 107–119 (2019)
34. Li, J., Nguyen, T., Ward, L.A., Wick, B.D.: The Cauchy integral, bounded and compact commutators. Studia Math. **250**, 193–216 (2020)
35. Macías, R.A., Segovia, C.: Lipschitz functions on spaces of homogeneous type. Adv. Math. **33**, 257–270 (1979)
36. Mao, S., Wu, H., Yang, D.: Boundedness and compactness characterizations of Riesz transform commutators on Morrey spaces in the Bessel setting. Anal. Appl. **17**, 145–178 (2019)
37. Moen, K.: Sharp weighted bounds without testing or extrapolation. Arch. Math. (Basel) **99**, 457–466 (2012)

38. Muckenhoupt, B., Wheeden, R.L.: Weighted bounded mean oscillation and the Hilbert transform. Studia Math. **54**, 221–237 (1975/1976)
39. Muckenhoupt, B., Stein, E.M.: Classical expansions and their relation to conjugate harmonic functions. Trans. Am. Math. Soc. **118**, 17–92 (1965)
40. Nehari, Z.: On bounded bilinear forms. Ann. Math. **65**, 153–162 (1957)
41. Petermichl, S., Pott, S.: An estimate for weighted Hilbert transform via square functions. Trans. Am. Math. Soc. **354**, 1699–1703 (2002)
42. Sawyer, E., Wheeden, R.L.: Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces. Am. J. Math. **114**, 813–874 (1992)
43. Stein, E.M.: Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals. Princeton University Press, Princeton, NJ (1993)
44. Tao, J., Yang, D., Yang, D.: Boundedness and compactness characterizations of the Cauchy integral commutators on Morrey spaces. Math. Methods Appl. Sci. **42**, 1631–1651 (2019)
45. Treil, S.: A remark on two weight estimates for positive dyadic operators. In: Operator-Related Function Theory and Time-Frequency Analysis, Abel Symp., vol. 9, pp. 185–195. Springer, Cham (2015)

Affiliations

Xuan Thinh Duong[1] · Ruming Gong[2] · Marie-Jose S. Kuffner[3] · Ji Li[1] · Brett D. Wick[4] · Dongyong Yang[5]

Xuan Thinh Duong
xuan.duong@mq.edu.au

Ruming Gong
gongruming@163.com

Marie-Jose S. Kuffner
mkuffne1@jhu.edu

Ji Li
ji.li@mq.edu.au

Brett D. Wick
wick@math.wustl.edu

1 Department of Mathematics, Macquarie University, Macquarie Park, NSW 2109, Australia

2 School of Mathematical Sciences, Guangzhou University, Guangzhou, China

3 Department of Mathematics, John Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA

4 Department of Mathematics, Washington University–St. Louis, St. Louis, MO 63130-4899, USA

5 School of Mathematical Sciences, Xiamen University, Xiamen, China