Michigan Math. J. 67 (2018), 757–786

Dyadic Representation and Boundedness of Nonhomogeneous Calderón–Zygmund Operators with Mild Kernel Regularity

Ana Grau de la Herrán & Tuomas Hytönen

Abstract. We prove a new dyadic representation theorem with applications to the $T(1)$ and A_2 theorems. In particular, we obtain the nonhomogeneous $T(1)$ theorem under weaker kernel regularity than in the earlier approaches.

1. Introduction

Various results in the theory of singular integrals are known "for all Calderón–Zygmund operators". Examples that we have in mind include the $T(1)$ theorem of David and Journé [3]:

$$\|T\|_{L^2\to L^2} \le C \Leftrightarrow \|T(1_Q)\|_{L^2} \le c|Q|^{1/2}, \qquad \|T^*(1_Q)\|_{L^2} \le c|Q|^{1/2} \quad \forall Q;$$

its extension to nonhomogeneous (nondoubling) measures by Nazarov, Treil, and Volberg [20]; and the A_2 theorem of the second author [7]:

$$\|T\|_{L^2(w)\to L^2(w)} \le c_T[w]_{A_2}, \quad [w]_{A_2} := \sup_Q \fint_Q w\,dx \fint_Q \frac{1}{w}\,dx.$$

However, when it comes to fine details of the definition of Calderón–Zygmund operators, it turns out that these theorems (seem to) require slightly different assumptions on the operator. We are particularly concerned about the minimal smoothness assumptions that we need to impose on the kernel of the operator.

The most common definition of Calderón–Zygmund operators involves Hölder-continuous (in a suitable scale-invariant fashion, detailed below) kernels with a power-type modulus of continuity $\omega(t) = t^\delta$ for $\delta \in (0, 1]$, and it is in this form that both the $T(1)$ and A_2 theorems first appeared. However, in many cases we can deal with more general continuity moduli with a modified Dini condition of the type

$$\int_0^1 \omega(t)\Big(1 + \log\frac{1}{t}\Big)^\alpha \frac{dt}{t} < \infty. \tag{1.1}$$

The usual Dini condition corresponds to $\alpha = 0$, and it is known to be enough for many classical results in the theory of singular integrals. It is only very recently

Received December 15, 2016. Revision received November 30, 2017.

Both authors were supported by the European Union through the ERC Starting Grant "Analytic-probabilistic methods for borderline singular integrals" (grant agreement No. 278558) and by the Academy of Finland through the Finnish Centre of Excellence in Analysis and Dynamics Research (projects Nos. 271983 and 307333).

that this was shown by Lacey [14] to be sufficient for the A_2 theorem, whereas the prior approaches needed $\alpha = 1$ (this is explicit in [12] and implicit in [16; 17]). The sharpest sufficient condition for the classical $T(1)$ theorem, to our knowledge, appears to be $\alpha = \frac{1}{2}$, which is implicit in Figiel [6] and explicit in Deng, Yan, and Yang [4]. However, the more recent extensions of the $T(1)$ theorem to nonhomogeneous measures are only available under the Hölder-continuity assumption. (See [20] for the full nonhomogeneous $T(b)$ theorem or [23], where a particular case of the $T(1)$ theorem is recovered by methods that are closely related both to the A_2 theorem and the present paper.)

Given that the critical cancellation properties of Calderón–Zygmund operators are somewhat hidden in their usual definition, all the results mentioned depend crucially on (implicit or explicit) *representation theorems* of Calderón–Zygmund operators as infinite superpositions of simpler (which often means dyadic) model operators

$$Tf = \sum_{i=1}^{\infty} \Phi_i(T, f).$$

In this way, the smoothness needed to bound an operator is linked with the smoothness required to obtain a convergent representation.

There are basically two kinds of representation theorems, linear and nonlinear. In the linear case, each Φ_i is linear in both T and f, or, as stated otherwise, $\Phi_i(T, f) = \Psi_i(T)f$, where Ψ_i is a linear transformation between suitable spaces of linear operators. Such a representation lies behind the usual proofs of the $T(1)$ theorem and the original proof of the A_2 theorem [7]. In contrast to this, the more recent approaches to the A_2 theorem [14; 17; 18] (and some recent ramifications of the $T(1)$ theorem [15]) are based on decompositions, where $\Phi_i(T, f)$ depends nonlinearly on both T and f. Typically, such nonlinear representations arise from some kind of stopping time arguments.

Although nonlinear representations appear to yield stronger results, at least in questions around the A_2 theorem, there is still independent interest toward linear representations, which are better suited, for example, for iterative applications as in [2] or multiparameter extensions as in [19; 21]. It is also of some theoretical interest whether the nonlinear methods are fundamentally stronger or whether the same results could also be recovered via linear representations.

In this paper, we prove a new (linear) dyadic representation formula with applications to both $T(1)$ and A_2 theorems. In the $T(1)$ direction, this leads to a nonhomogeneous $T(1)$ theorem under the same mild kernel regularity assumptions that were so far only known in a homogeneous setting. As for A_2, although we are not able to recover the largest class of kernels amenable to nonlinear methods, we come rather close to it and much closer than any of the previously known linear arguments.

We now turn to a more detailed description of our results. Let T be a Calderón–Zygmund operator of order n on $\mathbb{R}^d$ with respect to a Borel measure μ of order n. That is, T acts on a dense subspace of functions in $L^2(\mathbb{R}^d)$ (for the present purposes, this class should at least contain the indicators of cubes in $\mathbb{R}^d$) and has the kernel representation

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

where μ satisfies the growth condition

$$\mu(B(x,r)) \le C \cdot r^n$$

for every $x \in \mathbb{R}^d$ and every $r > 0$. Note that μ need not be a doubling measure.

Moreover, the kernel should satisfy the nth-order standard estimates, which we assume in the following form:

$$|K(x,y)| \le \frac{C}{|x-y|^n} \tag{1.2}$$

and

$$|K(x,y) - K(x',y)| + |K(y,x) - K(y,x')| \le \frac{C}{|x-y|^n} \omega\left(\frac{|x-x'|}{|x-y|}\right) \tag{1.3}$$

whenever $|x - x'| \le 1/2|x-y|$. Here ω is a modulus of continuity: an increasing and subadditive ($\omega(a+b) \le \omega(a) + \omega(b)$) function with $\omega(0) = 0$.

We also assume the "$T(1)$" conditions in the local form

$$\begin{cases} \|T1_Q\|_{L^2(\mu)} \le C\mu(Q)^{1/2}, \\ \|T^*1_Q\|_{L^2(\mu)} \le C\mu(Q)^{1/2}, \end{cases} \tag{1.4}$$

for all cubes $Q \subset \mathbb{R}^d$, where we also regard $Q = \mathbb{R}^d$ as a cube in the case $\mu(\mathbb{R}^d) < \infty$.

Our new representation theorem then takes the following form.

THEOREM 1.1. *Under the stated assumptions on T and μ and suitable test functions $f, g \in L^2(\mu)$ (as detailed in Section 2.1), the operator T admits a representation*

$$\langle Tf, g\rangle = \mathbb{E} \sum_{\substack{k \in \mathbb{Z} \\ k \ne 0}} \omega(2^{-|k|})(\langle R_k f, g\rangle + \langle Q_k f, g\rangle) + \mathbb{E}\langle \Pi_{T1} f, g\rangle + \mathbb{E}\langle f, \Pi_{T^*1} g\rangle,$$

where

(1) $\mathbb{E}$ *is the expectation over a random choice of dyadic systems on* $\mathbb{R}^d$,

(2) *each* Π_b, $b \in \{T1, T^*1\}$, *is a dyadic paraproduct with*

$$\|\Pi_b\|_{L^2(\mu) \to L^2(\mu)} \le C,$$

(3) *each R_k and Q_k is a dyadic operator with*

$$\|R_k\|_{L^2(\mu) \to L^2(\mu)} \le C, \qquad \|Q_k\|_{L^2(\mu) \to L^2(\mu)} \le C\sqrt{|k|}.$$

If μ is doubling, then each R_k and Q_k is a sum of $O(|k|)$ dyadic shifts of complexity $O(|k|)$.

Both dyadic paraproducts and dyadic shifts mentioned in the theorem have the "usual definition", which we recall further. As indicated, the dyadic operators R_k and Q_k of the new representation theorem are not precisely dyadic shifts in the sense of the usual definition, although they are closely related to them. But, in a sense, these operators have a fundamental nature, and their stated norm bounds are more efficient than what would follow from their decomposition into usual shifts and an application of known estimates.

The operators R_k and Q_k are related to, and inspired by, certain operators denoted by T_m and U_m and introduced by Figiel [5; 6]. These are, in fact, the first "dyadic shifts" in the literature, although somewhat different from the modern usage of the term.

Our main application of Theorem 1.1 is the $T(1)$ theorem, which shows that the same mild kernel regularity as in the case of the Lebesgue measure [4] is also admissible for the general nonhomogeneous (i.e., not necessarily doubling) measures μ considered here.

COROLLARY 1.2. *Under the stated assumptions on T and μ, if the modulus of continuity satisfies the Dini condition* (1.1) *with $\alpha = \frac{1}{2}$, then T acts boundedly on $L^2(\mu)$.*

Proof assuming Theorem 1.1. Using the decomposition and norm estimates provided by Theorem 1.1, it follows that

$$\begin{aligned}
|\langle Tf, g\rangle| &\le \sum_{\substack{k\in\mathbb{Z}\\ k\neq 0}} \omega(2^{-|k|})C\big(\|f\|_2\|g\|_2 + \sqrt{|k|}\,\|f\|_2\|g\|_2\big) + C\|f\|_2\|g\|_2 \\
&\le C\|f\|_2\|g\|_2\Big(1 + \sum_{k=1}^{\infty}\omega(2^{-k})\sqrt{k}\Big) \\
&\le C\|f\|_2\|g\|_2\Big(1 + \int_0^1 \omega(t)\sqrt{\log\frac{1}{t}}\,\frac{dt}{t}\Big) \\
&\le C\|f\|_2\|g\|_2
\end{aligned}$$

by an easy comparison of sums and integrals. □

Toward the A_2 theorem, we obtain the following:

COROLLARY 1.3. *Let $n = d$, and let μ be the Lebesgue measure on $\mathbb{R}^d$. Under the stated assumptions on T, if the modulus of continuity satisfies the Dini condition* (1.1) *with $\alpha = 2$, then T satisfies the A_2 inequality*

$$\|Tf\|_{L^2(w)} \le C[w]_{A_2}\|f\|_{L^2(w)}.$$

Although this does not quite recover the form of the A_2 theorem obtained by Lacey [14] with $\alpha = 0$ (see also the simplification by Lerner [18]) or even the earlier results [12; 16; 17] using nonlinear representations with $\alpha = 1$, this extends the class of Calderón–Zygmund operators that we can handle by any of the linear representation theorems currently known.

Proof assuming Theorem 1.1. Since dyadic shifts of complexity $O(|k|)$ are bounded on $L^2(w)$ with norm $O(|k| \cdot [w]_{A_2})$ (see [11, Theorem 2.10] or [22, Theorem 4.1]), it follows that

$$\|R_k\|_{L^2(w)\to L^2(w)} + \|Q_k\|_{L^2(w)\to L^2(w)} \leq Ck^2[w]_{A_2}.$$

Recalling that dyadic paraproducts are also bounded on $L^2(w)$ with norm $O([w]_{A_2})$ (see [1]), substituting all this into the representation formula of Theorem 1.1, we obtain

$$\begin{aligned}\|T\|_{L^2(w)\to L^2(w)} &\leq C\Big(1+\sum_{k=1}^{\infty}\omega(2^{-k})k^2\Big)[w]_{A_2}\\ &\leq C\Big(1+\int_0^1\omega(t)\Big(\log\frac{1}{t}\Big)^2\frac{dt}{t}\Big)[w]_{A_2}\\ &\leq C[w]_{A_2}.\end{aligned}$$ □

We actually suspect that a direct weighted analysis of the new operators R_k and Q_k (instead of their reduction to known results about dyadic shifts) could lead to the better weighted bounds

$$\|R_k\|_{L^2(w)\to L^2(w)} + \|Q_k\|_{L^2(w)\to L^2(w)} \overset{?}{\leq} C|k|[w]_{A_2}$$

and thus to a linear proof of the A_2 theorem under the Dini condition (1.1) with $\alpha = 1$. However, since this would still be weaker than Lacey's result [14] with $\alpha = 0$, we have not pushed hard on this point.

For future investigations, we point out that our nonhomogeneous results (for power bounded measures on $\mathbb{R}^d$) should extend to the case of upper doubling measures on geometrically doubling metric spaces (as in [13]) with essentially notational complications only. Somewhat less obvious may be the extension to representations appropriate for $T(b)$ (rather than just $T(1)$) theorems. Multilinear and multiparameter extensions should also be possible.

1.1. Plan of the Paper

We recall some preliminaries and notation in Section 2. Once the notation is available, we provide a more detailed statement of the main Theorem 1.1 in Section 3, including a precise formula for the various operators appearing in the decomposition. The proof of the theorem is then divided into the subsequent sections. We split the main part of the proof into *identities* and *estimates,* namely, writing T as a sum of the dyadic pieces (Section 4) and showing that these pieces satisfy the relevant norm bounds (Section 5). The proof of Theorem 1.1 is completed in Section 6, where we establish the asserted shift structure of the new operators in the homogeneous situation. The last Section 7 provides additional information about the weak $(1, 1)$ behavior of the new operators, again in the homogeneous situation only.

2. Preliminaries

Starting from a fixed reference system of dyadic cubes $\mathscr{D}^0$, we will consider new dyadic systems obtained by translating the reference system as follows. Let $\sigma = (\sigma_j)_{j\in\mathbb{Z}} \in (\{0,1\}^d)^{\mathbb{Z}}$ and

$$I \dot{+} \sigma := I + \sum_{j:2^{-j}<\ell(I)} 2^{-j}\sigma_j.$$

Then

$$\mathscr{D}^\sigma := \{I \dot{+} \sigma : I \in \mathscr{D}^0\},$$

and it is straightforward to check that $\mathscr{D}^\sigma$ inherits the important nestedness property of $\mathscr{D}^0$: if $I, J \in \mathscr{D}^\sigma$, then $I \cap J \in \{I, J, \emptyset\}$. When the particular σ is unimportant, the notation $\mathscr{D}$ is sometimes used for a generic dyadic system.

The reference system could be, but need not be, the standard system

$$\mathscr{D}^0 = \{2^{-k}([0,1)^d + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^d\};$$

we could equally well start from any other fixed reference system.

Within any fixed system of dyadic cubes $\mathscr{D}$, we use the following notation:

- $I^{(r)}$ is the rth ancestor I, that is, $I^{(r)} \supseteq I$ and $\ell(I^{(r)}) = 2^r\ell(I)$.
- $\mathrm{ch}(I)$ is the set of children of I, that is, $J \in \mathrm{ch}(I)$ if and only if $J^{(1)} = I$.
- $\mathscr{D}_k$ refers to the set of cubes $Q \in \mathscr{D}$ such that $\ell(Q) = 2^{-k}$.

Moreover, we write:

$$\langle f\rangle_I = \frac{1}{\mu(I)}\int_I f(x)\,d\mu(x), \qquad \langle f, g\rangle = \int_{\mathbb{R}^d} f(x)g(x)\,d\mu(x),$$
$$E_I f = 1_I\langle f\rangle_I, \qquad E_k f = \sum_{I\in\mathscr{D}_k} E_I f,$$
$$D_I f = \sum_{I'\in\mathrm{ch}(I)} E_{I'} f - E_I f, \qquad D_k f = \sum_{I\in\mathscr{D}_k} D_I f.$$

DEFINITION 2.1. For every $I \in \mathscr{D}$, we define the **Haar functions** as a collection of functions $\{\varphi_I^i\}_{i=1}^{2^d-1}$ such that

(1) $\operatorname{supp}\varphi_I^i \subseteq I$,
(2) φ_I^i is constant on each $I' \in \mathrm{ch}(I)$,
(3) $\int \varphi_I^i\,d\mu = 0$,
(4) $\|\varphi_I^i\|_\infty \cdot \|\varphi_I^i\|_1 \le C$ (a constant independent of I and i),
(5) $\|\varphi_I^i\|_2 \in \{0,1\}$, and
(6) $D_I f = \sum_{i=1}^{2^d-1}\langle f, \varphi_I^i\rangle\varphi_I^i$.

The proof of the existence of such functions can be found in [8, Section 4]. If μ is doubling, then the construction is well known, and in this case, (4) can be improved to $\|\varphi_I^i\|_\infty \le C\mu(I)^{-1/2}$.

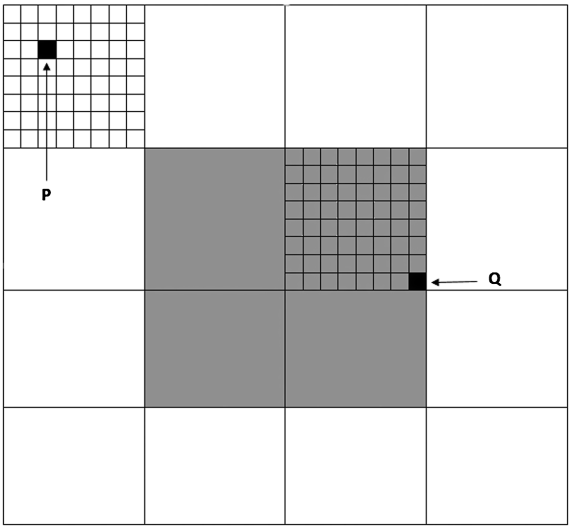

Figure 1 Illustration of k-goodness with $k = 5$. The small cubes in the picture (which have 2^5 times smaller side length than the big cube) are 5-good whenever they belong to the grey region (like Q) and 5-bad whenever they belong to the rest of the big cube (like P).

DEFINITION 2.2. For $k \in \mathbb{N}$ and $k \geq 2$, we say that a cube $I \in \mathscr{D}^\sigma$ is k**-good** if $\operatorname{dist}(I, \partial I^{(k)}) \geq \frac{1}{4}\ell(I^{(k)})$. (See Figure 1 for an illustration.)

The probability of a particular cube $I \dot{+} \sigma$ being k-good is equal for all cubes $I \in \mathscr{D}^0$ and for all k, so we denote

$$\pi_{\text{good}} := \mathbb{P}_\sigma(I \dot{+} \sigma \ k\text{-good}) = 2^{-d}.$$

Despite the easy numerical value of this expression, we use the notation π_{good} to stress the origin of this factor from goodness considerations in the relevant expressions.

REMARK 2.3. Let Q be a cube, and $r, k \in \mathbb{N}$, $r, k \geq 2$. The position and k-goodness of $Q \dot{+} \sigma$ are independent random variables. In addition, the r-goodness of $Q \dot{+} \sigma$ is independent of the k-goodness of $(Q \dot{+} \sigma)^{(r)}$.

Indeed, the position of $Q \dot{+} \sigma$ by definition depends only on σ_j for $2^{-j} < \ell(Q)$. On the other hand, the k-goodness of $Q \dot{+} \sigma$ depends on its relative position with respect of $(Q \dot{+} \sigma)^{(k)}$. Since the same translation components σ_j for $2^{-j} < \ell(Q)$ appear in both $Q \dot{+} \sigma$ and $(Q \dot{+} \sigma)^{(k)}$, this relative position depends only on σ_j for $\ell(Q) \leq 2^{-j} < \ell(Q^{(k)})$, which makes the position and k-goodness independent random variables. The other claim is proved similarly.

DEFINITION 2.4. Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, $r \geq 2$, and $H \in \mathscr{D}$. Then we define

$$D_H^{(r)} f := \sum_{I : I^{(r)} = H} D_I f, \qquad P_H^{(r)} f := \sum_{j=0}^{r-1} D_H^{(j)} f \tag{2.1}$$

and

$$D_H^{(r,\,\text{good})} f := \sum_{\substack{I : I^{(r)} = H \\ I\ r\text{-good}}} D_I f, \qquad P_H^{(r,\,\text{good})} f := \sum_{j=2}^{r-1} D_H^{(j,\,\text{good})} f. \tag{2.2}$$

Due to orthogonality, it is a standard computation to verify that, for any $k \geq 1$,

$$\Big(\sum_K \| D_K^{(k)} f \|_2^2 \Big)^{1/2} \leq \| f \|_2.$$

2.1. A Finitary Setup

The core of the proof of Theorem 1.1 is based on somewhat elaborate manipulations of the random martingale difference expansions of f and g in the duality pairing $\langle Tf, g \rangle$. To minimize the need of tiresome justifications of the rearrangements of sums and expectations, we choose the following setup, similar to [10] or [23], to carry out these manipulations.

Let $f, g \in L^2(\mu)$ be both supported on some big cube K_1 and constant on the dyadic subcubes of K_1 of side length $2^{-N}\ell(K_1)$. Clearly, such functions are dense in $L^2(\mu)$ when both K_1 and N are allowed to vary. Let K_0 be the cube with $\ell(K_0) = 2\ell(K_1)$, positioned so that K_1 is the "upper right" quadrant of K_0.

Let us then take as the reference dyadic system some $\mathscr{D}^0$ that contains K_0. (This determines the choice of dyadic cubes smaller than K_0 uniquely, but we can make an arbitrary choice of the dyadic ancestors of K_0 in the system $\mathscr{D}^0$.) Then we consider a modified version of the shifted systems with

$$K \dot{+} \sigma := K + \sum_{j : 2^{-N}\ell(K_1) \leq 2^{-j} < \min(\ell(K), \ell(K_1))} 2^{-j} \sigma_j;$$

that is, we only randomize on length scales larger than the minimal length scale $2^{-N}\ell(K_1)$, where the functions f and g are constant. Thus $\mathscr{D}^\sigma$ and $\mathscr{D}^0$ have the same small dyadic cubes, and since f and g are constant on these cubes, the corresponding martingale differences vanish.

Similarly, we only randomize on length scales smaller than the support of the functions. In particular, no matter the value of σ, we find that $K_0 \dot{+} \sigma \supset K_1$ contains the supports of f and g. Thus, irrespective of the value of σ, we have

$$f = E_{K_0 \dot{+} \sigma} f + \sum_{\substack{K \in \mathscr{D}^\sigma : K \subseteq K_0 \dot{+} \sigma \\ \ell(K) > 2^{-n}\ell(K_1)}} D_K f,$$

with a similar expansion for g. Now, there is only a fixed finite number of terms in this sum, independently of σ. In fact, there are only finitely many relevant values of σ, since only N coordinates σ_j with $2^{-N}\ell(K_1) \le 2^{-j} < \ell(K_1)$ have an impact on the definition of $K\dot{+}\sigma$. Thus, all our sums and expectations range over fixed finite ranges only, so that the rearrangement of the summation order is never an issue. Since the terms $E_{K_0\dot{+}\sigma} f$ and $E_{K_0\dot{+}\sigma} g$ are trivial to handle by the $T(1)$ conditions (1.4), we may also assume that these terms (equivalently, the integrals $\int f\,d\mu$ and $\int g\,d\mu$) vanish, so that both f and g are expanded in terms of the martingale differences alone.

With this reduction stated, however, we will not indicate it explicitly, but simply write $\sum_{K\in\mathscr{D}}$, or just $\sum_K$, for the relevant sum.

A cautious reader might notice that our fiddling with the definition of "$K\dot{+}\sigma$" will have some impact on the properties of the notion of goodness discussed. However, the changes are essentially immaterial, since goodness is really only relevant to us when both smaller and bigger cubes appear in our martingale difference expansion, and for such cubes, our randomization is essentially unchanged. We will leave the detailed verification of this "no harm done" assertion to the interested reader; see also the discussion of "good" and "really good" cubes in [23].

3. The Dyadic Representation Theorem: Detailed Statement

With the notation defined, we can provide additional details to Theorem 1.1 as follows:

THEOREM 3.1. *In the representation Theorem 1.1, the various operators have the following form*:

$$Q_k = \sum_{K\in\mathscr{D}} A_K^{(k)}, \qquad R_k = \sum_{K\in\mathscr{D}} B_K^{(k)},$$

where the operators $A_K^{(k)}$ and $B_K^{(k)}$ satisfy the orthogonality relations:

$$\begin{aligned}
A_K^{(k)} &= P_K^{(k+1)} A_K^{(k)} D_K^{(k+r)},\\
A_K^{(-k)} &= D_K^{(k+r)} A_K^{(-k)} P_K^{(k+1)},\\
B_K^{(k)} &= (P_K^{(k+r+1)} - P_K^{(k)}) B_K^{(k)} D_K^{(k+r)},\\
B_K^{(-k)} &= D_K^{(k+r)} B_K^{(-k)} (P_K^{(k+r)} - P_K^{(k)}), \quad k \ge 1,
\end{aligned}$$

and their kernels $a_K^{(k)}(x,y)$ and $b_K^{(k)}(x,y)$ satisfy the bounds

$$|a_K^{(k)}(x,y)| \le C\frac{1_K(x)1_K(y)}{\ell(K)^n} + C\sum_{H:H^{(|k|)}=K} \frac{1_H(x)1_H(y)}{\mu(H)}, \quad |k| \ge 1,$$

$$|b_K^{(k)}(x,y)| \le C\frac{1_K(x)1_K(y)}{\ell(K)^n}, \quad |k| \ge 2.$$

Moreover, the paraproducts have the form

$$\Pi_b f = \sum_K D_K^{(r,\text{good})} b(\langle f\rangle_K - E_{-\infty} f),$$

where

$$E_{-\infty} f := \begin{cases} \mu(\mathbb{R}^d)^{-1} \int_{\mathbb{R}^d} f \, d\mu & \text{if } \mu(\mathbb{R}^d) < \infty, \\ 0 & \text{otherwise}. \end{cases}$$

If μ is doubling, then the following additional statements hold:

- *The pointwise bound for $b_K^{(k)}(x, y)$ is also valid for $|k| = 1$.*
- *For each $U_k \in \{R_k, Q_k, R_k^*, Q_k^*\}$, we have $\|U_k\|_{L^1(\mu) \to L^{1,\infty}(\mu)} \le C|k|$.*

The proof of Theorems 1.1 and 3.1 naturally splits into two parts: the algebraic identities that give the desired decomposition and the estimates for the terms of this expansion. We deal with each of these tasks in turn in the following two sections.

4. Dyadic Representation: Identities

The following proposition is introduces goodness into the basic martingale difference expansion.

PROPOSITION 4.1. *Let T be a Calderón–Zygmund operator, $f, g \in L^2(\mu)$ as in the finitary setup of Section 2.1, and $r \in \mathbb{N}$ with $r \ge 2$. Then T has the following expansion:*

$$\begin{aligned} \langle Tf, g \rangle &= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \sum_{I, J \in \mathscr{D}^\sigma} 1_{r\text{-good}}(\text{smaller}\{I, J\}) \cdot \langle T D_I f, D_J g \rangle \\ &= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \Bigg(\sum_{i, j \in \mathbb{Z}} \langle T(\chi_{ij} D_i f), \psi_{ij} D_j g \rangle \Bigg), \end{aligned}$$

where

$$\text{smaller}\{I, J\} := \begin{cases} I & \text{if } \ell(I) \le \ell(J), \\ J & \text{if } \ell(I) > \ell(J), \end{cases} \qquad \phi_j = \sum_{\substack{I \in \mathscr{D}_j \\ I \ r\text{-good}}} 1_I,$$

and

$$\chi_{ij} := \begin{cases} \phi_i & \text{if } i \ge j, \\ 1 & \text{if } i < j, \end{cases} \qquad \psi_{ij} := \begin{cases} 1 & \text{if } i \ge j, \\ \phi_j & \text{if } i < j. \end{cases}$$

Proof. This result is similar to Proposition 3.5 of [9], which we are going to replicate here with our definition of goodness.

Recall that

$$f = \sum_{I \in \mathscr{D}^0} D_{I \dot{+} \sigma} f$$

for any $\sigma \in (\{0, 1\}^d)^{\mathbb{Z}}$; and we can also take the expectation $\mathbb{E}_\sigma$ of both sides of this identity.

We use this random $D_{I \dot{+} \sigma}$ expansion of f, multiply and divide by

$$\pi_{\text{good}} = \mathbb{E}_\sigma 1_{r\text{-good}}(I \dot{+} \sigma),$$

and use the independence from Remark 2.3 to get:

$$\begin{aligned}
\langle Tf, g\rangle &= \mathbb{E}_\sigma \sum_I \langle T D_{I\dot{+}\sigma} f, g\rangle \\
&= \frac{1}{\pi_{\text{good}}} \sum_I \mathbb{E}_\sigma 1_{r\text{-good}}(I\dot{+}\sigma)\mathbb{E}_\sigma \langle T D_{I\dot{+}\sigma} f, g\rangle \\
&= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \sum_I 1_{r\text{-good}}(I\dot{+}\sigma)\langle T D_{I\dot{+}\sigma} f, g\rangle \\
&= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \sum_{I,J} 1_{r\text{-good}}(I\dot{+}\sigma)\langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle.
\end{aligned}$$

On the other hand, using independence again in half of this double sum, we have

$$\begin{aligned}
&\frac{1}{\pi_{\text{good}}} \sum_{\ell(I)>\ell(J)} \mathbb{E}_\sigma 1_{r\text{-good}}(I\dot{+}\sigma)\langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle \\
&\quad= \frac{1}{\pi_{\text{good}}} \sum_{\ell(I)>\ell(J)} \mathbb{E}_\sigma 1_{r\text{-good}}(I\dot{+}\sigma)\mathbb{E}_\sigma \langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle \\
&\quad= \mathbb{E}_\sigma \sum_{\ell(I)>\ell(J)} \langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle,
\end{aligned}$$

and hence

$$\begin{aligned}
\langle Tf, g\rangle = &\frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \sum_{\ell(I)\le\ell(J)} 1_{r\text{-good}}(I\dot{+}\sigma)\langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle \\
&+ \mathbb{E}_\sigma \sum_{\ell(I)>\ell(J)} \langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle.
\end{aligned}$$

Comparison with the basic identity

$$\langle Tf, g\rangle = \mathbb{E}_\sigma \sum_{I,J} \langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle$$

shows that

$$\begin{aligned}
&\mathbb{E}_\sigma \sum_{\ell(I)\le\ell(J)} \langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle \\
&\quad= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \sum_{\ell(I)\le\ell(J)} 1_{r\text{-good}}(I\dot{+}\sigma)\langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle.
\end{aligned}$$

Symmetrically, we also have

$$\begin{aligned}
&\mathbb{E}_\sigma \sum_{\ell(I)>\ell(J)} \langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle \\
&\quad= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\sigma \sum_{\ell(I)>\ell(J)} 1_{r\text{-good}}(J\dot{+}\sigma)\langle T D_{I\dot{+}\sigma} f, D_{J\dot{+}\sigma} g\rangle,
\end{aligned}$$

and this completes the proof of the first asserted identity. The second one is a simple restatement, as seen by the computation

$$\begin{aligned}\sum_{I,J} 1_{r\text{-good}}(\text{smaller}\{I,J\})\langle TD_I f, D_J f\rangle \\ &= \sum_{i\geq j}\langle T(\phi_i D_i f), D_j g\rangle + \sum_{i<j}\langle TD_i f, \phi_j D_j g\rangle \\ &= \sum_{i,j}\langle T(\chi_{ij} D_i f), \psi_{ij} D_j g\rangle.\end{aligned}$$

□

We split the subsequent analysis of the series into four cases, depending on whether $i \geq j$ or $i < j$, and whether $|i-j| \leq r$ or $|i-j| > r$. Since the cases $i > j$ and $i < j$ are dual to each other, we only explicitly deal with $i \geq j$, which still splits into the two cases $0 \leq i-j \leq r$ and $j < i-r$.

PROPOSITION 4.2. *Let T be a Calderón–Zygmund operator, $f, g \in L^2(\mu)$ as in Section 2.1, and $r \in \mathbb{N}$ such that $r \geq 2$. Then*

$$\sum_{\substack{i,j\\0\leq i-j\leq r}} \langle T(\chi_{ij} D_i f), \psi_{ij} D_j g\rangle = \sum_{m\in\mathbb{Z}^d}\sum_{H}\langle TD_H^{(r,\,good)} f, P_{H\dot{+}m}^{(r+1)} g\rangle.$$

REMARK 4.3. A similar argument would show that

$$\sum_{\substack{i,j\\0< j-i\leq r}} \langle T(\chi_{ij} D_i f), \psi_{ij} D_j g\rangle = \sum_{m\in\mathbb{Z}^d}\sum_{H}\langle TP_{H\dot{+}m}^{(r)} f, D_H^{(r,\,\text{good})} g\rangle,$$

where the slight symmetry-break ($P_{H\dot{+}m}^{(r)}$ vs. $P_{H\dot{+}m}^{(r+1)}$) is caused by the fact that the diagonal $i = j$ is included in only one of the cases.

Proof of Proposition 4.2.

$$\begin{aligned}\sum_{0\leq i-j\leq r} \langle T(\phi_i D_i f), D_j g\rangle \\ &= \sum_{0\leq k\leq r}\sum_{I\ r\text{-good}}\sum_{J:2^{r-k}\ell(I)=\ell(J)} \langle TD_I f, D_J g\rangle \\ &= \sum_{m\in\mathbb{Z}^d}\sum_{0\leq k\leq r}\sum_{H}\sum_{\substack{I:I^{(r)}=H\\ I\ r\text{-good}}}\sum_{J:J^{(k)}=H\dot{+}m} \langle TD_I f, D_J g\rangle \\ &= \sum_{m\in\mathbb{Z}^d}\sum_{0\leq k\leq r}\sum_{H}\langle TD_H^{(r,\,\text{good})} f, D_{H\dot{+}m}^{(k)} g\rangle \\ &= \sum_{m\in\mathbb{Z}^d}\sum_{H}\langle TD_H^{(r,\,\text{good})} f, P_{H\dot{+}m}^{(r+1)} g\rangle.\end{aligned}$$

□

PROPOSITION 4.4. *Let T be a Calderón–Zygmund operator, $f, g \in L^2(\mu)$ as in Section 2.1, and $r \in \mathbb{N}$ such that $r \geq 2$. Then*

$$\begin{aligned}\sum_i \sum_{j<i-r} \langle T(\chi_{ij} D_i f), \psi_{ij} D_j g\rangle &= \sum_{\substack{m\in\mathbb{Z}^d\\ m\neq 0}} \sum_H \langle T D_H^{(r,\text{ good})} f, 1_{H\dot{+}m}\rangle(\langle g\rangle_{H\dot{+}m} - \langle g\rangle_H) \\ &\quad + \sum_H \langle f, D_H^{(r,\text{ good})} T^* 1\rangle(\langle g\rangle_H - E_{-\infty} g),\end{aligned}$$

where

$$E_{-\infty} g := \begin{cases} \frac{1}{\mu(\mathbb{R}^d)} \int g(x)\, d\mu(x) & \text{if } \mu(\mathbb{R}^d) < \infty, \\ 0 & \text{otherwise.}\end{cases}$$

Proof. Before starting, we want to point out that $\sum_{j<i-r} D_j g = E_{i-r} g - E_{-\infty} g$. Once this remark has been done, we can proceed to the proof:

$$\begin{aligned}\sum_i \sum_{j<i-r} \langle T(\chi_{ij} D_i f), \psi_{ij} D_j g\rangle &= \sum_i \sum_{j<i-r} \langle T\phi_i D_i f, D_j g\rangle \\ &= \sum_i \Big\langle T\phi_i D_i f, \sum_{j<i-r} D_j g\Big\rangle \\ &= \sum_i \langle T\phi_i D_i f, E_{i-r} g - E_{-\infty} g\rangle.\end{aligned}$$

Recalling the definition of ϕ_i, we can rewrite the second part of this sum as

$$\begin{aligned}\sum_i \langle T\phi_i D_i f, E_{-\infty} g\rangle &= \sum_i \langle T\phi_i D_i f, 1\rangle E_{-\infty} g \\ &= \sum_i \langle \phi_i D_i f, T^* 1\rangle E_{-\infty} g = \sum_H \langle D_H^{(r,\text{ good})} f, T^* 1\rangle E_{-\infty} g\end{aligned}$$

and

$$\begin{aligned}\sum_i \langle T\phi_i D_i f, E_{i-r} g\rangle &= \sum_{\substack{I\ r\text{-good}\\ J:\ell(J)=2^r\ell(I)}} \langle T D_I f, E_J g\rangle \\ &= \sum_{\substack{H,J\\ \ell(H)=\ell(J)}} \Big\langle T \sum_{\substack{I:I^{(r)}=H\\ I\ r\text{-good}}} D_I f, E_J g\Big\rangle \\ &= \sum_{m\in\mathbb{Z}^d} \sum_H \langle T D_H^{(r,\text{ good})} f, E_{H\dot{+}m} g\rangle \\ &= \sum_{m\in\mathbb{Z}^d} \sum_H \langle T D_H^{(r,\text{ good})} f, 1_{H\dot{+}m}\rangle(\langle g\rangle_{H\dot{+}m} - \langle g\rangle_H) \\ &\quad + \sum_H \langle T D_H^{(r,\text{ good})} f, 1\rangle \langle g\rangle_H.\end{aligned}$$

We can add the restriction $m \neq 0$, since the summand vanishes for $m = 0$ in any case. The self-adjointness of the operator $D_H^{(r,\,\text{good})}$ finishes the proof. □

The previous two propositions both introduce a double series over terms of the form $\Phi(H, H\dot{+}m)$. The following lemma provides a useful rearrangement of such summations.

LEMMA 4.5.

$$\begin{aligned}\mathbb{E} \sum_{m \in \mathbb{Z}^d \setminus \{0\}} \sum_H \Phi(H, H\dot{+}m) \\ = \frac{1}{\pi_{\text{good}}} \mathbb{E} \sum_{k=2}^{\infty} \sum_{|m| \sim 2^{k-2}} \sum_{H\ k\text{-good}} \Phi(H, H\dot{+}m),\end{aligned}$$

where

$$|m| \sim 2^{k-2} \quad \overset{\text{def}}{\Leftrightarrow} \quad 2^{k-3} < |m| \leq 2^{k-2}.$$

Proof. Since every $m \in \mathbb{Z}^d \setminus \{0\}$ satisfies $|m| \sim 2^{k-2}$ for a unique $k \geq 2$, and the k-goodness of H is independent of the position of H (and hence of $H\dot{+}m$), we have

$$\begin{aligned}&\mathbb{E} \sum_{m \in \mathbb{Z}^d \setminus \{0\}} \sum_{H \in \mathscr{D}^\sigma} \Phi(H, H\dot{+}m) \\ &= \frac{1}{\pi_{\text{good}}} \sum_{k=2}^{\infty} \sum_{|m| \sim 2^{k-2}} \sum_{H \in \mathscr{D}^0} \mathbb{E}(1_{k\text{-good}}(H\dot{+}\sigma)) \mathbb{E}\Phi(H\dot{+}\sigma, H\dot{+}\sigma\dot{+}m) \\ &= \frac{1}{\pi_{\text{good}}} \sum_{k=2}^{\infty} \sum_{|m| \sim 2^{k-2}} \sum_{H \in \mathscr{D}^0} \mathbb{E}(1_{k\text{-good}}(H\dot{+}\sigma) \Phi(H\dot{+}\sigma, H\dot{+}\sigma\dot{+}m)) \\ &= \frac{1}{\pi_{\text{good}}} \mathbb{E} \sum_{k=2}^{\infty} \sum_{|m| \sim 2^{k-2}} \sum_{H \in \mathscr{D}^\sigma\ k\text{-good}} \Phi(H, H\dot{+}m).\end{aligned}$$

□

The usefulness of the previous rearrangement is in the following:

LEMMA 4.6. *If H is k-good and $|m| \leq 2^{k-2}$, then $H\dot{+}m \subset H^{(k)}$.*

Proof. Let $K := H^{(k)}$, so that $\ell(K) = 2^k \ell(H)$ and $\operatorname{dist}(H, K^c) \geq \frac{1}{4}\ell(K) = 2^{k-2}\ell(H)$ by the definition of k-goodness. If x is any interior point of H, then this means that $\operatorname{dist}(x, K^c) > 2^{k-2}\ell(H)$. Every interior point y of $H\dot{+}m$ has the form $y = x + m\ell(H)$ for such an x, and hence $\operatorname{dist}(y, K^c) \geq \operatorname{dist}(x, K^c) - |m|\ell(H) > 2^{k-2}\ell(H) - 2^{k-2}\ell(H) = 0$. Thus $y \in K$ for every interior point $y \in H\dot{+}m$, and hence $H\dot{+}m \subset K$. □

A combination of Propositions 4.2 and 4.4 with Lemma 4.5 shows that

$$
\begin{aligned}
&\frac{\mathbb{E}}{\pi_{\text{good}}} \sum_{\substack{i,j\\ j\le i}} \langle T(\chi_{ij} D_i f), \psi_{ij} D_j g\rangle \\
&\quad = \frac{\mathbb{E}}{\pi_{\text{good}}} \Bigg(\sum_{\substack{i,j\\ 0\le i-j\le r}} + \sum_{\substack{i,j\\ j<i-r}} \Bigg) \langle T(\phi_i D_i f), D_j g\rangle \\
&\quad = \frac{\mathbb{E}}{\pi_{\text{good}}} \sum_H \langle T D_H^{(r,\text{good})} f, P_H^{(r+1)} g\rangle \\
&\qquad + \frac{\mathbb{E}}{\pi_{\text{good}}^2} \sum_{k=2}^{\infty} \sum_{|m|\sim 2^{k-2}} \sum_{H\ k\text{-good}} \langle T D_H^{(r,\text{good})} f, P_{H\dot{+}m}^{(r+1)} g\rangle \\
&\qquad + \frac{\mathbb{E}}{\pi_{\text{good}}^2} \sum_{k=2}^{\infty} \sum_{|m|\sim 2^{k-2}} \sum_{H\ k\text{-good}} \langle T D_H^{(r,\text{good})} f, 1_{H\dot{+}m}\rangle (\langle g\rangle_{H\dot{+}m} - \langle g\rangle_H) \\
&\qquad + \frac{\mathbb{E}}{\pi_{\text{good}}} \sum_H D_H^{(r,\text{good})} T^* 1(\langle g\rangle_H - E_{-\infty} g) \\
&\quad =: \mathbb{E}\omega(2^{-1})\langle R_1 f, g\rangle + \mathbb{E}\sum_{k=2}^{\infty} \omega(2^{-k})\langle R_k f, g\rangle \\
&\qquad + \mathbb{E}\sum_{k=2}^{\infty} \omega(2^{-k})\langle Q_k f, g\rangle + \mathbb{E}\langle f, \Pi_{T^*1} g\rangle,
\end{aligned}
$$

where

$$
\begin{aligned}
R_1 f &:= \frac{1}{\pi_{\text{good}}} \frac{1}{\omega(\frac12)} \sum_H P_H^{(r+1)} T D_H^{(r,\text{good})} f, \\
R_k f &:= \frac{1}{\pi_{\text{good}}} \sum_{|m|\sim 2^{k-2}} \sum_{H\ k\text{-good}} \frac{1}{\omega(2^{-k})} P_{H\dot{+}m}^{(r+1)} T D_H^{(r,\text{good})} f, \quad k\ge 2, \\
Q_k f &:= \frac{1}{\pi_{\text{good}}^2} \sum_{|m|\sim 2^{k-2}} \sum_{H\ k\text{-good}} \frac{\langle T D_H^{(r,\text{good})} f, 1_{H\dot{+}m}\rangle}{\omega(2^{-k})} \Bigg(\frac{1_{H\dot{+}m}}{\mu(H\dot{+}m)} - \frac{1_H}{\mu(H)} \Bigg),
\end{aligned}
$$

and

$$
\Pi_{T^*1} g := \sum_H D_H^{(r,\text{good})} T^* 1(\langle g\rangle_H - E_{-\infty} g).
$$

Note that we have incorporated the factor $\omega(2^{-k})$ into the definition of R_k and Q_k to achieve a favorable normalization of the series.

For reasons of symmetry, we also have a similar decomposition of the other half of the double sum, namely

$$
\begin{aligned}
&\frac{\mathbb{E}}{\pi_{\text{good}}}\sum_{\substack{i,j\\ i<j}}\langle T(\chi_{ij}D_i f), \psi_{ij}D_j g\rangle \\
&\qquad = \frac{\mathbb{E}}{\pi_{\text{good}}}\sum_{\substack{i,j\\ i<j}}\langle D_i f, T^*(\phi_j D_j g)\rangle \\
&\qquad = \mathbb{E}\omega(2^{-1})\langle f, \tilde{R}_1 f\rangle + \mathbb{E}\sum_{k=2}^{\infty}\omega(2^{-k})\langle f, \tilde{R}_k g\rangle + \mathbb{E}\sum_{k=2}^{\infty}\omega(2^{-k})\langle f, \tilde{Q}_k g\rangle \\
&\qquad\quad + \mathbb{E}\langle \Pi_{T1} f, g\rangle,
\end{aligned}
$$

where $\tilde{R}_k$ and $\tilde{Q}_k$ have the same form as R_k and Q_k, respectively, with the only difference that

- T is replaced by T^* throughout, and
- $P^{(r+1)}$ in R_k is replaced by $P^{(r)}$ in $\tilde{R}_k$ (cf. Remark 4.3).

Defining $Q_1 := 0$ and

$$
R_{-k} := \tilde{R}_k^*, \qquad Q_{-k} := \tilde{Q}_k^*,
$$

we then obtain the desired identity

$$
\langle Tf, g\rangle = \mathbb{E}\sum_{\substack{k\in\mathbb{Z}\\ k\neq 0}}\omega(2^{-|k|})(\langle R_k f, g\rangle + \langle Q_k f, g\rangle) + \mathbb{E}(\langle \Pi_{T1} f, g\rangle + \langle f, \Pi_{T^*1} g\rangle).
$$

It remains to establish the claimed properties of the operators.

5. Dyadic Representation: Estimates

Having established the dyadic representation on an algebraic level, we turn to the relevant estimates for the various terms in the obtained expansion. Since the operators R_k and Q_k are essentially similar for positive and negative values of k, we only explicitly deal with $k \geq 1$.

5.1. The Paraproducts

Our dyadic decomposition of the operator T led to two dyadic paraproducts of the form

$$
\begin{aligned}
\Pi_b f &= \sum_{H\in\mathscr{D}} D_H^{(r,\text{good})} b(\langle f\rangle_H - E_{-\infty} f), \quad b \in \{T1, T^*1\}, \\
&= \sum_{H\in\mathscr{D}}\Bigg(\sum_{\substack{I:I^{(r)}=H\\ \operatorname{dist}(I,H^c)\geq \frac{1}{4}\ell(H)}} D_I b\Bigg)\langle f\rangle_H - \Bigg(\sum_{H\in\mathscr{D}} D_H^{(r,\text{good})} b\Bigg) E_{-\infty} f \\
&=: \Pi_b^1 f - \Pi_b^2 f.
\end{aligned}
$$

These are a standard part of our decomposition, which have been studied in exactly the same form in [20], where the following result is proved.

PROPOSITION 5.1 ([20], Theorem 7.1). *For any $\lambda > 1$, we have*

$$\|\Pi_b^1 f\|_2 \le C\|b\|_{BMO_\lambda^2(\mu)}\|f\|_2,$$

where, for $p \in [1, \infty)$,

$$\|b\|_{BMO_\lambda^p(\mu)} := \sup_Q \inf_{a\in\mathbb{C}}\left(\frac{1}{\mu(\lambda Q)}\int_Q |f-a|^p\,d\mu\right)^{1/p},$$

and the supremum is over all cubes Q in $\mathbb{R}^d$ (including $Q = \lambda Q = \mathbb{R}^d$ if $\mu(\mathbb{R}^d) < \infty$).

For the part Π_b^2, the same estimate is easy.

LEMMA 5.2.

$$\|\Pi_b^2 f\|_2 \le C\|b\|_{BMO_\lambda^2(\mu)}\|f\|_2.$$

Proof. Note that this term is only nonzero if $\mu(\mathbb{R}^d) < \infty$, and in this case, noting that $D_H^{(r,\mathrm{good})}a = 0$ for any constant a,

$$\begin{aligned}\|\Pi_b^2 f\|_2 &= \Big\|\sum_H D_H^{(r,\mathrm{good})}(b-a)\Big\|_2 |E_{-\infty} f| \\ &\le \|b-a\|_2\cdot\mu(\mathbb{R}^d)^{-1/2}\|f\|_2 \le \|b\|_{BMO_\lambda^2(\mu)}\|f\|_2\end{aligned}$$

by a suitable choice of a. □

Thus, it remains to check the BMO conditions on $b \in \{T1, T^*1\}$ under our assumptions on the operator T. This is also reasonably standard and contained in the following:

PROPOSITION 5.3. *Under the assumptions of Theorem 1.1 and the standard Dini condition (1.1) with $\alpha = 0$, we have $T1 \in BMO_\lambda^2$ and $T^*1 \in BMO_\lambda^2$ for any $\lambda > 1$.*

Proof. We are going to prove that $T1 \in BMO_\lambda^2$ and the case of the adjoint is similar. Fix a cube Q and some $\lambda > 1$. We denote by x_Q the center of the cube Q.

Note that we have $|x - x_Q| < \frac{1}{2}\ell(Q) \le \frac{1}{2}|y-x|$ for $x \in Q$ and $y \in (2Q)^c$ (using the ℓ^∞ metric on $\mathbb{R}^d$ for convenience). So, for $x \in Q$ and $\tau \ge 2$, we have

$$\begin{aligned}&|T1_{(\tau Q)^c}(x) - T1_{(\tau Q)^c}(x_Q)| \\ &\quad\le \int_{(\tau Q)^c} |K(x,y) - K(x_Q,y)|\,d\mu(y) \\ &\quad\le C\int_{(2Q)^c}\omega\left(\frac{|x-x_Q|}{|y-x_Q|}\right)\frac{1}{|y-x_Q|^n}\,d\mu(y)\end{aligned}$$

$$\le C\sum_{k=1}^{\infty}\int_{2^{k+1}Q\setminus 2^kQ}\omega\left(\frac{2^{-1}\ell(Q)}{2^{k-1}\ell(Q)}\right)\frac{1}{(2^{k-1}\ell(Q))^n}\,d\mu(y)$$
$$\le C\sum_{k=1}^{\infty}\omega(2^{-k})\frac{\mu(2^{k+1}Q)}{(2^{k-1}\ell(Q))^n}$$
$$\le C\sum_{k=1}^{\infty}\omega(2^{-k})\le C\int_0^1\omega(t)\frac{dt}{t}.$$

Additionally, if $x\in Q$ and $\lambda\in(1,2)$, then

$$|T1_{2Q\setminus\lambda Q}(x)|\le\int_{2Q\setminus\lambda Q}|K(x,y)|\,d\mu(y)\le C\int_{2Q\setminus\lambda Q}\frac{1}{|x-y|^n}\,d\mu(y)$$
$$\le C\int_{2Q\setminus\lambda Q}\frac{1}{(\lambda-1)^n\ell(Q)^n}\,d\mu(y)\le C\frac{1}{(\lambda-1)^n}.$$

Set $\tau=\max\{2,\lambda\}$ and $a_Q=T1_{(\tau Q)^c}(x_Q)$. Then

$$\int_Q|T1(x)-a_Q|^2\,d\mu(x)$$
$$\le C\int_{\mathbb{R}^d}|T1_{\lambda Q}(x)|^2\,d\mu(x)+C\int_Q|T1_{\tau Q\setminus\lambda Q}(x)|^2\,d\mu(x)$$
$$+C\int_Q|T1_{(\tau Q)^c}(x)-T1_{(\tau Q)^c}(x_Q)|^2\,d\mu(x)$$
$$\le C\mu(\lambda Q)+C\frac{1}{(\lambda-1)^n}\mu(Q)+C\left(\sum_{k=0}^{\infty}\omega(2^{-k})\right)^2\mu(Q)$$
$$\le C\mu(\lambda Q),$$

where the middle term of the decomposition is equal to zero for $\lambda\ge 2$. □

5.2. The Operator R_1

We have

$$R_1=\sum_H B_H^{(1)},$$

where the operator $B_H^{(1)}$ is given by

$$\pi_{\text{good}}\omega\left(\frac{1}{2}\right)B_H^{(1)}f:=P_H^{(r+1)}TD_H^{(r,\text{good})}f=\sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}}\sum_{J:J\subseteq H\subseteq J^{(r)}}D_JTD_If$$
$$=\sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}}\sum_{J:J\subseteq H\subseteq J^{(r)}}\sum_{i,j}\langle T\varphi_I^i,\varphi_J^j\rangle\langle\varphi_I^i,f\rangle\varphi_J^j. \tag{5.1}$$

LEMMA 5.4.

$$\|T\varphi_I^i\|_2\le C.$$

Proof. Since φ_I^i takes a constant value $\langle\varphi_I^i\rangle_{I'}$ on each $I' \in \mathrm{ch}(I)$, it follows from the testing condition $\|T1_I\|_2 \le C\mu(I)^{1/2}$ that

$$\begin{aligned}
\|T\varphi_I^i\|_2 &= \Big\| T \sum_{I'\in\mathrm{ch}(I)} \langle\varphi_I^i\rangle_{I'} 1_{I'} \Big\|_2 \\
&\le \sum_{I'\in\mathrm{ch}(I)} |\langle\varphi_I^i\rangle_{I'}| \|T1_{I'}\|_2 \\
&\le C \sum_{I'\in\mathrm{ch}(I)} |\langle\varphi_I^i\rangle_{I'}| \mu(I')^{1/2} \\
&\le C2^{d/2} \Big(\sum_{I'\in\mathrm{ch}(I)} |\langle\varphi_I^i\rangle_{I'}|^2 \mu(I') \Big)^{1/2} \\
&\le C\|\varphi_I^i\|_2 = C,
\end{aligned}$$

where we incorporated the dimensional factor $2^{d/2}$ into C. □

It follows that

$$\begin{aligned}
|\langle B_H^{(1)} f, g\rangle| &\le C \sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}} \sum_{J:J\subseteq H\subseteq J^{(r)}} \sum_{i,j} |\langle\varphi_I^i, f\rangle||\langle\varphi_J^j, g\rangle| \\
&\le C \Big(\sum_{I:I^{(r)}=H} \sum_i |\langle\varphi_I^i, f\rangle|^2 \Big)^{1/2} \Big(\sum_{J:J\subseteq H\subseteq J^{(r)}} \sum_j |\langle\varphi_I^i, f\rangle|^2 \Big)^{1/2} \\
&= C\|D_H^{(r)} f\|_2 \|P_H^{(r+1)} g\|_2
\end{aligned}$$

using the Cauchy–Schwarz inequality and the fact that the total number of terms is bounded by a dimensional constant. Hence

$$\begin{aligned}
|\langle R_1 f, g\rangle| &\le \sum_H |\langle B_H^{(1)} f, g\rangle| \\
&\le C \Big(\sum_H \|D_H^{(r)} f\|_2^2 \Big)^{1/2} \Big(\sum_H \|P_H^{(r+1)} g\|_2 \Big)^{1/2} \\
&\le C\|f\|_2 \sqrt{r+1} \|g\|_2 \le C\|f\|_2\|g\|_2,
\end{aligned}$$

incorporating the fixed constant $\sqrt{1+r}$ into C.

By (5.1) the kernel $b_H^{(1)}$ of $B_H^{(1)}$ is given by

$$b_H^{(1)}(x, y) = \frac{1}{\pi_{\mathrm{good}}\omega(\frac{1}{2})} \sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}} \sum_{J:J\subseteq H\subseteq J^{(r)}} \sum_{i,j} \langle T\varphi_I^i, \varphi_J^j\rangle \varphi_I^i(y)\varphi_J^j(x).$$

In general, there are no good pointwise bounds for this expression; however, if μ is a doubling measure, then

$$|\varphi_I^i(y)| \le \frac{C}{\mu(I)^{1/2}} 1_I(y) \le \frac{C}{\mu(H)^{1/2}} 1_I(y),$$

with a similar bound for $|\varphi_J^j(x)|$. Thus

$$\begin{aligned}|b_H^{(1)}(x,y)| &\leq C \sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}} \frac{1_I(y)}{\mu(H)^{1/2}} \sum_{J:J\subseteq H\subseteq J^{(r)}} \frac{1_J(x)}{\mu(H)^{1/2}} \sum_{i,j} 1\\ &\leq \frac{C}{\mu(H)} 1_H(x) 1_H(y).\end{aligned}$$

5.3. *The Operators R_k, $k \geq 2$*

For the analysis of R_k (and further of Q_k), it is convenient to take the new summation variable $K := H^{(k)}$, which is a common ancestor of both H and $H\dot{+}m$ for k-good H and $|m| \sim 2^{k-2}$. This leads to the decomposition

$$R_k f = \sum_K \frac{1}{\pi_{\text{good}}} \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good}\\ H^{(k)}=K}} \frac{1}{\omega(2^{-k})} P_{H\dot{+}m}^{(r+1)} T D_H^{(r,\text{good})} f =: \sum_K B_K^{(k)} f,$$

where

$$B_K^{(k)} = \frac{1}{\pi_{\text{good}}} \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good}\\ H^{(k)}=K}} \sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}} \sum_{\substack{J:J\subseteq H\dot{+}m\\ \subseteq J^{(r)}}} \frac{1}{\omega(2^{-k})} D_J T D_I$$

has the kernel

$$b_K^{(k)}(x,y) = \frac{1}{\pi_{\text{good}}} \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good}\\ H^{(k)}=K}} \sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}} \sum_{\substack{J:J\subseteq H\dot{+}m\\ \subseteq J^{(r)}}} \sum_{i,j} \frac{\langle T\varphi_I^i, \varphi_J^j\rangle}{\omega(2^{-k})} \varphi_I^i(y)\varphi_J^j(x).$$

LEMMA 5.5. *In the last sum, we have*

$$\left|\frac{\langle T\varphi_I^i, \varphi_J^j\rangle}{\omega(2^{-k})} \varphi_I^i(y)\varphi_J^j(x)\right| \leq \frac{C}{\ell(K)^n} 1_I(y) 1_J(x).$$

Proof. Since I is r-good and $I^{(r)} = H$, it follows from the definition that $\operatorname{dist}(I, H^c) \geq \frac{1}{4}\ell(H)$. Since $J \subset H\dot{+}m$ and $|m| \sim 2^{k-2}$, we further deduce that $\operatorname{dist}(I, J) \gtrsim |m|\ell(H) \gtrsim 2^k \ell(H) = \ell(K)$. Denoting by c_I the center of I and using the vanishing integral of φ_I^i and kernel regularity, we obtain

$$\begin{aligned}|\langle T\varphi_I^i, \varphi_J^j\rangle| &= \left|\iint K(x,y)\varphi_I^i(y)\varphi_J^j(x)\,d\mu(y)\,d\mu(x)\right|\\ &= \left|\iint (K(x,y) - K(x,c_I))\varphi_I^i(y)\varphi_J^j(x)\,d\mu(y)\,d\mu(x)\right|\\ &\leq \iint C\omega\left(\frac{|y-c_I|}{|x-c_I|}\right)\frac{1}{|x-c_I|^d} |\varphi_I^i(y)|\,|\varphi_J^j(x)|\,d\mu(y)\,d\mu(x)\\ &\leq C\frac{\omega(2^{-k})}{\ell(K)^n} \|\varphi_I^i\|_1 \|\varphi_J^j\|_1.\end{aligned}$$

The proof is completed by recalling that φ_I^i is supported on I and

$$\|\varphi_I^i\|_1\|\varphi_I^i\|_\infty \le C,$$

with similar observations for φ_J^j. □

LEMMA 5.6. *The kernel $b_K^{(k)}$ of $B_K^{(k)}$ satisfies*

$$|b_K^{(k)}(x,y)| \le \frac{C}{\ell(K)^n} 1_K(x)1_K(y),$$

and hence

$$|\langle B_K^{(k)} f, g\rangle| \le \frac{C}{\ell(K)^n}\|1_K f\|_1\|1_K g\|_1 \le C\|f\|_2\|g\|_2.$$

Proof. From the previous lemma we conclude that

$$\begin{aligned}
|b_K^{(k)}(x,y)| &\le \frac{C}{\ell(K)^n} \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good}\\ H^{(k)}=K}} \sum_{\substack{I\ r\text{-good}\\ I^{(r)}=H}} \sum_{\substack{J: J\subseteq H\dot{+}m\\ \subseteq J^{(r)}}} 1_I(y)1_J(x)\\
&\le \frac{C}{\ell(K)^n} \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good}\\ H^{(k)}=K}} 1_H(y)1_{H\dot{+}m}(x),
\end{aligned}$$

since the cubes $I : I^{(r)} = H$ are pairwise disjoint, and the cubes $J : J \subseteq H\dot{+}m \subseteq J^{(r)}$ have overlap at most $r+1$ times at any point, and we have incorporated this fixed constant into C.

Consider a fixed pair (x, y). Then there is at most one H such that $H^{(k)} = K$ and $H \ni y$. For this H, there is at most one integer m such that $H\dot{+}m \ni x$. So altogether there is at most one nonzero summand. Since both H and $H\dot{+}m$ are subsets of K, the nonzero summand can only exist if $(x, y) \in K \times K$. This proves the required kernel bound.

In the operator bound, the first estimate is immediate, and the second one follows by $\|1_K f\|_1 \le \mu(K)^{1/2}\|f\|_2$ (applied to g as well) and $\mu(K) \le C\ell(K)^n$. □

LEMMA 5.7.

$$|\langle R_k f, g\rangle| \le C\|f\|_2\|g\|_2.$$

Proof. In the summation defining $B_K^{(k)}$, we observe that $I^{(r)} = H$ and $H^{(k)} = K$, and hence $I^{(r+k)} = K$, and $J^{(j)} = H\dot{+}m$ for some $j = 0, \ldots, r$ and $(H\dot{+}m)^{(k)} = K$, so that $J^{(j+k)} = K$ for some $j = 0, \ldots, r$. It follows that

$$\begin{aligned}
B_K^{(k)} &= \sum_{j=0}^{r} D_K^{(k+j)} B_K^{(k)} D_K^{(k+r)}\\
&= (P_K^{(k+r+1)} - P_K^{(k)}) B_K^{(k)} D_K^{(k+r)},
\end{aligned}$$

and hence

$$
\begin{aligned}
|\langle R_k f, g\rangle| &\le \sum_K |\langle (B_K^{(k)} D_K^{(k+r)} f, (P_K^{(k+r+1)} - P_K^{(k)})g\rangle| \\
&\le C\Big(\sum_K \|D_K^{(k+r)} f\|_2^2\Big)^{1/2}\Big(\sum_K \|(P_K^{(k+r+1)} - P_K^{(k)})g\|_2^2\Big)^{1/2} \\
&\le C\|f\|_2\Big(\sum_{j=0}^{r}\sum_K \|D_K^{(k+j)} g\|_2^2\Big)^{1/2} \\
&\le C\|f\|_2\sqrt{r+1}\|g\|_2 \\
&\le C\|f\|_2\|g\|_2. \qquad \square
\end{aligned}
$$

5.4. *The Operators Q_k*

Noting that $Q_1 := 0$ trivially satisfies the required estimates, we concentrate on Q_k with $k \ge 2$.

Taking the new summation variable $K := H^{(k)}$ as for R_k, we are led to the decomposition

$$
\begin{aligned}
Q_k f &= \sum_K \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good} \\ H^{(k)}=K}} \frac{\langle T D_H^{(r,\text{good})} f, 1_{H\dot{+}m}\rangle}{\pi_{\text{good}}^2 \omega(2^{-k})}\left(\frac{1_{H\dot{+}m}}{\mu(H\dot{+}m)} - \frac{1_H}{\mu(H)}\right) \\
&=: \sum_K A_K^{(k)} f.
\end{aligned}
$$

LEMMA 5.8.

$$A_K^{(k)} = P_K^{(k+1)} A_K^{(k)} D_K^{(k+r)}.$$

Proof. Expanding

$$D_H^{(r,\text{good})} = \sum_{\substack{I\ r\text{-good} \\ I^{(r)}=H}} D_I$$

and noting that $I^{(r)} = H$ and $H^{(k)} = K$ imply $I^{(r+k)} = K$, the identity $A_K^{(k)} = A_K^{(k)} D_K^{(k+r)}$ is immediate.

Concerning the postcomposition of $A_K^{(k)}$ by $P_K^{(k+1)}$, we make the following observations. First, $A_K^{(k)} f$ is supported on K, which again depends on the fact that the k-goodness of H, together with $|m| \sim 2^{k-2}$, implies that $H\dot{+}m$ is contained in $H^{(k)} = K$. Second, $A_K^{(k)} f$ is constant on the kth-order descendants of K, which is immediate from its expression as a superposition of the relevant indicator functions. Third, $\int A_K^{(k)} f \, d\mu = 0$, which is also immediate from the fact that this property clearly holds for each of the summands,

$$\int \left(\frac{1_{H\dot{+}m}}{\mu(H\dot{+}m)} - \frac{1_H}{\mu(H)}\right) d\mu = 0.$$

These three properties characterize the range of the projection $P_K^{(k+1)}$, and hence $A_K^{(k)} = P_K^{(k+1)} A_K^{(k)}$. □

The kernel of $A_K^{(k)}$ is given by

$$\begin{aligned} &a_K^{(k)}(x,y) \\ &\quad = \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good} \\ H^{(k)}=K}} \sum_{\substack{I\ r\text{-good} \\ I^{(r)}=H}} \sum_i \frac{\langle T\varphi_I^i, 1_{H\dot{+}m}\rangle}{\pi_{\mathrm{good}}^2\omega(2^{-k})} \varphi_I^i(y) \bigg(\frac{1_{H\dot{+}m}(x)}{\mu(H\dot{+}m)} - \frac{1_H(x)}{\mu(H)} \bigg) \\ &\quad =: a_{K,1}^{(k)}(x,y) - a_{K,2}^{(k)}(x,y), \end{aligned}$$

where the last two kernels are defined in a natural way by taking all the terms of the form $1_{H\dot{+}m}(x)/\mu(H\dot{+}m)$ to the first one and all those of the form $1_H(x)/\mu(H)$ to the second.

LEMMA 5.9. *In the last sum, we have*

$$\bigg| \frac{\langle T\varphi_I^i, 1_{H\dot{+}m}\rangle}{\omega(2^{-k})} \varphi_I^i(y) \bigg| \le \frac{C}{\ell(K)^n} \mu(H\dot{+}m) 1_I(y).$$

Proof. By essentially the same considerations as in the beginning of the proof of Lemma 5.5, we check that $\mathrm{dist}(I, H\dot{+}m) \gtrsim 2^k \ell(H) = \ell(K)$. As in the same proof, using the vanishing integral of φ_I^i to insert $K(x, c_I)$, where c_I is the center of I, we have

$$\begin{aligned} |\langle T\varphi_I^i, 1_{H\dot{+}m}\rangle| &= \bigg| \iint K(x,y)\varphi_I^i(y) 1_{H\dot{+}m}(x)\, d\mu(y)\, d\mu(x) \bigg| \\ &= \bigg| \iint (K(x,y) - K(x,c_I))\varphi_I^i(y) 1_{H\dot{+}m}(x)\, d\mu(y)\, d\mu(x) \bigg| \\ &\le \iint \omega\bigg(\frac{|y-c_I|}{|x-c_I|} \bigg) \frac{C}{|x-c_I|^d} |\varphi_I^i(y)|\, 1_{H\dot{+}m}(x)\, d\mu(y)\, d\mu(x) \\ &\le C\omega(2^{-k}) \frac{\|\varphi_I^i\|_1 \mu(H\dot{+}m)}{\ell(K)^n}. \end{aligned}$$

The proof is completed by recalling that φ_I^i is supported on I and using the estimate $\|\varphi_I^i\|_1 \|\varphi_I^i\|_\infty \le C$. □

LEMMA 5.10.

$$|a_{K,1}^{(k)}(x,y)| \le \frac{C}{\ell(K)^n} 1_K(x) 1_K(y).$$

Proof. From the previous lemma and the disjointness of I with $I^{(r)} = H$ we have

$$|a_{K,1}^{(k)}(x,y)| \le \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good} \\ H^{(k)}=K}} \frac{C}{\ell(K)^n} 1_H(y) 1_{H\dot{+}m}(x).$$

For a pair (x, y), there is at most one H such that $H^{(k)} = K$ and $H \ni y$, and fixing this H, there is at most one integer m such that $H \dot{+} m \ni x$. Moreover, $(x, y) \in K \times K$ is a necessary condition for the existence of such H and m. This proves the asserted bound. □

LEMMA 5.11.

$$|a_{K,2}^{(k)}(x,y)| \le C \sum_{H:H^{(k)}=K} \frac{1_H(x)1_H(y)}{\mu(H)}.$$

Proof. By the same initial considerations as in the previous lemma, we have

$$|a_{K,2}^{(k)}(x,y)| \le \sum_{|m|\sim 2^{k-2}} \sum_{\substack{H\ k\text{-good} \\ H^{(k)}=K}} \frac{C}{\ell(K)^n} \mu(H \dot{+} m) 1_H(y) \frac{1_H(x)}{\mu(H)}.$$

For each fixed H, we observe that the cubes $H \dot{+} m$, with $|m| \sim 2^{k-2}$ are pairwise disjoint and contained in K, and thus

$$\sum_{|m|\sim 2^{k-2}} \mu(H \dot{+} m) \le \mu(K) \le C\ell(K)^n.$$

□

LEMMA 5.12.

$$|\langle A_K^{(k)} f, g\rangle| \le C\|f\|_2\|g\|_2.$$

Proof. Based on the kernel bounds for the two part of $A_K^{(k)}$, it follows that

$$|\langle A_K^{(k)} f, g\rangle| \le \frac{C}{\ell(K)^n}\|1_K f\|_1\|1_K g\|_1 + C \sum_{H:H^{(k)}=K} \frac{\|1_H f\|_1\|1_H g\|_1}{\mu(H)}.$$

The first term has the desired bound by the easy argument that we already gave in Lemma 5.6. Concerning the second term, two applications of the Cauchy–Schwarz inequality give

$$\begin{aligned} \sum_{H:H^{(k)}=K} \frac{\|1_H f\|_1\|1_H g\|_1}{\mu(H)} &\le \sum_{H:H^{(k)}=K} \|1_H f\|_2\|1_H g\|_2 \\ &\le \Big(\sum_{H:H^{(k)}=K} \|1_H f\|_2^2\Big)^{1/2} \Big(\sum_{H:H^{(k)}=K} \|1_H g\|_2^2\Big)^{1/2} \\ &= \|1_K f\|_2\|1_K g\|_2, \end{aligned}$$

and this completes the proof. □

LEMMA 5.13.

$$|\langle Q_k f, g\rangle| \le C\sqrt{k}\|f\|_2\|g\|_2.$$

Proof.

$$\begin{aligned}
|\langle Q_k f, g\rangle| &\leq \sum_K |\langle A_K D_K^{(k+r)} f, P_K^{(k+1)} g\rangle| \\
&\leq C\Big(\sum_K \|D_K^{(k+r)} f\|_2^2\Big)^{1/2}\Big(\sum_K \|P_K^{(k+1)} g\|_2^2\Big)^{1/2} \\
&\leq C\|f\|_2\Big(\sum_{j=0}^{k}\sum_K \|D_K^{(j)} g\|_2^2\Big)^{1/2} \\
&\leq C\|f\|_2\sqrt{k+1}\|g\|_2 \leq C\sqrt{k}\|f\|_2\|g\|_2,
\end{aligned}$$

using $k \geq 2$ in the last step. □

We have now completed the proof of Theorem 1.1, except for the last claim concerning the representation of the new operators Q_k and R_k as dyadic shifts. Note that this is already enough for the deduction of Corollary 1.2.

6. The Shift Structure of the New Operators

A dyadic shift of type (u, v) can be defined as an operator of the form

$$S_{u,v} = \sum_{K\in\mathscr{D}} C_K^{(u,v)},$$

where the operators $C_K^{(u,v)}$ satisfy the orthogonality property

$$C_K^{(u,v)} = D_K^{(u)} C_K^{(u,v)} D_K^{(v)},$$

and their kernels $c_K^{(u,v)}$ have the pointwise bound

$$|c_K^{(u,v)}(x, y)| \leq C\frac{1_K(x)1_K(y)}{\mu(K)}.$$

Since

$$D_K^{(u)} f = \sum_{I:I^{(u)}=K}\sum_i \varphi_I^i\langle\varphi_I^i, f\rangle$$

and $\|\varphi_I^i\|_1 \leq C\mu(I)^{1/2}$, this is easily seen to coincide with the definition in terms of Haar functions that is used in several papers.

6.1. *The Operators R_k*

We concentrate on $k \geq 1$, since the case of negative k is similar and essentially dual to this one. (Note that the adjoint of a dyadic shift of type (u, v) is a shift of type (v, u).) Then

$$R_k = \sum_{K\in\mathscr{D}} B_K^{(k)}, \qquad B_K^{(k)} = \sum_{j=k}^{k+r} D_K^{(j)} B_K^{(k)} D_K^{(k+r)},$$

and the kernel $b_K^{(k)}$ of $B_K^{(k)}$ satisfies

$$|b_K^{(k)}(x,y)| \le C\frac{1_K(x)1_K(y)}{\ell(K)^n} \le C\frac{1_K(x)1_K(y)}{\mu(K)}.$$

Since $D_K^{(j)}$ is a difference of averaging operators (or by the pointwise bounds for the Haar functions), it follows that the kernel of $D_K^{(j)}B_K^{(k)}$ satisfies the same bound. Thus, letting

$$C_K^{(j,k+r)} := D_K^{(j)}B_K^{(k)}, \qquad S_{j,k+r} := \sum_K C_K^{(j,k+r)},$$

it is immediate that $S_{j,k+r}$ is a shift of type $(j,k+r)$ and

$$R_k = \sum_{j=k}^{k+r} S_{j,k+r}$$

is a sum of $r+1$ shifts of complexity at most $k+r$.

6.2. *The Operators Q_k*

For $k \ge 1$ again, recall that

$$Q_k = \sum_{K\in\mathscr{D}} A_K^{(k)}, \qquad A_K^{(k)} = \sum_{j=0}^{k} D_K^{(j)}A_K^{(k)}D_K^{(k+r)},$$

where the kernel $a_K^{(k)}$ of $A_K^{(k)}$ satisfies

$$|a_K^{(k)}(x,y)| \le C\frac{1_K(x)1_K(y)}{\ell(K)^n} + C\sum_{H:H^{(k)}=K}\frac{1_H(x)1_H(y)}{\mu(H)}.$$

Then we can split

$$A_K^{(k)} = A_K^{(k.1)} + \sum_{H:H^{(k)}=K} A_H^{(k,2)},$$

where the corresponding kernels $a_K^{(k,1)}$ and $a_H^{(k,2)}$ satisfy

$$|a_K^{(k,1)}(x,y)| \le C\frac{1_K(x)1_K(y)}{\mu(K)}, \qquad |a_H^{(k,2)}(x,y)| \le C\frac{1_H(x)1_H(y)}{\mu(H)}.$$

Thus

$$Q_k = \sum_{j=0}^{k}\sum_K D_K^{(j)}\Big(A_{K,1}^{(k)} + \sum_{H:H^{(k)}=K} A_H^{(k,2)}\Big)D_K^{(k+r)}, \tag{6.1}$$

and it is immediate that each

$$\sum_K D_K^{(j)}A_{K,1}^{(k)}D_K^{(k+r)}$$

is a dyadic shift of order $(j,k+r)$. We will prove the following:

LEMMA 6.1.

$$\sum_K D_K^{(j)} \sum_{H:H^{(k)}=K} A_H^{(k,2)} D_K^{(k+r)}$$

is a dyadic shift of order $(0, k-j+r)$.

Once this is proved, it follows from (6.1) that Q_k is a sum of $2(k+1)$ shifts of complexity at most $k+r$.

Proof of Lemma 6.1. As written, the series looks formally more like a shift of order $(j, k+r)$, but the kernel of $\sum_{H:H^{(k)}=K} A_H^{(k,2)}$ does not have a correct bound. Thus we need to reorganize the summation:

$$\begin{aligned}
&\sum_K D_K^{(j)} \sum_{H:H^{(k)}=K} A_H^{(k,2)} D_K^{(k+r)} \\
&\quad= \sum_K \Big(\sum_{J:J^{(j)}=K} D_J \Big)\Big(\sum_{L:L^{(j)}=K} \sum_{H:H^{(k-j)}=L} A_H^{(k.2)} \Big)\Big(\sum_{M:M^{(j)}=K} D_M^{(k-j+r)} \Big) \\
&\quad= \sum_J D_J \sum_{H:H^{(k-j)}=J} A_H^{(k,2)} D_J^{(k-j+r)},
\end{aligned}$$

using support considerations to see that only the terms with $M = L = J$ are nonzero and simplifying $\sum_K \sum_{J:J^{(k)}=K} = \sum_J$.

Now, the right-hand side has the structure of a shift of order $(0, k-j+r)$, and we only need to verify the bound for the kernel. The kernel of $D_J \sum_H A_H^{(k,2)}$ is given by

$$\begin{aligned}
&\Big| \sum_i \varphi_J^i(x) \int \varphi_J^i(z) \sum_{H:H^{(k-j)}=J} a_H^{(k,2)}(z,y)\, d\mu(z) \Big| \\
&\qquad\le C \sum_i \frac{1_J(x)}{\mu(J)^{1/2}} \sum_{H:H^{(k-j)}=J} \int \frac{1_J(z)}{\mu(J)^{1/2}} \frac{1_H(z) 1_H(y)}{\mu(H)}\, d\mu(z) \\
&\qquad\le C \frac{1_J(x)}{\mu(J)} \sum_{H:H^{(k-j)}=J} 1_H(y) = C \frac{1_J(x) 1_J(y)}{\mu(J)},
\end{aligned}$$

which is the correct bound. Since $D_J^{(k-j+r)}$ is a difference of averaging operators (or from the bounds for Haar functions), we find that the kernel of $D_J \sum_H A_H^{(k,2)} D_J^{(k-j+r)}$ has the same bound. This completes the proof. □

Now we have completed the proof of all claims in Theorem 1.1 and therefore also the proof of Corollary 1.3.

7. The Weak-Type Bounds

In this final section, we complete the proof of Theorem 3.1 by proving the asserted weak-type estimates in the case that μ is a doubling measure.

PROPOSITION 7.1. *Let μ be a doubling measure on $\mathbb{R}^d$, and let U_k be an operator of the form*

$$U_k = \sum_{K\in\mathscr{D}} V_K^{(k)}, \qquad V_K^{(k)} = 1_K V_K^{(k)} P_K^{(k+r+1)}.$$

Suppose further that

$$\|V_K^{(k)}\|_{L^1(\mu)\to L^1(\mu)} \le C, \qquad \|U_k\|_{L^2(\mu)\to L^2(\mu)} \le C\sqrt{k}.$$

Then

$$\|U_k\|_{L^1(\mu)\to L^{1,\infty}(\mu)} \le Ck.$$

Observe that each of the operators Q_k, R_k, Q_k^*, R_k^* is of the form $U_{|k|}$ considered in the proposition.

Proof. Let $f \in L^1(\mu)$ and $\lambda > 0$. We make the usual Calderón–Zygmund decomposition: Consider the maximal cubes $J \in \mathscr{D}$ such that $\langle |f|\rangle_J > \lambda$; then by the doubling property $\langle |f|\rangle_J \le C\lambda$. Thus the good part

$$g := \sum_J \langle f\rangle_J 1_J + f 1_{\Omega^c}, \quad \Omega := \bigcup J,$$

satisfies $\|g\|_\infty \le C\lambda$, $\|g\|_1 \le \|f\|_1$, and hence $\|g\|_2^2 \le C\lambda\|f\|_1$. It follows that

$$\mu(\{|U_k g| > \lambda\}) \le \lambda^{-2}\|U_k g\|_2^2 \le C\lambda^{-2}\big(\sqrt{k}\|g\|_2\big)^2 \le Ck\lambda^{-1}\|f\|_1.$$

Also, the set $\Omega := \bigcup J$ satisfies $\mu(\Omega) \le \lambda^{-1}\|f\|_1$. So it remains to estimate

$$\mu(\{|U_k b| > \lambda\} \setminus \Omega), \quad b := \sum_J b_J := \sum_J (f - \langle f\rangle_J)1_J.$$

We have

$$\mu(\{|U_k b| > \lambda\} \setminus \Omega) \le \frac{1}{\lambda}\int_{\Omega^c} |U_k b|\, d\mu \le \frac{1}{\lambda}\sum_J \int_{J^c} |U_k b_J|\, d\mu$$

and

$$1_{J^c} U_k b_J = 1_{J^c}\sum_K V_K^{(k)} b_J = 1_{J^c}\sum_{K\supsetneq J} V_K^{(k)} b_J,$$

since K must intersect both J and J^c to get a nonzero contribution. Moreover, noting that

$$V_K^{(k)} b_J = V_K^{(k)} P_K^{(k+r+1)} b_J = V_K^{(k)}\Bigg(\sum_{I: I^{(k+r+1)}=K} 1_I\langle b_J\rangle_I - 1_K\langle b_J\rangle_K\Bigg)$$

and recalling that $\int_J b_J\, d\mu = 0$, we find that $\langle b_J\rangle_I = 0$ for all dyadic cubes I with $\ell(I) \ge \ell(J)$. So the only nonzero contribution can arise if $\ell(I) = 2^{-k-r-1}\ell(K) < \ell(J)$. In combination with $J \subsetneq K$, this implies that $J \subsetneq K \subseteq J^{(k+r)}$. So altogether

$$\begin{aligned}\int_{J^c} |U_k b_J|\, d\mu &\le \sum_{K: J\subsetneq K\subseteq J^{(k+r)}} \|V_K^{(k)} b_J\|_1 \\ &\le \sum_{K: J\subsetneq K\subseteq J^{(k+r)}} C\|b_J\|_1 = (k+r)C\|b_J\|_1 \le Ck\|b_J\|_1,\end{aligned}$$

since $k \geq 1$ and r is fixed. Summing over J, we conclude that

$$\mu(\{|U_k b| > \lambda\} \setminus \Omega) \leq \frac{1}{\lambda} \sum_J \int_{J^c} |U_k b_J| \, d\mu \leq \frac{1}{\lambda} \sum_J Ck \|b_J\|_1 \leq \frac{Ck}{\lambda} \|f\|_1,$$

which completes the proof. □

References

[1] O. V. Beznosova, *Linear bound for the dyadic paraproduct on weighted Lebesgue space $L_2(w)$,* J. Funct. Anal. 255 (2008), no. 4, 994–1007.

[2] L. Dalenc and Y. Ou, *Upper bound for multi-parameter iterated commutators,* Publ. Mat. 60 (2016), no. 1, 191–220.

[3] G. David and J.-L. Journé, *A boundedness criterion for generalized Calderón–Zygmund operators,* Ann. of Math. (2) 120 (1984), no. 2, 371–397.

[4] D. Deng, L. Yan, and Q. Yang, *Blocking analysis and $T(1)$ theorem,* Sci. China Ser. A 41 (1998), no. 8, 801–808.

[5] T. Figiel, *On equivalence of some bases to the Haar system in spaces of vector-valued functions,* Bull. Pol. Acad. Sci. Math. 36 (1988), no. 3–4, 119–131.

[6] ———, *Singular integral operators: a martingale approach,* Geometry of Banach spaces, Strobl, 1989, London Math. Soc. Lecture Note Ser., 158, pp. 95–110, Cambridge Univ. Press, Cambridge, 1990.

[7] T. P. Hytönen, *The sharp weighted bound for general Calderón–Zygmund operators,* Ann. of Math. (2) 175 (2012), no. 3, 1473–1506.

[8] ———, *The vector-valued nonhomogeneous Tb theorem,* Int. Math. Res. Not. IMRN 2 (2014), 451–511.

[9] ———, *Representation of singular integrals by dyadic operators, and the A_2 theorem,* Expo. Math. 35 (2017), no. 2, 166–205.

[10] ———, *The two-weight inequality for the Hilbert transform with general measures,* Proc. London Math. Soc. (2018, to appear), doi:10.1112/plms.12136.

[11] T. P. Hytönen, M. T. Lacey, H. Martikainen, T. Orponen, M. C. Reguera, E. T. Sawyer, and I. Uriarte-Tuero, *Weak and strong type estimates for maximal truncations of Calderón–Zygmund operators on A_p weighted spaces,* J. Anal. Math. 118 (2012), no. 1, 177–220.

[12] T. P. Hytönen, M. T. Lacey, and C. Pérez, *Sharp weighted bounds for the q-variation of singular integrals,* Bull. Lond. Math. Soc. 45 (2013), no. 3, 529–540.

[13] T. P. Hytönen and H. Martikainen, *Non-homogeneous Tb theorem and random dyadic cubes on metric measure spaces,* J. Geom. Anal. 22 (2012), no. 4, 1071–1107.

[14] M. T. Lacey, *An elementary proof of the A_2 bound,* Israel J. Math. 217 (2017), no. 1, 181–195.

[15] M. T. Lacey and H. Martikainen, *Local Tb theorem with L^2 testing conditions and general measures: Calderón–Zygmund operators,* Ann. Sci. Éc. Norm. Supér. (4) 49 (2016), no. 1, 57–86.

[16] A. K. Lerner, *On an estimate of Calderón–Zygmund operators by dyadic positive operators,* J. Anal. Math. 121 (2013), 141–161.

[17] ———, *A simple proof of the A_2 conjecture,* Int. Math. Res. Not. IMRN 14 (2013), 3159–3170.

[18] ———, *On pointwise estimates involving sparse operators,* New York J. Math. 22 (2016), 341–349.

[19] H. Martikainen, *Representation of bi-parameter singular integrals by dyadic operators,* Adv. Math. 229 (2012), no. 3, 1734–1761.

[20] F. Nazarov, S. Treil, and A. Volberg, *The Tb-theorem on non-homogeneous spaces,* Acta Math. 190 (2003), no. 2, 151–239.

[21] Y. Ou, *Multi-parameter singular integral operators and representation theorem,* Rev. Mat. Iberoam. 33 (2017), no. 1, 325–350.

[22] S. Treil, *Sharp A_2 estimates of Haar shifts via Bellman function,* Recent trends in analysis. Proceedings of the conference in honor of Nikolai Nikolski on the occasion of his 70th birthday, Bordeaux, France, August 31–September 2, 2011, pp. 187–208, The Theta Foundation, Bucharest, 2013.

[23] A. Volberg, *The proof of the nonhomogeneous $T1$ theorem via averaging of dyadic shifts,* St. Petersburg Math. J. 27 (2016), no. 3, 399–413.

A. G. de la Herrán
Department of Mathematics and Statistics
University of Helsinki
P.O. Box 68
FI-00014 Helsinki
Finland

anagrauherran@gmail.com

T. Hytönen
Department of Mathematics and Statistics
University of Helsinki
P.O. Box 68
FI-00014 Helsinki
Finland

tuomas.hytonen@helsinki.fi