OPERATOR-VALUED DYADIC SHIFTS AND THE $T(1)$ THEOREM

TIMO S. HÄNNINEN AND TUOMAS P. HYTÖNEN

Abstract. In this paper we extend dyadic shifts and the dyadic representation theorem to an operator-valued setting: We first define operator-valued dyadic shifts and prove that they are bounded. We then extend the dyadic representation theorem, which states that every scalar-valued Calderón–Zygmund operator can be represented as a series of dyadic shifts and paraproducts averaged over randomized dyadic systems, to operator-valued Calderón–Zygmund operators. As a corollary, we obtain another proof of the operator-valued, global $T1$ theorem.

We work in the setting of integral operators that have $R$-bounded operator-valued kernels and act on functions taking values in $UMD$-spaces. The domain of the functions is the Euclidean space equipped with the Lebesgue measure.

In addition, we give new proofs for the following known theorems: Boundedness of the dyadic (operator-valued) paraproduct, a variant of Pythagoras’ theorem for (vector-valued) functions adapted to a sparse collection of dyadic cubes, and a decoupling inequality for (UMD-valued) martingale differences.

CONTENTS

1. Introduction 2
2. Preliminaries 7
   2.1. Sum of stochastically independent conditional expectations 7
   2.2. Properties of $R$-bounds 8
   2.3. Pythagoras’ theorem for functions adapted to a sparse collection 8
3. Decoupling of the sum of martingale differences 11
4. Vector-valued dyadic shifts are bounded 16
5. Sufficient condition for the boundedness of dyadic paraproducts 17
6. Vector-valued dyadic representation theorem 21
   6.1. Expanding the dual pairing by means of dyadic shifts 21
   6.2. Estimating the $R$-bounds of the matrix elements 23
Acknowledgments 28
References 28

Date: December 2, 2014.

2010 Mathematics Subject Classification. Primary 42B20; Secondary 46E40.

Key words and phrases. operator-valued, vector-valued, dyadic shift, dyadic representation, paraproduct, $T1$, $UMD$, $R$-boundedness, decoupling, Pythagoras.
1. Introduction

In this paper we extend dyadic shifts and the dyadic representation theorem to an operator-valued setting. We work with integral operators that have $R$-bounded operator-valued kernels and act on functions taking values in $UMD$-spaces. The domain of the functions is the Euclidean space equipped with the Lebesgue measure.

First, we summarize what is known in the scalar-valued setting. A dyadic shift $S^{ji}$ with parameters $i$ and $j$ (and complexity $\max\{i, j\} + 1$) is defined by

$$S^{ji} f := \sum_{K \in \mathcal{D}} D^K f,$$

which involves the following ingredients:

- the *shifted Haar projection* $D^K$ associated with a dyadic cube $K \in \mathcal{D}$ is defined by
  $$D^K f := \sum_{I \in \mathcal{D}: I \subseteq K, \ell(I) = 2^{-i}(K)} D_I f,$$

- the *Haar projection* $D_I$ associated with a dyadic cube $I \in \mathcal{D}$ is defined by
  $$D_I f := \sum_{I' \in \text{child}(I)} (f)_{I'} 1_{I'} - (f)_{I} 1_{I} = \sum_{\eta \in (0, 1)^d \setminus \{0\}} \langle f, h^\eta \rangle h^\eta_I,$$

where $\text{child}(I)$ denotes the dyadic children of $I$, $(f)_I := \frac{1}{|I|} \int_I f(x) \, dx$, and $\{h^\eta_I\}_{\eta \in (0, 1)^d}$ are the Haar functions associated with $I$,

- the averaging operator $A_K$ is defined by
  $$A_K f(x) := \frac{1}{|K|} \int_K a_K(x, x') f(x') \, dx',$$

where it is assumed that the kernels satisfy $|a_K(x, x')| \leq 1$ for all $K \in \mathcal{D}$, $x \in K$, and $x' \in K$.

The *dyadic paraproduct* associated with a function $b : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$\Pi_b f := \sum_{Q \in \mathcal{D}} D_Q b(f)_Q = \sum_{Q \in \mathcal{D}} \sum_{\eta \in (0, 1)^d \setminus \{0\}} \langle b, h_Q^\eta \rangle (f)_{Q} h_Q^\eta.$$

Dyadic shifts are bounded on $L^p$ for $1 < p < \infty$. Indeed, by Pythagoras’ theorem, they are bounded on $L^2$, and, by using the Calderón–Zygmund decomposition, from $L^1$ to $L^{1,\infty}$. From the Marcinkiewicz interpolation theorem, it follows that dyadic shifts are bounded on $L^p$ for $1 < p \leq 2$, and hence, by duality, on $L^p$ for $2 \leq p < \infty$. The weak-$L^1$ bound with an exponential dependence on the complexity was proven by Lacey, Petermichl, and Reguera [12] and with a linear dependence by Hytönen [9]. It is a classical result that a dyadic paraproduct associated with a function $b$ is bounded on $L^p$ if and only if $b$ is a BMO function.

Dyadic shifts are dyadic model operators for Calderón–Zygmund operators: Petermichl [14 Lemma 2.1] proved that the Hilbert transform can be represented as a particular dyadic shift averaged over randomized dyadic systems, and Hytönen [9 Theorem 4.2] that every Calderón–Zygmund operator can be represented as a series of dyadic shifts and paraproducts averaged over randomized dyadic systems. The dyadic representation theorem for Calderón–Zygmund operators together with the boundedness of dyadic shifts and paraproducts yields another proof of the global $T1$ theorem for Calderón–Zygmund operators. For a detailed proof of the dyadic representation theorem, see the lecture notes on the $A_2$ theorem [8].
The operator-valued setting in this paper follows the by-now-usual paradigm of doing Banach-space valued harmonic analysis beyond Hilbert space: Orthogonality of vectors is replaced with unconditionality of martingale differences, and uniform boundedness of operators with $R$-boundedness. Pioneering examples of this are the result by Burkholder [2] and Bourgain [1] that the Hilbert transform is bounded on $L^p$ if and only if the Banach space $E$ has the UMD property, and the operator-valued Fourier multiplier theorems by Weis [17].

A family of operators $\mathcal{T} \subseteq \mathcal{L}(E, F)$ from a Banach space $(E, \cdot, |E|)$ to a Banach space $(F, \cdot, |F|)$ is said to be $R$-bounded if there exists a constant $\mathcal{R}_p(\mathcal{T})$ such that

$$\left( |E| \sum_{n=1}^{N} |\epsilon_n T_n e_n|^p \right)^{1/p} \leq \mathcal{R}_p(\mathcal{T}) \left( |E| \sum_{n=1}^{N} |e_n|^p \right)^{1/p}$$

for all choices of operators $(T_n)_{n=1}^{N} \subseteq \mathcal{T}$ and vectors $(e_n)_{n=1}^{N} \subseteq E$, where the expectation is taken over independent, unbiased random signs $(\epsilon_n)_{n=1}^{N}$. A Banach space $(E, \cdot, |E|)$ is said to be a UMD (unconditional martingale difference) space if there exists a constant $\beta_p(E)$ such that

$$\left\| \sum_{n=1}^{N} \epsilon_n d_n \right\|_{L^p(E)} \leq \beta_p(E) \left\| \sum_{n=1}^{N} d_n \right\|_{L^p(E)}$$

for all $E$-valued $L^p$-martingale difference sequences $(d_n)_{n=1}^{N}$ and for all choices of signs $(\epsilon_n)_{n=1}^{N} \subseteq \{-1, +1\}^N$. It is well-known that $R$-boundedness and UMD-property are independent (up to the involved constants) of the exponent $p \in (1, \infty)$: For an exposition on Banach-space-valued martingales, UMD spaces, and $R$-boundedness, among other things, see Neerven’s lecture notes [16].

We conclude the introduction by precisely fixing the operator-valued setting and stating the results. First, we define the operator-valued dyadic shifts and state their boundedness.

**Definition 1.1** (Operator-valued dyadic shift). Let $E$ be a UMD space. An operator-valued dyadic shift associated with parameter $j$ and $i$ is defined by

$$S_{ji} f := \sum_{K} D^j_K A_K D^i_K f$$

for every locally integrable function $f : \mathbb{R}^d \rightarrow E$, where, for each $K \in \mathcal{D}$, the averaging operator $A_K$ associated with an operator-valued kernel $a_K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathcal{L}(E)$ is defined by

$$A_K f(x) := \frac{1}{|K|} \int_{K} a_K(x, x') f(x') \, dx.$$  

The family of the operator-valued kernels is assumed to be $R$-bounded so that there exists a positive constant $\mathcal{R}_p(\{a\})$ such that

$$\mathcal{R}_p(\{a_K(x, x') \in \mathcal{L}(E) : K \in \mathcal{D}, \ x \in K, \ x' \in K'\}) \leq \mathcal{R}_p(\{a\}).$$

Let $L^p(\mathbb{R}^d; E)$ denote the Lebesgue–Bochner space, which is equipped with the norm

$$\|f\|_{L^p(\mathbb{R}^d; E)} = \left( \int_{\mathbb{R}^d} |f(x)|^p \, dx \right)^{1/p}.$$  

We prove the following theorem:
Theorem 1.2 (Operator-valued dyadic shifts are bounded). Let $1 < p < \infty$. Let $E$ be a UMD space. Let $S^{ij}$ be a dyadic shift with parameters $i$ and $j$ and associated with the operator-valued kernels $a_K$. Then

$$\|S^{ij} f\|_{L^p(\mathbb{R}^d, E)} \leq 4(\max\{i, j\} + 1) \mathcal{R}_p(\{j\}) \beta_p(E)^2 \|f\|_{L^p(\mathbb{R}^d, E)}$$

for all $f \in L^p(\mathbb{R}^d, E)$.

Next, we define the operator-valued Calderón–Zygmund operators and state the dyadic representation theorem for them. Following the paradigm of replacing orthogonality by unconditionality of martingale differences and uniform boundedness by $R$-boundedness, the standard estimates and the weak boundedness property are replaced by the Rademacher standard estimates and Rademacher weak boundedness property.

Definition 1.3 (Rademacher standard estimates). An operator-valued singular kernel $k : \mathbb{R}^d \times \mathbb{R}^d \setminus \{(x, x) : x \in \mathbb{R}^d\} \to \mathcal{L}(E)$ satisfies the Rademacher standard estimates if and only if:

(i) The kernel $k$ satisfies the decay estimate

$$\mathcal{R}\left(\{k(x, y)|x - y|^d : x \in \mathbb{R}^d, y \in \mathbb{R}^d \text{ with } x \neq y\}\right) \leq \mathcal{R}_{CZ_0}$$

for some constant $\mathcal{R}_{CZ_0}$.

(ii) The kernel $k$ satisfies the Hölder-type estimates

$$\mathcal{R}\left(\left\{\left(\frac{|x - y|}{|x - x'|}\right)^{\alpha} |x - y|^d : x \in \mathbb{R}^d, x' \in \mathbb{R}^d, y \in \mathbb{R}^d \text{ with } 0 < |x - x'| < \frac{1}{2}|x - y|\right\}\right) \leq \mathcal{R}_{CZ_0}$$

and

$$\mathcal{R}\left(\left\{\left(\frac{|x - y|}{|y - y'|}\right)^{\alpha} |x - y|^d : x \in \mathbb{R}^d, y \in \mathbb{R}^d, y' \in \mathbb{R}^d \text{ with } 0 < |y - y'| < \frac{1}{2}|x - y|\right\}\right) \leq \mathcal{R}_{CZ_0}$$

for some Hölder exponent $\alpha \in (0, 1]$ and for some constant $\mathcal{R}_{CZ_0}$.

Definition 1.4 (Rademacher weak boundedness property). An operator $T$ mapping locally integrable $E$-valued functions to locally integrable $E$-valued functions satisfies the Rademacher weak boundedness property if and only if

$$\mathcal{R}\left(\left\{\frac{1}{|I|} \int_{\mathbb{R}^d} 1_I(x) T(\cdot; 1_I)(x) \, dx \in \mathcal{L}(E) : I \in \mathcal{D}\right\}\right) \leq \mathcal{R}_{WBP}$$

for some constant $\mathcal{R}_{WBP}$.

The randomized dyadic systems are defined as follows. Let $\mathcal{D}^\omega$ designate the standard dyadic system. For every parameter $(\omega_j)_{j \in \mathbb{Z}} \in (\{0, 1\}^d)^\mathbb{Z} =: \Omega$ and every $I \in \mathcal{D}^0$, the translated dyadic cube $I + \omega$ is defined by

$$I + \omega := I + \sum_{j \in \mathcal{D}(I)} 2^{-j} \omega_j.$$
For each $\omega \in \Omega$, the translated dyadic system $D^{\omega}$ is defined by $D^{\omega} := \{I+\omega : I \in D^0\}$. We equip the parameter set with the natural probability measure: Each component $\omega_j \in \{0,1\}^d$ has an equal probability $2^{-d}$ of taking any of the $2^d$ values and all components are stochastically independent.

**Theorem 1.5** (Operator-valued dyadic representation theorem). Let $E$ be a Banach space. Let $T$ be a singular integral operator that satisfies the Rademacher weak boundedness property and whose operator-valued kernel satisfies the Rademacher standard estimates with the Hölder exponent $\alpha$. Assume that $T : L^p(\mathbb{R}^d; E) \to L^p(\mathbb{R}^d; E)$ is bounded. Then, for some dyadic shifts $S_{D^{\omega}}^{ij}$ and for the dyadic paraproducts $\Pi_{T_1}^{D^\omega}$ and $\Pi_{T_1}^{D^\omega}$, we have

$$\langle g, Tf \rangle = \mathbb{E}_\omega \left( C_T \sum_{i \geq 0, j \geq 0} 2^{(1/\epsilon)\alpha \max\{i,j\}} \langle g, S_{D^{\omega}}^{ij} f \rangle + \langle g, (\Pi_{T_1}^{D^\omega})^* f \rangle \right)$$

for all $g \in C_0^\alpha(\mathbb{R}^d; \mathbb{R}) \otimes E^*$ and $f \in C_0^\alpha(\mathbb{R}^d; \mathbb{R}) \otimes E$. Moreover,

$$C_T \lesssim_{\epsilon, \alpha} R_{CZ_\alpha} + R_{CZ_\alpha} + R_{WBP}.$$  

**Remark.** The statement contains an auxiliary parameter $\epsilon$ with $0 < \epsilon < 1$. The factor $2^{(1/\epsilon)\alpha \max\{i,j\}}$ can be replaced with the factor $(1+\max\{i,j\})^{\gamma(d+\alpha)2^{-\alpha \max\{i,j\}}}$. This is achieved by replacing the ‘boundary’ function $t \mapsto t^\gamma$ with the function $t \mapsto (1 + a^{-1}\log(t^{-1}))^{-\gamma}$ in the definition of a good dyadic cube, Definition 6.1 which then results in the decay $2^{-\alpha \max\{i,j\}}$ in the estimates for the matrix elements, Lemma 6.4 and Lemma 6.7. For the details, see the lecture notes on the $A_2$ theorem [8]. For simplicity, we use the function $t \mapsto t^\gamma$.

For a Banach space $(\mathcal{T}, \cdot | \cdot)$, the BMO$_p(\mathbb{R}^d, \mathcal{T})$-norm is defined by

$$\|b\|_{\text{BMO}_p(\mathbb{R}^d, \mathcal{T})} := \sup_{Q \subset \mathbb{R}^d} \left( \frac{1}{|Q|} \int_Q |b(x) - \langle b \rangle_Q |^p \, dx \right)^{1/p}.$$  

The following sufficient condition for the boundedness of the paraproduct $\Pi_b$ associated with an operator-valued function $b$ was proven by Hytönen [10] by using interpolation and decoupling of martingale differences. Predecessors of this operator-valued result (under stronger assumptions) were obtained by Hytönen and Weis [7], based on unpublished ideas of Bourgain recorded by Figiel and Wojtaszczuk [4] in the case of a scalar-valued function $b$.

**Theorem 1.6** (Sufficient conditions for the boundedness of a paraproduct). Let $E$ be a UMD space. Let $\mathcal{T} \subseteq \mathcal{L}(E)$ be a UMD subspace of $\mathcal{L}(E)$. Then

$$\|\Pi_b f\|_{L^p(\mathbb{R}^d; E)} \leq 6 \cdot 2^d p^\prime \beta_p(E)^2 \beta_p(\mathcal{T}) \|b\|_{\text{BMO}_p(\mathbb{R}^d, \mathcal{T})} \|f\|_{L^p(\mathbb{R}^d; E)}$$

for all $b \in \text{BMO}_p(\mathbb{R}^d, \mathcal{T})$ and $f \in L^p(\mathbb{R}^d; E)$.

In this paper we give a different proof of Theorem 1.6. This proof is elementary in that neither interpolation nor decoupling of martingale differences is used.

By combining Theorem 1.2, Theorem 1.5 and Theorem 1.6, we obtain a new proof for the following corollary, which is a special case of Hytönen’s vector-valued, non-homogeneous, global $Tb$ theorem [10] $Tb$ theorem 4.1. Earlier results of this type include the first vector-valued $T1$ theorem by Figiel [3], and the first operator-valued $T1$ theorem by Hytönen and Weis [7]. Several related results have appeared in the literature.
Corollary 1.7 (T1 theorem for operator-valued kernels). Let \( T \) be a singular integral operator that satisfies the Rademacher weak boundedness property and whose operator-valued kernel satisfies the Rademacher standard estimates. Assume that \( T^1 \in BMO_p(\mathbb{R}^d; \mathcal{T}) \) and \( T^*1 \in BMO_p(\mathbb{R}^d; \mathcal{T}^*) \) for some UMD subspaces \( \mathcal{T} \subset \mathcal{L}(E) \) and \( \mathcal{T}^* \subset \mathcal{L}(E^*) \). Then
\[
\|T\|_{L^p(\mathbb{R}^d; E)} \leq T_{d,p,\alpha} (R_{CZ_0} + R_{CZ_0} + R_{WBP} + \|T^1\|_{BMO_p(\mathbb{R}^d; \mathcal{T})} + \|T^*1\|_{BMO_p(\mathbb{R}^d; \mathcal{T}^*)}) \beta_p(E)^2.
\]

Here the condition \( T^1 \in BMO_p(\mathbb{R}^d; \mathcal{T}^*) \) is interpreted via duality as follows: There exists \( b \in BMO_p(\mathbb{R}^d; \mathcal{T}^*) \) such that \( \int_{\mathbb{R}^d} T(\cdot, h_1) dx = \int_{\mathbb{R}^d} b(x) h_1(x) dx \).

This interpretation originates from extracting the paraproducts as in the equation (6.1) in Section 6.

Finally, let us compare our results with Pott and Stoica’s results [15]. They study the question how the operator norm of a general vector-valued Calderón–Zygmund operator depends on the UMD constant. The purpose of their paper is to prove that this dependence is linear for a large class of Calderón–Zygmund operators. They prove the following estimate for vector-valued dyadic shifts:

Theorem 1.8 (Self-adjoint vector-valued dyadic shifts depend linearly on the UMD constant [15]). Let \( 1 < p < \infty \). Let \( E \) be a UMD space. Let \( S^{ij} \) be a self-adjoint dyadic shift with parameters \( i \) and \( j \). Then
\[
\|S^{ij} f\|_{L^p(\mathbb{R}; E)} \lesssim (\max\{i, j\} + 1) 2^{\max\{i, j\}/2} \beta_p(E) \|f\|_{L^p(\mathbb{R}; E)}
\]
for all \( f \in L^p(\mathbb{R}; E) \).

By the fact that an estimate for dyadic shifts can be transferred to an estimate for Calderón–Zygmund operators via the dyadic representation theorem (Theorem 1.8), their estimate for dyadic shifts then transfers to the following estimate for vector-valued Calderón–Zygmund operators:

Theorem 1.9 (Calderón–Zygmund operators that have even kernel with sufficiently smoothness, and vanishing paraproduct depend linearly on the UMD constant [15]). Let \( 1 < p < \infty \). Let \( E \) be a UMD space. Let \( T \) be a singular integral operator that satisfies the weak boundedness property and whose kernel satisfies the standard estimates with the Hölder exponent \( \alpha \). Assume that the kernel is even and has smoothness \( \alpha > 1/2 \). Assume that \( T \) satisfies the vanishing paraproduct condition \( T(1) = T^*(1) = 0 \). Then
\[
\|T\|_{L^p(\mathbb{R}; E)} \leq C_{\alpha, d} \beta_p(E),
\]
where \( C_T \) depends only on the constants in the standard estimates and the weak boundedness property.

Now, let us compare our estimate for dyadic shifts (Theorem 1.2) with Pott and Stoica’s estimate (Theorem 1.8). We note that the dependence on the complexity dictates whether the series in the dyadic representation theorem (Theorem 1.5) converges. On the one hand, our estimate depends linearly on the complexity, whereas theirs exponentially, which then translates into the smoothness condition \( \alpha > 1/2 \) in their estimate for Calderón–Zygmund operators (Theorem 1.9). On the other hand, their estimate depends linearly on the UMD constant, whereas ours depends quadratically. We remark that by interpolating between our estimate and
their (by multiplying the inequalities \( \|S\|_{L^p(E)} - L^p(E) \leq k^{1-\theta} \beta_p(E)^{2(1-\theta)} \) and
\( \|S\|_{L^p(E)} - L^p(E) \leq 2^{\theta k/2} \beta_p(E)\theta \), we obtain that
\[ \|S\|_{L^p(E)} - L^p(E) \leq \beta_p(E)^{2-\theta} k^{1-\theta} 2^{\theta k/2}, \]
which then transfers to:

**Corollary 2.10** (Calderón–Zygmund operators that have even kernel and vanishing paraproduct depend subquadratically on the UMD constant). Let \( 1 < p < \infty \). Let \( E \) be a UMD space. Let \( T \) be a singular integral operator that satisfies the weak boundedness property and whose kernel satisfies the standard estimates with the Hölder-exponent \( \alpha \). Assume that the kernel is even. Assume that \( T \) satisfies the paraproduct condition \( T(1) = T^*(1) = 0 \). Then
\[ \|T\|_{L^p(E)} - L^p(E) \leq C_{\alpha, \theta} \frac{1}{(1-\theta)^{2(1-\theta)}} \beta_p(E)^{2(1-\theta)} \]
for \( \alpha \leq 1/2 \),
\[ \|T\|_{L^p(E)} - L^p(E) \leq \beta_p(E) \]
for \( \alpha > 1/2 \),
for every \( \theta \) with \( 0 < \theta < \alpha \). Here \( C_{\alpha, \theta} \) depends only on the constants in the standard estimates and the weak boundedness property.

Lastly, we remark that we prove our estimate for dyadic shifts by using a martingale decoupling equality, whereas Pott and Stoica prove theirs by using the Bellman function method. At the moment, we do not know how to reproduce their result by our method nor our result by their method. A more complete understanding of both methods could yield interesting further results.

## 2. Preliminaries

### 2.1. Sum of stochastically independent conditional expectations.

**Lemma 2.1** (Sum of stochastically independent conditional expectations). Let \( (X_n, F_n, \mu_n) \) be a probability space for each \( n = 1, \ldots, N \). Let \( (X, F, \mu) \) denote the product probability space \( (\Pi_{n=1}^N X_n, \mathcal{F}_{n=1}^N F_n, \mathcal{F}_{n=1}^N \mu_n) \). Let \( 1 \leq p \leq \infty \). Assume that \( f_n \in L^p(X_n, F_n, \mu_n; E) \) and that \( G_n \) is a sub-\( \sigma \)-algebra of \( F_n \) for each \( n = 1, \ldots, N \). Then
\[ \| \sum_{n=1}^N \mathbb{E}[f_n|G_n] \|_{L^p(X, F, \mu; E)} \leq \sum_{n=1}^N \| f_n \|_{L^p(X_n, F_n, \mu_n; E)}. \]

**Proof.** We prove that \( \mathbb{E}[f_n|G_n] = \mathbb{E}[f_n|\chi_{m=1}^N G_n] \), from which the estimate follows by the linearity and the \( L^p \)-contractivity of the conditional expectation operator,
\[ \| \sum_{n=1}^N \mathbb{E}[f_n|G_n] \|_{L^p(X, F, \mu; E)} = \| \mathbb{E}\left[ \sum_{n=1}^N f_n | \chi_{m=1}^N G_n \right] \|_{L^p(X, F, \mu; E)} \leq \sum_{n=1}^N \| f_n \|_{L^p(X, F, \mu; E)}. \]

By Kolmogorov’s definition of the conditional expectation, we have \( \mathbb{E}[f_n|G_n] = \mathbb{E}[f_n|\chi_{m=1}^N G_n] \) if and only if
\[ (2.1) \int_G \mathbb{E}[f_n|G_n] d\mu = \int_G f_n d\mu \]
for all \( G \in \chi_{m=1}^N G_n \). The collection of sets \( G \in \chi_{m=1}^N G_n \) satisfying the condition \([2.1]\) is a \( \lambda \)-system (which means that the collection contains the empty set, is closed under taking complements and is closed under taking countable disjoint unions). The \( \sigma \)-algebra \( \chi_{m=1}^N G_n \) is generated by the collection of sets \( G_1 \times \cdots \times G_N \) with each \( G_n \in G_n \), which is a \( \pi \)-system (which means that the collection is closed under taking finite intersections). Dynkin’s \( \pi - \lambda \) theorem (for a proof, see, for
example, the appendix of Durrett’s textbook [4] states that the \( \lambda \)-system and the \( \sigma \)-algebra both generated by the same \( \pi \)-system coalesce. Hence it suffices to check the condition \[24\] for the sets \( G_1 \times \cdots \times G_N \) with each \( G_n \in \mathcal{G}_n \), which is done by using Fubini’s theorem and Kolmogorov’s definition of the conditional expectation,

\[
\int_{G_1 \times \cdots \times G_N} \mathbb{E}[f_n | G_n] \, d\mu = \int_{G_n} \mathbb{E}[f_n | G_n] \, d\mu_n \prod_{m \neq n} d\mu_m(G_m) = \int_{G_n} f_n \, d\mu_n \prod_{m \neq n} d\mu_m(G_m) = \int_{G_1 \times \cdots \times G_N} f_n \, d\mu.
\]

\[ \square \]

2.2. Properties of \( R \)-bounds. In this section we have collected some properties of \( R \)-bounds. For the proofs and references, see Neerven’s lecture notes [10]. Let \((X, \mathcal{F}, \mu)\) be a \( \sigma \)-finite measure space. Let \( E \) be a Banach space. Assume that \( x \mapsto L(x) \) is an \( \mathcal{L}(E) \)-valued function defined on \( X \) such that the function \( x \mapsto L(x)e \) defined on \( X \) is strongly measurable for each \( e \in E \). We define the operator \( \int_X L(x) \lambda(x) \, d\mu(x) : E \to E \) by

\[
\left( \int_X L(x) \lambda(x) \, d\mu(x) \right)e := \int_X L(x)e \lambda(x) \, d\mu(x) \quad \text{for all } e \in E.
\]

**Proposition 2.2** (Averaging preserves \( R \)-bounds). Let \((X, \mathcal{F}, \mu)\) be a \( \sigma \)-finite measure space. Let \( S \) be an index set. Let \( \{L_s\}_{s \in S} \) be an indexed family of \( \mathcal{L}(E) \)-valued functions defined on \( X \) such that the \( E \)-valued function \( x \mapsto L_s(x)e \) defined on \( X \) is strongly \( \mu \)-measurable for every \( e \in E \) and every \( s \in S \). Let \( \{\lambda_s\}_{s \in S} \) be an indexed family of integrable real-valued functions. Then

\[
\mathcal{R}(\{\int_X L_s(x) \lambda_s(x) \, d\mu(x) : s \in S\}) \leq \sup\{\int_X |\lambda_s(x)| \, d\mu(x) : s \in S\} \cdot \mathcal{R}(\{L_s(x) : s \in S \text{ and } x \in X\}).
\]

**Proposition 2.3** (Triangle inequality for \( R \)-bounds). Let \( S \) and \( T \) be index sets.

Let \( \{L_s\}_{s \in S} \) and \( \{M_t\}_{t \in T} \) be indexed families of operators. Then

\[
\mathcal{R}(\{M_s + L_t : s \in S, t \in T\}) \leq \mathcal{R}(\{M_s : s \in S\}) + \mathcal{R}(\{L_t : t \in T\}).
\]

**Proposition 2.4** (Vector-valued Stein’s inequality). Let \( E \) be a UMD space. Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space. Let \( (\mathcal{G}_n)_{n=1}^\infty \) be a refining sequence of \( \sigma \)-algebras. Then the family

\[
\{\mathbb{E}[\cdot | \mathcal{G}_n] : L^p(\Omega; E) \to L^p(\Omega; E)\}_{n=1}^\infty
\]

is \( R \)-bounded. Moreover,

\[
\mathcal{R}_\mathbb{P}(\{\mathbb{E}[\cdot | \mathcal{G}_n] : L^p(\Omega; E) \to L^p(\Omega; E)\}_{n=1}^\infty) \leq \beta_p(E).
\]

2.3. Pythagoras’ theorem for functions adapted to a sparse collection. Let \( \mu \) be a Borel measure on \( \mathbb{R}^d \). We use the notation \( \langle f \rangle^\mu_Q := \frac{1}{\mu(Q)} \int_Q f \, d\mu \). Let \( S \) be a collection of dyadic cubes. For each \( S \in \mathcal{S} \), let \( \text{ch}_S(S) \) denote the collection of all maximal \( S' \in \mathcal{S} \) such that \( S' \nsubseteq S \) and let \( E_S(S) \) denote the set \( E_S(S) := S \setminus \bigcup_{S' \in \text{ch}_S(S)} S' \). For each \( Q \in \mathcal{D} \), let \( \pi_S(Q) \) denote the minimal dyadic cube \( S \in \mathcal{S} \) such that \( S \supseteq Q \). We say that the collection \( S \) is sparse if \( \mu(E_S(S)) \geq \frac{1}{2} \mu(S) \) for every \( S \in \mathcal{S} \).
Lemma 2.5 (Special case of the dyadic Carleson embedding theorem). Let $E$ be a Banach space. Let $1 < p < \infty$. Assume that $\mathcal{S}$ is a sparse collection. Then

$$(\sum_{S \in \mathcal{S}} (\langle |f|_E \rangle_S^p \mu(S))^1/p) \leq 2p' \|f\|_{L^p(E)}.$$

Proof. For the dyadic Hardy–Littlewood maximal function $M^\mu f = \sup_{Q \subseteq E} \langle f \rangle_Q$, we have $\langle |f|_E \rangle_S \leq \inf_S M^\mu |f|_E$, and, moreover, $\|M|f|_E\|_{L^p(\mu)} \leq p' \|f\|_{L^p(\mu)}$. These facts together with the assumptions yield

$$(\sum_{S \in \mathcal{S}} (\langle |f|_E \rangle_S^p \mu(S))^1/p) \leq 2^{1/p'} (\sum_{S \in \mathcal{S}} \int_{E_\delta(S)} (\inf_S M^\mu |f|_E) \, d\mu)^{1/p}$$

$$\leq 2^{1/p'} \|M^\mu |f|_E\|_{L^p(\mu)} \leq 2^{1/p'} p' \|f\|_{L^p(E)} = 2^{1/p'} p' \|f\|_{L^p(E)}.$$

\[ \square \]

For each $S \in \mathcal{S}$, we define the operator $P_S$ by setting

$$P_S f := \sum_{Q \subseteq E} D_Q f$$

for every locally integrable $f : \mathbb{R}^d \to E$.

Lemma 2.6 (Properties of the operators $P_S$). For each $S \in \mathcal{S}$, the operator $P_S$ has the following properties:

(i) $P_S f = \sum_{S' \in \text{ch}_S(S)} \langle f \rangle_{S'} 1_{S'} + f 1_{E_\delta(S)} - \langle f \rangle S 1_S$.

(ii) $P_S f = f$ if and only if $f$ is supported on $S$, constant on each $S' \in \text{ch}_S(S)$, and satisfies $\int_S f \, d\mu = 0$.

(iii) $P_S^2 = P_S$, $P_S P_T = 0$ whenever $T \in \mathcal{S}$ with $T \neq S$.

(iv) $\int g P_S f \, d\mu = \int P_S g f \, d\mu$ for every $f \in L^p(E)$ and $g \in L^{p'}(E^*)$.

(v) $\|P_S f\|_{L^p(E)} \leq 2 \|1_S f\|_{L^p(E)}$.

Proof. We prove the property (i), from which the other properties follow. On the one hand,

$$f 1_S = \sum_{Q \subseteq E} D_Q f + \langle f \rangle S 1_S,$$

on the other hand,

$$f 1_S = f 1_{E_\delta(S)} + \sum_{S'} f 1_{S'}, \quad f 1_{E_\delta(S)} = \sum_{S' \in \text{ch}_S(S)} (\sum_{Q \subseteq E} D_Q f + \langle f \rangle S' 1_{S'})$$

Thus, by comparing,

$$\sum_{Q \subseteq E} D_Q f - \sum_{S' \in \text{ch}_S(S)} \sum_{Q \subseteq E} D_Q f = f 1_{E_\delta(S)} + \sum_{S' \in \text{ch}_S(S)} (\langle f \rangle S' 1_{S'} - \langle f \rangle S 1_S).$$

Observing that

$$\sum_{Q \subseteq E} - \sum_{S' \in \text{ch}_S(S)} \sum_{Q \subseteq E} = \sum_{Q \pi_S(Q) = S}$$

completes the proof. \[ \square \]

The following variant of Pythagoras’ theorem in the case $E = \mathbb{R}$ was proven by Katz and Pereyra [11, Lemma 7] by using a multilinear estimate. We next give a different proof of the theorem, which extends it to an arbitrary Banach space $E$. 

Lemma 2.7 (Pythagoras’ theorem for sparsely supported, piecewise constant functions). Let $E$ be a Banach space. Let $1 \leq p < \infty$. Let $\mathcal{S}$ be a sparse collection of dyadic cubes. For each $S \in \mathcal{S}$, assume that $f_S$ is a function that is supported on $S$ and constant on each $S' \in \text{ch}_S(S)$. Then
\[
\|\sum_S f_S\|_{L^p(E)} \leq 3p \left( \sum_S \|f_S\|_{L^p(E)}^p \right)^{1/p}.
\]

Moreover, the reverse estimate
\[
\left( \sum_S \|f_S\|_{L^p(E)}^p \right)^{1/p} \leq 6p' \|\sum_S f_S\|_{L^{p'}(E)}
\]
holds if, in addition, one of the following conditions is satisfied:

(i) $\int f_S \, d\mu = 0$, or (ii) $E = \mathbb{R}$ and $f_S \geq 0$,

but may in general fail otherwise.

Proof. First, we prove the direct estimate. By duality, it is equivalent to the estimate
\[
\int \sum_S f_S g \, d\mu \leq 3p \left( \sum_S \|f_S\|_{L^p(E)}^p \right)^{1/p} \|g\|_{L^{p'}(E')}.
\]

Since $f_S$ is supported on $S$, since $f_S$ is constant on $S' \in \text{ch}_S(S)$, and since $S$ is partitioned by $\text{ch}_S(S)$ and $E_S(S)$, we have
\[
\int \sum_S f_S g \, d\mu = \sum_S \int_S f_S g \, d\mu = \sum_{S \in \text{ch}_S(S)} \int_{S'} f_S \, d\mu + \int \sum_S 1_{E_S(S)} f_S g \, d\mu.
\]

We can estimate the second term by Hölder’s inequality and the pairwise disjointness of the sets $E_S(S)$,
\[
\left| \int \sum_S 1_{E_S(S)} f_S g \, d\mu \right| \leq \sum_S 1_{E_S(S)} \|f_S\|_{L^p(E)} \|g\|_{L^{p'}(E')} = \left( \sum_S 1_{E_S(S)} \|f_S\|_{L^p(E)}^p \right)^{1/p} \|g\|_{L^{p'}(E')}.
\]

We can estimate the first term as follows.
\[
\left| \sum_{S \in \text{ch}_S(S)} \langle f_S \rangle_{S'} \int_{S'} g \, d\mu \right|
\leq \sum_S \sum_{S' \in \text{ch}_S(S)} |\langle f_S \rangle_{S'}| \mu(S')^{1/p} \left| \int_{S'} g \, d\mu \right|_{L^{p'}(S')} \mu(S')^{1/p'}
\leq \left( \sum_S \sum_{S' \in \text{ch}_S(S)} \langle f_S \rangle_{S'} \mu(S')^{1/p} \right)^{1/p} \left( \sum_S \sum_{S' \in \text{ch}_S(S)} \left( \frac{|g|_{L^{p'}(S')} \, d\mu}{\mu(S')} \right)^{p'} \mu(S')^{1/p'} \right)^{1/p'}
\leq \left( \sum_S \int |f_S|_{L^p} \, d\mu \right)^{1/p} \left( \sum_S \sum_{S' \in \text{ch}_S(S)} \langle |g|_{L^{p'}(S')} \rangle_{S'} \mu(S')^{1/p'} \right)^{1/p'}.
\]

The proof of the direct estimate is completed by the special case of the dyadic Carleson embedding theorem, Lemma 2.4.

Next, we prove the reverse estimate under the assumption that $\int f_S \, d\mu = 0$. By duality, this estimate is equivalent to the estimate
\[
\sum_S \int f_S g \, d\mu \leq 6p' \sum S \|f_S\|_{L^{p'}(E)} \left( \sum_S \|g_S\|_{L^{p'}(E')} \right)^{1/p'}. 
\]
for arbitrary functions $g_S \in L^{p'}(E')$. By the properties of the operators $P_S$, Lemma 2.6, we have that

$$\int f S g_S d\mu = \int P_S f S g_S d\mu = \int P_S^2 f S g_S d\mu = \int P_S f S \sum_T P_T g_T d\mu = \int f_S \sum_T P_T g_T d\mu.$$  

Note that, although the functions $g_T$ are arbitrary, the functions $P_T g_T$ satisfy the assumptions for the direct estimate; each $P_T g_T$ is supported on $T$, and constant on each $T' \in \text{ch}_S(T)$. Thus, by Hölder’s inequality and the direct estimate,

$$\sum_S \int f_S g_S d\mu = \int \sum_S f_S \sum_T P_T g_T d\mu \leq \|\sum_S f_S\|_{L^p(E)} \|\sum_T P_T g_T\|_{L^{p'}(E')} \leq 3p' \|\sum_S f_S\|_{L^p(E)} \left(\sum_T \|P_T g_T\|_{L^{p'}(E')}\right)^{1/p'} \leq 6p' \|\sum_S f_S\|_{L^p(E)} \left(\sum_T \|g_T\|_{L^{p'}(E')}\right)^{1/p'}.$$  

Next, we prove the reverse estimate under the assumption that $E = \mathbb{R}$ and $f_S \geq 0$. Since $f_S$ is supported on $S$, since $f_S$ is constant on $S' \in \text{ch}_S(S)$, since $S$ is partitioned by $\text{ch}_S(S)$ and $E_S(S)$ and since $\mu(S') \leq 2\mu(E_S(S'))$, we can write

$$\|f_S\|_{L^p(\mathbb{R})}^p = \sum_{S' \in \text{ch}_S(S)} |\langle f_S S'|\mu(S')| + \int 1_{E_S(S)}|f_S|^p d\mu \leq 2 \sum_{S' \in \text{ch}_S(S)} |\langle f_S S'|\mu(E_S(S'))| + \int 1_{E_S(S)}|f_S|^p d\mu = 2 \sum_{S' \in \text{ch}_S(S)} 1_{E_S(S')}|f_S|^p d\mu + \int 1_{E_S(S)}|f_S|^p d\mu.$$  

Summing over $S$ and taking into account that $E_S(S)$ are pairwise disjoint yields

$$\sum_S |f_S|_{L^p(\mathbb{R})}^p \leq 3 \int \left( \sum_{S' \in \text{ch}_S(S)} 1_{E_S(S')}|f_S|\right)^p d\mu.$$  

Using the assumption that $f_S \geq 0$ completes the proof.

Lastly, we note that a simple example shows that the reverse estimate may fail in general. Indeed, let $S := [0,1)$, $S_+ := [0,1/2)$, $f_S := 1_{S_+}$, and $f_S := -1_{S_+}$. Then $\|f_S\|_{L^p(\mathbb{R})}^p + \|f_S\|_{L^p(\mathbb{R})}^p = 2|S| - 2|S_+|$ but $\|f_S\|_{L^p(\mathbb{R})}^p + \|f_S\|_{L^p(\mathbb{R})}^p = 0$. $\square$

3. Decoupling of the sum of martingale differences

Let $(X, \mathcal{F}, \mu)$ be a $\sigma$-finite measure space. Let $(A_n)_{n=-\infty}^\infty$ be a refining sequence of countable partitions of $X$ into measurable sets of finite positive measure. Let $A := \bigcup_{n=\infty}^{\infty} A_n$. For each $K \in A$, let $\text{child}_A(K) := \{K' \in A_{n+1} : K' \subseteq K\}$. For each $K \in A$, let $f_K$ be a function that is supported on $K$ and constant on $K' \in \text{child}_A(A)$ and such that $\int_K f_K d\mu = 0$. Let $(Y_K, \mathcal{G}_K, \nu_K)$ be the probability space such that $Y_K := K$, $\mathcal{G}_K$ is the $\sigma$-algebra generated by $\{K\} \cup \text{child}_A(K)$, and $\nu_K = \mu(K)^{-1} \mu|_K$. Let $(Y, \mathcal{G}, \nu)$ be the product probability space of the spaces $(Y_K, \mathcal{G}_K, \nu_K)_{K \in A}$.

We notice that the sequence $(d_k)_{k=-\infty}^\infty$ with $d_k(x,y) := \sum_{K' \in A_{k+1}} f_K(x)$ is a martingale difference sequence adapted to the filtration $(\mathcal{F}_k)_{k=-\infty}^\infty$ generated by the refining sequence of partitions $(A_k)_{k=-\infty}^\infty$. Conversely, each martingale difference sequence $(d_k)_{k=-\infty}^\infty$ adapted to the filtration $(\mathcal{F}_k)_{k=-\infty}^\infty$ can be written as $d_k := \sum_{K' \in A_{k+1}} f_K$, where for each $K \in A_{k-1}$ the function $f_K$ is defined by $f_K := 1_K d_k = \sum_{K' \in A_k : K' \subseteq K} \langle d_k \rangle_{K'} 1_{K'}$. 


Let $1 < p < \infty$. Let $E$ be a UMD space. Then

\[
\begin{align*}
\frac{1}{\beta_p(E)} (\mathbb{E} \left[ \sum_{K \in \mathcal{A}} \epsilon_K 1_K(x) f_K(y) \right]_{L^p(d\mu(x) \times d\nu(y); E)})^{1/p} \\
\leq \| \sum_{K \in \mathcal{A}} f_K(x) \|_{L^p(d\mu(x); E)} \\
\leq \beta_p(E) (\mathbb{E} \left[ \sum_{K \in \mathcal{A}} \epsilon_K 1_K(x) f_K(y) \right]_{L^p(d\mu(x) \times d\nu(y); E)})^{1/p}.
\end{align*}
\]

Here we give another proof of the equality: Roughly speaking, we construct auxiliary martingale differences $u_K(x, y_K)$ and $v_K(x, y_K)$ such that $f_K(x) = u_K(x, y_K) + v_K(x, y_K)$ and $1_K(x) f_K(y_K) = u_K(x, y_K) - v_K(x, y_K)$, from which the decoupling equality follows by the definition of the UMD property. Let $d_k$ be a martingale difference sequence adapted to the filtration $\mathcal{F}_k$. We write

\[
d_k(x, y) = \sum_{K \in \mathcal{A}_{k-1}} 1_K(x) d_k(x) 1_K(y_K),
\]

and

\[
\tilde{d}_k(x, y) := \sum_{K \in \mathcal{A}_{k-1}} 1_K(x) \tilde{d}_k(y_K) 1_K(y_K).
\]

**Proposition 3.2** (Constructing auxiliary martingale differences). There exists a martingale difference sequence $(u_k)_{k \in \mathbb{Z}}$ on the product measure space $(X \times Y, \mathcal{F} \times \mathcal{G}, \mu \times \nu)$ such that

\[
d_k = u_k + u_{k+1/2}, \quad \text{and} \quad \tilde{d}_k = u_k - u_{k+1/2}.
\]

**Proof.** Let $d_k$ be a martingale difference sequence $d_k$ adapted to the filtration $\mathcal{F}_k$ generated by a refining sequence of partitions $\mathcal{A}_k$. The $\mathcal{F}_k$-measurability of $d_k$ means that $d_k$ equals to a constant $(d_k)_K$ on $K \in \mathcal{A}_k$. Thus, we can write

\[
d_k = \sum_{K \in \mathcal{A}_{k-1}} 1_K d_k = \sum_{K \in \mathcal{A}_{k-1}} 1_K \sum_{K' \subseteq K} 1_{K'} d_k = \sum_{K \in \mathcal{A}_{k-1}} 1_K \sum_{K' \subseteq K} \langle d_k \rangle_{K'}.
\]

The martingale difference property $\mathbb{E}[d_k | \mathcal{F}_{k-1}] = 0$ means that for every $K \in \mathcal{A}_{k-1}$ we have

\[
\int_K d_k d\mu = \sum_{K' \subseteq K} \langle d_k \rangle_{K'} \mu(K') = 0.
\]

First, we consider a fixed $K \in \mathcal{A}_{k-1}$. Let $\nu_K$ be the the measure $\nu_K := \mu(K)^{-1} \mu|_K$ restricted to the sub-$\sigma$-algebra $\mathcal{G}_K$ that is generated by the collection $\{K \cup \{K' \in \mathcal{A}_k : K' \subseteq K\}\}$. Note that the functions

\[
d_K(x, y_K) := 1_K(x) d_k(x) 1_K(y_K) = \sum_{A, B \in \mathcal{A}_k, A \subseteq K} \langle d_k \rangle_A 1_A(x) 1_B(y_K)
\]

and

\[
\tilde{d}_K(x, y_K) := 1_K(x) \tilde{d}_k(y_K) 1_K(y_K) = \sum_{A, B \in \mathcal{A}_k, A \subseteq K} \langle d_k \rangle_B 1_A(x) 1_B(y_K)
\]
are equally distributed in the measure space \((\mathbb{R}^d \times K, \mu \times \nu_K, \mathcal{F} \times \mathcal{G}_K)\), which is to say that the functions take the same values in sets of equal measure. We define the functions \(u_K(x,y_K)\) and \(v_K(x,y_K)\) by the pair of equations
\[
    d_K(x,y_K) = u_K(x,y_K) + v_K(x,y_K), \\
    \tilde{d}_K(x,y_K) = u_K(x,y_K) - v_K(x,y_K).
\]
Therefore, the function \(u_K(x,y_K)\) can be written out as
\[
    u_K(x,y_K) = \frac{1}{2}(d_K(x,y_K) + \tilde{d}_K(x,y_K))
\]
\[
    = \sum_{A,B \in A_k} \left[ \frac{1}{2}((d_k)_A + (d_k)_B)1_A(x)1_B(y_K) \right]
\]
\[
    = \sum_{A \in A_k} \frac{1}{2}((d_k)_A 1_{A \times A}(x,y_K) + \sum_{A,B \in A_k \cap A < B} \frac{1}{2}((d_k)_A - (d_k)_B)1_{A \times B; A < B}(x,y_K),
\]
where in the last step we introduced some order among the finite family
\[
    \{A \in A_k : A \subseteq K\} = \{A_i\}_{i=1}^{l(K)},
\]
and defined \(A < B\) if and only if \(A = A_i, B = A_j,\) and \(i < j\). The function \(v_K(x,y_K)\) can be written out as
\[
    v_K(x,y_K) = \frac{1}{2}(d_K(x,y_K) - \tilde{d}_K(x,y_K))
\]
\[
    = \sum_{A,B \in A_k} \left[ \frac{1}{2}((d_k)_A - (d_k)_B)1_A(x)1_B(y_K) \right]
\]
\[
    = \sum_{A,B \in A_k \cap A < B} \frac{1}{2}((d_k)_A - (d_k)_B)(1_{A \times B}(x,y_K) - 1_{B \times A}(x,y_K)).
\]
Next, we define the product measure space. For each \( k \in \mathbb{Z} \) and \( K \in \mathcal{A}_{k-1} \), let \((Y_K, \mathcal{G}_K, \nu_K)\) be the probability space such that \( Y_K := K, \mathcal{G}_K \) is the \( \sigma \)-algebra generated by \( \{K\} \cup \{K' \in \mathcal{A}_K : K' \subseteq K\} \), and \( \nu_K = \nu(K)^{-1}\mu|_K \). Let \((Y, \mathcal{G}, \nu)\) be the product probability space of the spaces \((Y_K, \mathcal{G}_K, \nu_K)\) for all \( K \in \mathcal{A} \). We recall that the product space \( Y \) is the Cartesian product \( Y = \prod_{K \in \mathcal{A}} Y_K \), the product \( \sigma \)-algebra \( \mathcal{G} \) (in the case of a countable index set) is the collection
\[
\mathcal{G} := \left\{ \prod_{K \in \mathcal{A}} G_K : G_K \in \mathcal{G}_K \right\},
\]
and the product measure \( \nu \) is the unique measure on \((Y, \mathcal{G})\) that satisfies
\[
\nu\left( \prod_{K \in \mathcal{A}} G_K \right) = \prod_{K \in \mathcal{A}} \nu_K(G_K),
\]
where \(*\) in the product indicates that for only finitely many \( G_K \) we have \( G_K \neq Y_K \).

Next, we prove that the sequence \((\ldots,u_k,u_{k+1/2},u_{k+1},u_{k+1+1/2},\ldots)\) defined by
\[
u_k(x,y) := \sum_{K \in \mathcal{A}_{k-1}} u_K(x,y)
\]
and
\[
u_k(x,y) := u_{k+1/2}(x,y) := \sum_{K \in \mathcal{A}_{k-1}} v_K(x,y)
\]
is a martingale difference sequence in the measure space \((X \times Y, \mathcal{F} \times \mathcal{G}, \mu \times \nu)\). Proving this is based on the following observations:

(a) For each \( K \in \mathcal{A}_{k-1} \), the function \( u_K(x,y_K) \) depends on \( x \) and \( y_K \) “in a symmetric way” (see Figure 1):

(b) For each \( K \in \mathcal{A}_{k-1} \), the function \( v_K(x,y_K) \) depends on \( x \) and \( y_K \) “in an anti-symmetric way” (see Figure 1):

(c) The function \( u_K \) averages to zero on the set \( K \times K \) because \( d_k \) itself is a martingale difference; indeed,
\[
\int_K \int_K u_k(x,y_k) \, d\mu(x) \, d\nu_K(y_K) = \frac{1}{\mu(K)} \sum_{A \subseteq A_k} \{d_k\}_A \mu(A) \mu(A) + \frac{1}{2} \left( \frac{1}{\mu(K)} \sum_{A \subseteq A_k} \{d_k\}_A \mu(A) \mu(B) \right)
\]
\[
= \frac{1}{\mu(K)} \sum_{A \subseteq A_k} \{d_k\}_A \mu(A) \sum_{B \subseteq A_k} \mu(B) = \sum_{A \subseteq A_k} \{d_k\}_A \mu(A) = \int_K d_k(x) \, d\mu(x) = 0.
\]

(d) The function \( v_K \) takes equal positive and negative values on two halves of the symmetric sets \( A \times B \cup B \times A \) with \( A < B \), whereas the function \( u_K \) takes equal values on both the halves. Moreover, the function \( v_K \) takes zero value on the symmetric sets \( A \times A \). Thus, for any function \( \phi(u_K) \), we have
\[
\int_{K \times K} v_K \phi(u_K) \, d\mu \, d\nu_K = \sum_{A \subseteq A_k, B \subseteq A_k} \left( \int_{A \times B} v_K \, d\mu \, d\nu_K + \int_{B \times A} v_K \, d\mu \, d\nu_K \right) \phi(u_K) = 0,
\]
where the average \( \langle \phi(u_K) \rangle_{A \times B, B \times A} \) denotes the constant value of \( \phi(u_K) \) on the set \( A \times B \cup B \times A \).
We define the filtration \((\mathcal{U}_k)_{k \in \mathbb{Z}}\) as follows. For each \(k \in \mathbb{Z}\), we define the \(\sigma\)-algebra \(\mathcal{U}_k\) as the \(\sigma\)-algebra generated by the functions \(\{u_t\}_{t \in \mathbb{R}}\) and \(\{1_K\}_{K \in \mathcal{A}_k}\), and, similarly, the \(\sigma\)-algebra \(\mathcal{U}_{k+1/2}\) as the \(\sigma\)-algebra generated by the functions \(\{u_t\}_{t \in \mathbb{R}+1/2}\) and \(\{1_K\}_{K \in \mathcal{A}_{k+1}}\). We note that the functions \(\{1_K\}_{K \in \mathcal{A}_k}\) are included for technical reasons: They ensure that each \(\mathcal{U}_k\), with \(k \in \frac{1}{2}\mathbb{Z}\), is \(\sigma\)-finite so that taking the conditional expectation with respect to it makes sense. Now, by definition, each \(u_k\) is measurable with respect to \(\mathcal{U}_k\), and, furthermore, \((\mathcal{U}_k)_{k \in \frac{1}{2}\mathbb{Z}}\) is a filtration. Next, we check that \(\mathbb{E}[u_k|\mathcal{U}_{k-1/2}] = 0\), which is equivalent to checking that

\[
\int_{X \times Y} u_k \varphi\left(\{u_t\}_{t \in \mathbb{R}+1/2}, \{1_K\}_{K \in \mathcal{A}_{k-1/2}}\right) d\mu d\nu = 0
\]

for all functions \(\varphi(\ldots, u_t, \ldots, u_{k-1/2}, 1_K, \ldots) =: \varphi(\{u_t\}_{t \in \mathbb{R}+1/2}, \{1_K\}_{K \in \mathcal{A}_{k-1/2}})\).

First, we check that \(\mathbb{E}[u_k|\mathcal{U}_{k-1/2}] = 0\) for \(k \in \mathbb{Z}\). We have

\[
\int_{X \times Y} u_k \varphi\left(\{u_t\}_{t \in \mathbb{R}+1/2}, \{1_K\}_{K \in \mathcal{A}_{k-2}}\right) d\mu d\nu
\]

\[
= \sum_{K \in \mathcal{A}_{k-1}} \int_{X \times Y} u_K \varphi\left(\{u_t\}_{t \in \mathbb{R}+2}, \{1_K\}_{K \in \mathcal{A}_{k-2}}\right) d\mu d\nu.
\]

We note that each of the functions \(\{u_t(x, y)\}_{t \in \mathbb{R}+1/2}\) and \(\{1_K(x)\}_{K \in \mathcal{A}_{k-2}}\) is constant with respect to \(x \in K \in \mathcal{A}_{k-1}\); we denote these constant values by their averages. Moreover, \(u_K(x, y)\) depends on \(y\) only via \(y_K\). Therefore, by pulling out the constant, and integrating out the independent variables, we obtain

\[
\int_{X \times Y} u_K \varphi\left(\{u_t\}_{t \in \mathbb{R}+1/2}, \{1_K\}_{K \in \mathcal{A}_{k-2}}\right) d\mu d\nu
\]

\[
= \int_{X \times Y} u_K d\mu d\nu_K
\]

\[
\cdot \prod_{K \in \mathcal{A}_{k-1}} \Phi\left(\{u_t\}_{t \in \mathbb{R}+1/2}, \{1_K\}_{K \in \mathcal{A}_{k-2}}\right) \prod_{K \in \mathcal{A}_{k-1}} d\nu_K.
\]

The observation (c) states that \(\int_{X \times Y} u_K d\mu d\nu_K = 0\).

Finally, we check that \(\mathbb{E}[u_{k+1/2}|\mathcal{U}_k] = 0\) for \(k \in \mathbb{Z}\). Again, we have

\[
\int_{X \times Y} u_{k+1/2} \varphi\left(\{u_t\}_{t \in \mathbb{R}}, \{1_K\}_{K \in \mathcal{A}_{k-1}}\right) d\mu d\nu
\]

\[
= \sum_{K \in \mathcal{A}_{k-1}} \int_{X \times Y} u_K \varphi\left(\{u_t\}_{t \in \mathbb{R}}, \{1_K\}_{K \in \mathcal{A}_{k-1}}\right) d\mu d\nu.
\]

We note that each of the functions \(\{u_t(x, y)\}_{t \in \mathbb{R}}\) and \(\{1_K(x)\}_{K \in \mathcal{A}_{k-1}}\) is constant with respect to \(x \in K \in \mathcal{A}_{k-1}\); we denote these constant values by their averages. Furthermore, \(u_K(x, y) = u_K(x, y) + \sum_{K \in \mathcal{A}_{k-1}} u_K(x, y) = u_K(x, y)\) for \(x \in K \in \mathcal{A}_{k-1}\). Moreover, \(v_K(x, y)\) depends on \(y\) only via \(y_K\). Therefore, again by pulling out the constant, and integrating out the independent variables, we obtain

\[
\int_{X \times Y} u_K \varphi\left(\{u_t\}_{t \in \mathbb{R}}, \{1_K\}_{K \in \mathcal{A}_{k-1}}\right) d\mu d\nu
\]

\[
= \int_{X \times Y} v_K \Phi_K(u_K) d\mu d\nu_K,
\]
where
\[ \Phi_K(u_K) := \prod_{K' \in A^1} Y_{K'} \phi\left(\{(u_i^1)_{K'}\}_{i \leq k-1}, \{(u_j^2)_{K'}\}_{j \leq k-1}\right) \prod_{K' \in A^1} d\nu_{K'}. \]

The observation (d) states that \( \int_{K \times Y_K} v_K \Phi_K(u_K) \, d\mu \, d\nu_K = 0. \)

4. VECTOR-VALUED DYADIC SHIFTS ARE BOUNDED

Let \( L := \max\{i, j\} + 1 \). By picking every \( L \)-th length scale, we decompose the collection \( \mathcal{D} \) of dyadic cubes to subcollections \( \mathcal{D}_{l \mod L} \), with \( l = 0, \ldots, L - 1 \), such that for every \( K \in \mathcal{D}_{l \mod L} \), we have that both \( D_L^j f \) and \( D_L^j g \) are constant on \( K' \in \text{child}_{l \mod L}(K) \) and have zero average on \( K \). More specifically, for each \( l = 1, \ldots, L - 1 \), let \( \mathcal{D}_{l \mod L} = \bigcup_{K \in \mathcal{D}} \{K \in \mathcal{D} : \ell(K) = 2^{-k L^1}\} \). Then
\[
S^j f = \sum_{K \in \mathcal{D}} D_L^j A_D^j D_L^j f = \sum_{l=0}^{L-1} \sum_{K \in \mathcal{D}_{l \mod L}} D_L^j A_D^j D_L^j f.
\]

This decomposition is done in order to decouple by using Theorem 3.1. From now on we consider a fixed \( l \). We write \( e_K := D_L^j A_D^j D_L^j f \). Let \( d\nu_K(x) = |K|^{-1} 1_K(x) \, dx \) be the Lebesgue measure restricted and normalized to the dyadic cube \( K \). Let \( \nu \) denote the product measure \( \times_{K \in A^1} \nu_K \) on the product space \( Y := \prod_{K \in A^1} K \). By Theorem 3.1
\[
\left\| \sum_{K \in \mathcal{D}_{l \mod L}} 1_K(x)e_K(x) \right\|_{L^p(dx; E)}^p \leq \beta_p(\mathcal{E})^p E \left\| \sum_{K \in \mathcal{D}_{l \mod L}} \epsilon_K 1_K(x) e_K(y_K) \right\|_{L^p(dx \times d\nu(y); E)}^p.
\]

We write \( e_K(y_K) = D_L^j A_K D_L^j f(y_K) =: D_L^j g_K(y_K) \). By using Lemma 2.1 together with the fact that \( D_L^j \) is a difference of two conditional expectations, we obtain
\[
\left\| \epsilon_K 1_K(x) \sum_{K \in \mathcal{D}_{l \mod L}} D_L^j g_K(y_K) \right\|_{L^p(dx \times d\nu(y); E)}^p \leq 2^p \left\| \sum_{K \in \mathcal{D}_{l \mod L}} \epsilon_K 1_K(x) g_K(y_K) \right\|_{L^p(dx \times d\nu(y); E)}^p.
\]

We write \( g_K(y_K) = A_K D_L^j f(y_K) =: A_K f_K(y_K) \). By introducing an independent copy \( (Y, \tilde{\nu}) \) of the probability space \( (Y, \nu) \), we write
\[
A_K f_K(y_K) = \frac{1_K(y_K)}{|K|} \int_K a_K(y_K, x') f_K(x') \, dx',
\]
\[
= \int_{Y} 1_K(y_K) a_K(y_K, \tilde{y}_K) f_K(x') \, d\tilde{\nu}(\tilde{y}).
\]

By Jensen’s inequality,
\[
\left| \int_{Y} \sum_{K \in \mathcal{D}_{l \mod L}} \epsilon_K 1_K(x) a_K(y_K, \tilde{y}_K) f_K(\tilde{y}_K) \, d\tilde{\nu}(\tilde{y}) \right|_E^p \leq \int_{Y} \sum_{K \in \mathcal{D}_{l \mod L}} \epsilon_K 1_K(x) a_K(y_K, \tilde{y}_K) f_K(\tilde{y}_K) \, d\tilde{\nu}(\tilde{y})_E^p.
\]

Since the family of operators \( \{a_K(x, x') : K \in \mathcal{D}, x \in K, x' \in K\} \) is \( R \)-bounded, we have
\[
E \left| \sum_{K \in \mathcal{D}_{l \mod L}} \epsilon_K 1_K(x) a_K(y_K, \tilde{y}_K) f_K(\tilde{y}_K) \right|_E^p \leq R_p(\{a\})^p E \left| \sum_{K \in \mathcal{D}_{l \mod L}} \epsilon_K 1_K(x) f_K(\tilde{y}_K) \right|_E^p.
\]
Altogether, we have obtained that
\[ \| K \|_{L^p(\mathbb{D}^1_{\text{mod} L})} \leq (2\beta_p(E)\mathcal{R}_p(\{a\}))^p \int_Y \mathbb{E} \sum_{K \in \mathbb{D}^1_{\text{mod} L}} \epsilon_K 1_K(x) \mathcal{D}_K f(\bar{y}_K) \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p \]

Since \( \mathcal{D}_K^i = 0 \) whenever \( i \neq m \), we can write \( \mathcal{D}_K^i = \mathcal{D}_K^i \sum_{m=0}^{L-1} \mathcal{D}_K^m \). By using Lemma 2.4 together with the fact that \( \mathcal{D}_K^i \) is a difference of two conditional expectations, we obtain

\[ \mathbb{E} \sum_{K \in \mathbb{D}^1_{\text{mod} L}} \epsilon_K 1_K(x) \mathcal{D}_K^i \sum_{m=0}^{L-1} \mathcal{D}_K^m f(\bar{y}_K) \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p \leq 2^p \mathbb{E} \sum_{K \in \mathbb{D}^1_{\text{mod} L}} \epsilon_K 1_K(x) \sum_{m=0}^{L-1} \mathcal{D}_K^m f(\bar{y}_K) \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p \]

We have that \( \sum_{m=0}^{L-1} \mathcal{D}_K^m f \) is constant on \( K' \in \text{child}_{\mathbb{D}^1_{\text{mod} L}}(K) \) and has zero average on \( K \). Therefore, by removing the decoupling using Theorem 3.1, we obtain

\[ \mathbb{E} \sum_{K \in \mathbb{D}^1_{\text{mod} L}} \epsilon_K 1_K(x) \sum_{m=0}^{L-1} \mathcal{D}_K^m f(\bar{y}_K) \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p \leq \| \sum_{K \in \mathbb{D}} \mathcal{D}_K f \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p = \| f \|_{L^p(\mathbb{R}^d;\mathbb{E})}^p. \]

The proof is completed.

5. SUFFICIENT CONDITION FOR THE BOUNDEDNESS OF DYADIC PARAPRODUCTS

From the fact that \( \| f \|_{L^p(\mathbb{R}^d;\mathbb{E})} \rightarrow 0 \) as \( \ell(Q) \rightarrow \infty \), it follows that the functions of the form \( f := \sum_{Q \in \mathcal{Q}} f_Q \equiv \sum_{Q \in \mathcal{Q}} (f - (f)_Q)1_Q \), where \( Q_0 \) are disjoint dyadic cubes, are dense in \( L^p(\mathbb{E}) \). Hence it suffices to prove the estimate

\[ \| \sum_{Q \in \mathcal{D}(Q_0)} \mathcal{D}_Q b(f) \|_{L^p(\mathbb{E})} \leq 6 \cdot 2^p 2^{p'p} \beta_p(E)^2 \beta_p(T) ||b||_{\text{BMO}_{\mathbb{R}^d;\mathbb{E}}} \| f \|_{L^p(\mathbb{R}^d;\mathbb{E})} \]

uniformly for all \( Q_0 \in \mathcal{D} \). Now, we fix a dyadic cube \( Q_0 \). Let \( \mathcal{D}(Q_0) := \{ Q \in \mathcal{D} : Q \in Q_0 \} \). Let \( S := S(Q_0) \subset \mathcal{D}(Q_0) \) be a sparse collection that contains the cube \( Q_0 \). For each \( Q \in \mathcal{D} \), let \( \pi_S(Q) \) denote the minimal dyadic cube \( S \in S \) such that \( S \supseteq Q \). We rearrange the summation as \( \sum_{Q \in \mathcal{D}(Q_0)} = \sum_{S \in S} \sum_{Q \in \mathcal{D}(Q_0)} \). By the variant of Pythagoras’ theorem, Theorem 2.7, we obtain

\[ \| \sum_{Q \in \mathcal{D}(Q_0)} \mathcal{D}_Q b(f) \|_{L^p(\mathbb{E})} = \sum_{S \in S} \sum_{Q \in \mathcal{D}(Q_0) : \pi_S(Q) = S} \mathcal{D}_Q b(f) \|_{L^p(\mathbb{E})} \leq 3p(\| \sum_{S \in S} \sum_{Q \in \mathcal{D}(Q_0) : \pi_S(Q) = S} \mathcal{D}_Q b(f) \|_{L^p(\mathbb{E})} \|_{L^p(\mathbb{E})})^{1/p}. \]

It remains to choose the sparse collection \( S \) so that

\[ \| \sum_{Q \in \mathcal{D}(Q_0) : \pi_S(Q) = S} \mathcal{D}_Q b(f) \|_{L^p(\mathbb{E})} \leq C_{b, E, p, d} \| f \|_{L^p(\mathbb{E})} S^{1/p}, \]
which, by the special case of the dyadic Carleson embedding theorem, Lemma 2.5 completes the proof by the estimate

$$
\left( \sum_{S \in \mathcal{S}} \sum_{Q \in D(Q_0) : \pi(Q) = S} DQ b(f)_Q \| f \|_{L^p(E)}^p \right)^{1/p} \leq C_{b,p,d} \left( \sum_{S \in \mathcal{S}} \| \langle f \rangle_{L^p(S)} \|_{S}^p \right)^{1/p} \leq C_{b,p,d} 2^p \| f \|_{L^p(E)}.
$$

Next, we choose the collection $\mathcal{S}$ so that the estimate (5.1) is satisfied. For each $S \in \mathcal{D}$, let $\text{ch}_S(S)$ be the collection of all the maximal dyadic subcubes $S' \not\subset S$ such that

(5.2) \[ \langle |f|_{L^p(E)} \rangle_{S'} > 2 \langle |f|_{L^p(E)} \rangle_S. \]

By the dyadic nestedness and maximality, the collection $\text{ch}_S(S)$ is pairwise disjoint. We define recursively $S_0 := \{ Q_0 \}$ and $S_{n+1} := \bigcup_{S \in S_n} \text{ch}_S(S)$. Let $\mathcal{S} := \bigcup_{n=0}^{\infty} S_n$. We define the pairwise disjoint sets $E_S(S) := S \setminus \bigcup_{S' \in \text{ch}_S(S)} S'$. By construction,

$$
\sum_{S' \in \text{ch}_S(S)} |S'| \leq \frac{1}{2} |S|,
$$

which is to say that $|E_S(S)| \geq \frac{1}{2} |S|$. Hence the collection $\mathcal{S}$ is sparse.

Next, we check that $\int_Q f \, dx = \int_Q f_S \, dx$ for $f_S := f 1_{E_S(S)} + \sum_{S' \in \text{ch}_S(S)} (f)_{S'} 1_{S'}$ whenever $\pi(Q) = S$. Firstly, the set $Q$ is partitioned by $E_S(S) \cap Q$ and $\{ S' \in \text{ch}_S(S) : S' \cap Q \neq \emptyset \}$. Secondly, by the dyadic nestedness, $S' \cap Q \neq \emptyset$ implies that either $Q \subseteq S'$ or $S' \not\subset Q$. The alternative $Q \not\subseteq S'$ is excluded because $\pi_S(S) = Q$ means that $S$ is the minimal $S'' \in \mathcal{S}$ such that $Q \subseteq S''$. Hence $S' \not\subset Q$ for all $S' \in \text{ch}_S(S)$ with $S' \cap Q \neq \emptyset$. Therefore

$$
\int_Q f \, dx = \int_Q f 1_{E_S(S)} + \sum_{S' \in \text{ch}_S(S) : S' \subseteq Q} \int_{S'} f \, dx.
$$

Next, we check that $|f_S|_E \leq 2 - 2^d (|f|_{L^p(E)})_S$ almost everywhere. First, let $x \in E_S(S)$. Then, by construction, for all $Q \in \mathcal{D}$ such that $Q \ni x$ we have $(|f|_{L^p(E)})_Q \leq 2 (|f|_{L^p(E)})_S$. Therefore, by the Lebesgue differentiation theorem, $|f(x)|_E \leq 2 (|f|_{L^p(E)})_S$ for almost every such $x$. Let $S' \in \text{ch}_S(S)$. By the maximality of $S'$, the dyadic parent of $S'$ satisfies the opposite $(|f|_{L^p(E)})_{S'} \leq 2 (|f|_{L^p(E)})_S$ of the inequality (5.2). By doubling, $(|f|_{L^p(E)})_{S'} \leq 2^d (|f|_{L^p(E)})_S$. Altogether, $(|f|_{L^p(E)})_{S'} \leq 2 \cdot 2^d (|f|_{L^p(E)})_S$.

Altogether, we have that

$$
\| \sum_{Q \in D(Q_0) : \pi(Q) = S} DQ b(f)_Q \|_{L^p(E)} = \| \sum_{Q \in D(Q_0) : \pi(Q) = S} DQ b(f_S)_Q \|_{L^p(E)}
$$

with $\| f_S \|_{L^\infty(E)} \leq 2 \cdot 2^d (|f|_{L^p(E)})_S$. The proof is completed by Lemma 5.1.

**Lemma 5.1.** Let $1 < p < \infty$. Let $E$ be a UMD space. Assume that $T$ is a UMD subspace of $L^p(E)$. Let $S$ be a dyadic cube and let $Q(S)$ be a collection of dyadic subcubes of $S$. Then

$$
\| \sum_{Q \in Q(S)} DQ b(f)_Q \|_{L^p(E)} \leq \beta_p(E)^2 \beta_p(T) \| b \|_{BMO_p(T)} \| f \|_{L^\infty(S;E)} |S|^{1/p}
$$

for any $f \in L^\infty(S;E)$ and $b \in BMO_p(\mathbb{R}^d;T)$. 

Proof without the decoupling equality. By the UMD property and the Kahane contraction principle, we obtain

$$\| \sum_{Q \in \mathcal{Q}(S)} D_Q b(f) \|_{L^p(E)} \leq \beta_p(E)^p \| \sum_{Q \in \mathcal{Q}(S)} \epsilon_Q D_Q b(f) \|_{L^p(E)}^p.$$ 

We expand

$$D_Q b = \sum_{\eta \in \{0,1\}^d \setminus \{0\}} (b, h^\eta_Q) h^\eta_Q,$$

where, for each $Q = I_1 \times \cdots \times I_d$ and $\eta = (\eta_1, \ldots, \eta_d) \in \{0,1\}^d$, we have $h^\eta_Q = h^\eta_Q \cdot h^\eta_Q$ with

$$h^\eta_Q = \frac{1}{\sqrt{|Q|}} \left( 1 - 1_{t_{ext}} \right) \text{ and } h^\eta_Q = \frac{1}{\sqrt{|Q|}} 1_I.$$ 

Therefore

$$\left( \mathbb{E} \left| \sum_{Q \in \mathcal{Q}} \epsilon_Q D_Q b(f) Q \|_{L^p(E)}^p \right\|^p \right)^{1/p} \leq \sum_{\eta \in \{0,1\}^d} \left( \mathbb{E} \left| \sum_{Q \in \mathcal{Q}} \epsilon_Q h^\eta_Q Q \|_{L^p(E)}^p \right\|^p \right)^{1/p}.$$

Next, we consider a fixed $\eta$. We observe that, at each point $x \in \mathbb{R}^d$, we have $h^\eta_Q(x) = \pm h^\eta_Q(x)$ and that $|h^\eta_Q|$ is constant on $Q$. Hence

$$\mathbb{E} \left| \sum_{Q \in \mathcal{Q}} \epsilon_Q h^\eta_Q \langle b, h^\eta_Q f \rangle Q \|_{L^p(E)}^p \right\|^p = \mathbb{E} \left| \sum_{Q \in \mathcal{Q}} \epsilon_Q \langle b, h^\eta_Q |f| \rangle Q \|_{L^p(E)}^p \right\|^p.$$

By the vector-valued Stein inequality, and the observation that, at each point $x \in \mathbb{R}^d$, we have $h^\eta_Q(x) = \pm h^\eta_Q(x)$, we obtain

$$\mathbb{E} \left| \sum_{Q \in \mathcal{Q}} \epsilon_Q \langle b, h^\eta_Q |f| \rangle Q \|_{L^p(E)}^p \right\|^p \leq \beta_p(E)^p \| \sum_{Q \in \mathcal{Q}} \epsilon_Q \langle b, h^\eta_Q |f| \rangle Q \|_{L^p(E)}^p$$

By assumption, we have $b : \mathbb{R}^d \to \mathcal{T}$ with $\mathcal{T} \subseteq L(E)$. By the pointwise norm estimate,

$$\| \sum_{Q \in \mathcal{Q}} \epsilon_Q \langle b, h^\eta_Q |f| \rangle Q \|_{L^p(E)} \leq \| f \|_{L^\infty(E)} \| \sum_{Q \in \mathcal{Q}} \epsilon_Q \langle b, h^\eta_Q |f| \rangle Q \|_{L^p(T)}.$$

We can view $\langle b, h^\eta_Q \rangle h^\eta_Q$ as a subsequence of a martingale difference sequence (thanks to Emil Vuorinen for pointing this out!). We split $Q$ into two subsets $Q_1^+$ and $Q_2$ according to the value of $h^\eta_Q$.

$$Q_1^+ := \bigcup_{Q' \in \text{child}(Q) : (h^\eta_Q|_{Q'})^2 > |Q|^{-1/2}} Q' \quad \text{and} \quad Q_2 := \bigcup_{Q' \in \text{child}(Q) : (h^\eta_Q|_{Q'})^2 < |Q|^{-1/2}} Q'.$$

The corresponding martingale differences are

$$U_Q^b := \langle b \rangle_Q 1_Q + \langle b \rangle_Q 1_{Q_2} + \langle b \rangle Q_1^+ 1_{Q_1^+},$$

and

$$V_Q^b := -\langle (b)Q^2 1_{Q^2} + \langle b \rangle Q_2 1_{Q_2} \rangle + \sum_{Q' \in \text{child}(Q)} \langle f \rangle Q' 1_{Q'}.$$ 

By construction, $D_Q b = U_Q^b + V_Q^b$ and $U_Q^b = \langle b, h^\eta_Q \rangle h^\eta_Q$. Hence, for any signs $\epsilon_Q$, we have

$$\| \sum_{Q \in \mathcal{Q}} \epsilon_Q \langle b, h^\eta_Q |f| \rangle Q \|_{L^p(T)} = \| \sum_{Q \in \mathcal{Q}} (\epsilon_Q U_Q^b + 0 \cdot V_Q^b) \|_{L^p(T)}$$

$$\leq \beta_p(T) \| \sum_{Q \in \mathcal{Q}} (U_Q^b + V_Q^b) \|_{L^p(T)} = \beta_p(T) \| \sum_{Q \in \mathcal{Q}} D_Q b \|_{L^p(T)}.$$
We can expand $\sum_{Q \in S} D_Q b = 1_S(b - (b)_S)$. By the definition of the BMO space,
\[ \|1_S(b - (f)_S)\|_{L^p(T)} \leq \|b\|_{\text{BMO}_p(T)} |S|^{1/p}. \]

**Proof with the decoupling equality.** By the decoupling equality, Theorem 3.1
\[ \| \sum_{Q \in S} D_Q b(f)_Q \|_{L^p(\{dx \, d\mu(y)\})}^p \leq \beta_p(E) \| \sum_{Q \in S} \epsilon_Q D_Q b(y_Q)1_Q(x)(f)_Q \|_{L^p(\{dx \, d\mu(y)\})}^p. \]
Now, at each point $y_Q \in Q$, we have $D_Q b(y_Q)(f)_Q = (D_Q b(y_Q)f)_Q$. By the vector-valued Stein inequality,
\[ \mathbb{E}\| \sum_{Q \in S} \epsilon_Q 1_Q(x)(D_Q b(y_Q)f)_Q \|_{L^p(\{dx, E\})}^p \leq \beta_p(E) \| \sum_{Q \in S} \epsilon_Q 1_Q(x)D_Q b(y_Q)f(x) \|_{L^p(\{dx \, d\mu(y)\}; E)}^p. \]
By the pointwise norm estimate,
\[ \mathbb{E}\| \sum_{Q \in S} \epsilon_Q 1_Q(x)D_Q b(y_Q)f(x) \|_{L^p(\{dx \, d\mu(y)\}; E)}^p \leq \|f\|_{L^p(E)} \| \sum_{Q \in S} \epsilon_Q 1_Q(x)D_Q b(y_Q) \|_{L^p(\{dx \, d\mu(y)\}; T)}^p. \]
By the decoupling equality, Theorem 3.1
\[ \| \sum_{Q \in S} \epsilon_Q 1_Q(x)D_Q b(y_Q) \|_{L^p(\{dx \, d\mu(y)\}; T)}^p \leq \beta_p(T) \| \sum_{Q \in S} D_Q b(x) \|_{L^p(\{dx\}; T)}^p. \]

**Remark.** In the scalar-valued setting, we obtain the following proof of the boundedness of the dyadic paraproduct: Let $S$ be the collection of dyadic cubes that is iteratively chosen by the condition $|\{f\}|_S > 2|\{f\}|_S$. Hence $|\{f\}|_S \leq 2|\{f\}|_S$ whenever $\pi_S(Q) = S$. From the variant of Pythagoras’ theorem (Lemma 2.7), Burkholder’s inequality, and the special case of the dyadic Carleson embedding theorem (Lemma 2.5), it follows that
\[ \| \sum_{Q} (f)_Q D_Q b \|_{L^p(\mathbb{R}^d; \mathbb{R})} \leq 3 \| \sum_{\pi_S(Q) = S} (f)_Q D_Q b \|_{L^p(\mathbb{R}^d; \mathbb{R})}^{1/p} \]
\[ \leq 3 \| \sum_{\pi_S(Q) = S} (f)_Q D_Q b \|_{L^p(\mathbb{R}^d; \mathbb{R})} \]
\[ \leq 6 \| \sum_{\pi_S(Q) = S} (f)_Q D_Q b \|_{L^p(\mathbb{R}^d; \mathbb{R})} \]
\[ \leq 6 \| \sum_{\pi_S(Q) = S} (f)_Q D_Q b \|_{L^p(\mathbb{R}^d; \mathbb{R})} 2^{p'} \| f \|_{L^r(\mathbb{R}^d; \mathbb{R})}. \]
Note that $\beta_p(\mathbb{R}) = \max\{p, p'\} - 1$, which was proven by Burkholder [3].
6. Vector-valued dyadic representation theorem

The proof of the vector-valued dyadic representation theorem follows verbatim
the proof of the scalar-valued one that is given in Hytönen’s lecture notes on the A₂
theorem [8], except for the estimation of matrix elements: In the scalar-valued case,
the absolute value of the matrix elements (which are real numbers) is estimated,
whereas in the vector-valued case, the R-bound of the matrix elements (which are
operators) needs to be estimated. For readability, we have sketched the whole proof
here.

6.1. Expanding the dual pairing by means of dyadic shifts. By expanding
g ∈ Lᵖ(ℝᵈ; E*) as
g = ∑_{J ∈ D} D_J g = ∑_{J ∈ D} ∑_{η=1}^{2^d-1} (g, h_η^J) h_η^J,
where h_η^J with η = 1, . . . , 2ᵈ − 1 and J ∈ D are L²-normalized Haar functions, and
f ∈ Lᵖ(ℝᵈ; E) similarly, the dual pairing is written as
⟨g, Tf⟩ = ∑_{J ∈ D, J' ∈ D} (g, h_J)(h_{J'}, Th_{J})(h_{J'}, f).
The index η will be suppressed from now on. To control the relative arrangement
of I and J and whence the size of matrix elements, the notion of a good dyadic
cube is introduced.

Definition 6.1 (Good dyadic cube). Fix a boundary exponent γ ∈ (0, 1) and an
ancestor threshold r ∈ N. A dyadic cube I ∈ D is good if we have
dist(I, K⁺) > \left( \frac{ℓ(I)}{ℓ(K)} \right)^γ ℓ(K)
for every dyadic ancestor K ∈ D of the dyadic cube I such that ℓ(K) ≥ 2rℓ(I).

To restrict to the good cubes in the dual pairing, the randomized dyadic systems
are introduced. Let D⁰ designate the standard dyadic system. For every parameter
(ω_j)_{j ∈ Z} ∈ \{0, 1\}^Z =: Ω and every I ∈ D⁰, the translated dyadic cube I + ω is defined by
I + ω := I + ∑_{j:2^{-i} < ℓ(I)} 2⁻^j ω_j.
For each ω ∈ Ω, the translated dyadic system Dω is defined by Dω := \{I + ω : I ∈ D⁰\}. The parameter set is equipped with the natural probability measure: Each component ω_j ∈ \{0, 1\}^d has an equal probability 2⁻^d of taking any of the 2ᵈ values
and all components are stochastically independent. By construction, the position
and the goodness of a dyadic cube I + ω are stochastically independent. Also by
construction, the probability P_ω(\{I + ω ∈ Dω is good\}) =: πgood does not depend on
I ∈ D⁰, and, as calculated in [8] Lemma 2.3,
πgood ≥ 1 − \frac{8d}{γ} 2⁻^rγ.
In particular, for any boundary exponent γ ∈ (0, 1) we can make the probability
πgood strictly positive by choosing the ancestor threshold r ∈ N sufficiently large.

The following proposition was proven by Hytönen [8] Proposition 3.5. (For an
earlier version of the proposition, see [9] Theorem 3.1.)
Proposition 6.2 (Discarding the bad cubes). Assume that $T : L^p(\mathbb{R}^d; \mathbb{R}) \to L^p(\mathbb{R}^d; \mathbb{R})$ is bounded. Then

$$
\langle \beta, T\alpha \rangle = \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{I \in \mathcal{D}^\omega, J \in \mathcal{D}^\omega: \text{smaller}(I,J) \text{ is good}} \langle \beta, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, \alpha \rangle
$$

for all $\beta \in C^1_0(\mathbb{R}^d; \mathbb{R})$ and $\alpha \in C^1_0(\mathbb{R}^d; \mathbb{R})$.

Let $C^1_0(\mathbb{R}^d; \mathbb{R}) \otimes E$ denote the set of all finite linear combinations of the form

$$
f = \sum_{i=1}^{I} \alpha_i e_i \text{ with } \alpha_i \in C^1_0(\mathbb{R}^d; \mathbb{R}) \text{ and } e_i \in E,
$$

set which is dense in $L^p(\mathbb{R}^d; E)$. By linearity, Theorem 6.2 extends to vector-valued functions.

Corollary 6.3. Let $E$ be a Banach space. Assume that $T : L^p(\mathbb{R}^d; E) \to L^p(\mathbb{R}^d; E)$ is bounded. Then

$$
\langle g, Tf \rangle = \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{I \in \mathcal{D}^\omega, J \in \mathcal{D}^\omega: \text{smaller}(I,J) \text{ is good}} \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle
$$

for all $g \in C^1_0(\mathbb{R}^d; \mathbb{R}) \otimes E^*$ and $f \in C^1_0(\mathbb{R}^d; \mathbb{R}) \otimes E$.

Next, the paraproducts are extracted. The dyadic system $\mathcal{D}^\omega$ is suppressed in the notation from now on. Consider the summation

$$
\sum_{I \cap J \neq \varnothing \text{ is good}} \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle.
$$

In the case $'I \notin J'$, the paraproduct $\Pi_{T,1}$ is extracted as follows: Let $J_I$ denote the dyadic child of $J$ that contains $I$. Then

$$
\langle h_J, Th_I \rangle = \langle 1_{J_I'} h_J, Th_I \rangle + \langle h_J \rangle \langle 1_{J_I}, Th_I \rangle
$$

$$
= \langle 1_{J_I'} (h_J - \langle h_I \rangle), Th_I \rangle + \langle h_J \rangle \langle 1_{J_I} + 1_{J_I'}, Th_I \rangle.
$$

Summing the last term yields

$$
\sum_{I, J \in \mathcal{D}^\omega: I \cap J \neq \varnothing} \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle = \sum_I \sum_{J, J \neq J} \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle 1_{J_I'} (h_J - \langle h_I \rangle), Th_I \rangle
$$

$$
= \sum_I \langle g, Th_I \rangle \langle h_I, f \rangle := \langle \Pi_{T,1} g, f \rangle.
$$

Similarly, in the case $'J \notin I'$ the paraproduct $\Pi_{T,1}$ is extracted. For the remaining, it is supposed that the paraproducts are extracted, and hence the convention

$$
\langle h_J, Th_I \rangle := \langle 1_{J_I'} (h_J - \langle h_I \rangle), Th_I \rangle \text{ whenever } I \notin J,
$$

is used together with the similar convention whenever $J \notin I$.

Next, the summation is rearranged according to the minimal common dyadic ancestor of $I$ and $J$, which is denoted by $I \vee J$. (If $I \subseteq J$, then $I \vee J = J$. If $I \cap J = \varnothing$, then a common dyadic ancestor exists because one of the cubes is good.)

By splitting the summation according to which one of the cubes $I$ and $J$ has smaller side length (and hence is good), and by rearranging the summation according to which cube $K$ is the minimal common dyadic ancestor $I \vee J$ and what is the
size of $I$ and $J$ relative to $I \cup J$, one obtains

$$
\sum_{I,J} \sum_{\text{smaller } I,J \text{ is good}} = \sum_{I,J \in J} \sum_{K \text{ good, } \ell(I) = 2^{-ix} \ell(K), \ell(J) = 2^{-ix} \ell(K)} + \sum_{I,J \not\in J} \sum_{K \text{ good, } \ell(I) = 2^{-ix} \ell(K), \ell(J) = 2^{-ix} \ell(K)}.
$$

Note that, for $K = I \cup J$, one can write

$$
\sum_{I,J \notin J} \langle g, h_J \rangle \langle h_J, Th_1 \rangle \langle h_I, f \rangle = \langle g, D_K^I A_J^K D_K^I f \rangle
$$

by defining

$$
A_{K}^{i,j} f(x') = \frac{1_K(x')}{|K|} \int_K a_{K}^{i,j}(x', x) f(x) \, dx
$$

with

$$
a_{K}^{i,j}(x', x) := |K| \sum_{I,J \subset J = K} \sum_{\text{smaller } I,J \text{ is good, } \ell(I) = 2^{-ix} \ell(K), \ell(J) = 2^{-ix} \ell(K)} h_J(x') h_I(x) \langle h_J, Th_1 \rangle.
$$

Altogether, it is obtained that

$$
\langle g, T f \rangle = \frac{1}{\pi_{\text{good}}} \sum_{i,j} \langle g, \sum_{K \in \mathcal{D}} D_K^i A_K^{i,j} D_K^j f \rangle + \frac{1}{\pi_{\text{good}}} \sum_{i} \langle g, (\Pi_{T,1}^{2} + (\Pi_{T,1}^{2})^*) f \rangle.
$$

6.2. Estimating the $R$-bounds of the matrix elements. We may consider the case $i \geq j$ (which means $\ell(I) \leq \ell(J)$), since, by duality, the case $i < j$ can be treated similarly. It remains to estimate the $R$-bound of the family $\{a_{K}^{i,j}(x, x') : K \in \mathcal{D}, x \in K, x' \in K\}$ of the operator-valued kernels defined by

$$
a_{K}^{i,j}(x', x) := |K| \sum_{I,J \subset J = K} \sum_{\text{smaller } I,J \text{ is good, } \ell(I) = 2^{-ix} \ell(K), \ell(J) = 2^{-ix} \ell(K)} h_J(x') h_I(x) \langle h_J, Th_1 \rangle
$$

with $i \geq j$ (and hence $\ell(I) \leq \ell(J)$). We divide this into cases according to two criteria. The first criterion is whether $K$ is much bigger than $I$. The second criterion is how the cubes $I$ and $J$ intersect: Whether $I \notin J$ (in which case $K = J$), $I = J$ (in which case $K = J$), or $I \cap J = \emptyset$. In total, we have five cases:

- $\ell(K) > 2^\varepsilon \ell(I)$ and $I \cap J = \emptyset$,
- $\ell(J) > 2^\varepsilon \ell(I)$ and $I \notin J$ (in this case $K = J$),
- $\ell(K) \leq 2^\varepsilon \ell(I)$ and $I \cap J = \emptyset$,
- $\ell(J) \leq 2^\varepsilon \ell(I)$ and $I \notin J$ (in this case $K = J$), and
- $I = J$ (in this case $K = I = J$).

These cases are tackled in Lemmas 6.4 through 6.8 which complete the proof of the representation theorem by assuring that

$$
\mathcal{R}(\{a_{K}^{i,j}(x', x) : K \in \mathcal{D}, x \in K, x' \in K\}) \lesssim_{\gamma,d} (\mathcal{R}_{\text{CZ}_d} + \mathcal{R}_{\text{CZ}_d} + \mathcal{R}_{\text{WBP}}) 2^{-(1+\varepsilon)\max\{i,j\}},
$$

under the choice $\gamma := \frac{\alpha}{\alpha + d}$ of the boundary exponent $\gamma \in (0, 1)$.
Lemma 6.4 (Case \( \ell'(I) \leq \ell(J), \ell(K) \geq 2^\epsilon \ell(I), \text{ and } I \cap J = \emptyset \)). Suppose that \( i \) and \( j \) are nonnegative integers such that \( i > r \) and \( i \geq j \). Let

\[
a_{K}^{ij}(x', x) := |K| \sum_{I} |h_{I}(x')h_{J}(x)|,
\]

where the summation is over all the dyadic cubes \( I \) such that \( I \cap J = \emptyset \) and \( I \vee J = K \), \( \ell(I) = 2^{-i} \ell(K) \), \( \ell(J) = 2^{-j} \ell(K) \), and \( I \) is good with threshold \( r \) and exponent \( \gamma \). Then

\[
\mathcal{R}(|a_{K}^{ij}(x', y)| : K \in \mathcal{D}, x \in K \text{ and } x' \in K) \lesssim \mathcal{R}_{CZ} 2^{-i(1-\gamma) - \gamma d}.
\]

Proof. We observe that for each triplet \((K, x, x')\) either the sum is empty or there is a unique \( I_{K,x} \) and a unique \( J_{K,x'} \) satisfying the summation condition. Let \( y_{I_{K,x}} \) denote the center of the dyadic interval \( I_{K,x} \). By using the integral representation of the Calderón–Zygmund operator \( T \), and by using the cancellation of the Haar functions, we write

\[
a_{K}^{ij}(x', x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} (k(y', y) - k(y', y_{I_{K,x}})) \left( \frac{|y - y'|}{|y - y_{I_{K,x}}|} \right)^\alpha \left| y - y' \right|^d 1_{I_{K,x}}(y) 1_{J_{K,x'}}(y') \times |K| \left( \frac{|y - y_{I_{K,x}}|}{|y - y'|} \right)^\alpha \frac{1}{|y - y'|^d} h_{I_{K,x}}(y)h_{J_{K,x'}}(y')h_{I_{K,x}}(x)h_{J_{K,x'}}(x') \, dy \, dy'.
\]

Under the assumptions, we have \( |y - y_{I_{K,x}}| < \frac{1}{2} |y - y'| \), which is checked in the following paragraph. Hence, by the Rademacher standard estimates,

\[
\mathcal{R}(|L_{K,x,x'}(y, y') : x \in K, x' \in K \text{ and } y \in \mathbb{R}^d, y' \in \mathbb{R}) \lesssim \mathcal{R}_{CZ}.
\]

Next, we show that

\[
\sup_{x \in K, x' \in K} \int_{\mathbb{R}^d} |\lambda_{K,x,x'}(y, y')| \, dy \, dy' \lesssim_{\gamma} 2^{-i(1-\gamma) - \gamma d},
\]

which, by Theorem 2.2, completes the proof. For the remaining, we suppress the dependence on the triplet \((K, x, x')\) in the notation. Since \( y \in I \) and \( y_{I} \in I \), we have \( |y - y_{I}| \leq \frac{1}{2} \ell(I) \), and since \( y \in I \) and \( y' \in J \), we have \( |y - y'| \geq \text{dist}(I,J) \); hence

\[
\left( \frac{|y - y_{I}|}{|y - y'|} \right)^\alpha \frac{1}{|y - y'|^d} \leq \left( \frac{\ell(I)}{\text{dist}(I,J)} \right)^\alpha \frac{1}{\text{dist}(I,J)^d}.
\]

Therefore

\[
\int_{\mathbb{R}^d} |\lambda_{K}(y, y')| \, dy \, dy' \leq \|h_{I}\|_{\infty} \|h_{J}\|_{\infty} \|h_{I}\|_{1} \|h_{J}\|_{1} |K| \left( \frac{\ell(I)}{\text{dist}(I,J)} \right)^\alpha \frac{1}{\text{dist}(I,J)^d}.
\]

It remains to check that

\[
\text{dist}(I,J) \geq 2^{\gamma} \left( \frac{\ell(I)}{\ell(K)} \right)^{\gamma} \ell(K).
\]

In particular, this implies that \( |y - y_{I}| \leq \frac{1}{2} |y - y'| \). Let \( K_{I} \) denote the dyadic child of \( K \) that contains \( I \). Since \( \ell(K) > 2^{\epsilon} \ell(I) \), we have \( \ell(K_{I}) \geq 2^{\epsilon} \ell(I) \). Therefore, since \( I \) is good, we have that

\[
\text{dist}(I, K_{I}) > \left( \frac{\ell(I)}{\ell(K_{I})} \right)^{\gamma} \ell(K_{I}) = 2^{\gamma} \left( \frac{\ell(I)}{\ell(K)} \right)^{\gamma} \ell(K).
\]
If $K_I$ intersected $J$, then either $K_I \not\subset J$ (which is not true because we assume that $I$ and $J$ are disjoint) or $K_I \supset J$ (which is not true because we assume that $K$ is the minimal dyadic ancestor of $I$ that contains $J$). Therefore $K_I$ does not intersect $J$, and hence
\[ \text{dist}(I, J) > \text{dist}(I, K_I^c). \]
The proof is completed. \hfill $\square$

**Lemma 6.5** (Case $\ell(I) \leq \ell(J)$, $\ell(K) \leq 2^{r+1} \ell(I)$, and $I \cap J = \emptyset$). Suppose that $i$ and $j$ are nonnegative integers such that $i \leq r$ and $i \geq j$. Let
\[ a_{ij}^{ij}(x', x) := |K| \sum_{I, J} |h_{IJ}(x)h_{J,x}(x')|, \]
where the summation is over all the dyadic cubes $I$ and $J$ such that $I \cap J = \emptyset$, $I \cap J = K$, $\ell(I) = 2^{-i} \ell(K)$, $\ell(J) = 2^{-j} \ell(K)$, and $I$ is good with threshold $r$ and exponent $\gamma$. Then
\[ \mathcal{R}(\{a_{ij}^{ij}(x', x) : K \in D, x \in K \text{ and } x' \in K\}) \leq r_{c_0} \mathcal{R}_{CZ_0}. \]

**Proof.** We note that for each triplet $(K, x, x')$ either the sum is empty or there is a unique $I_{K,x}$ and a unique $J_{K,x}$ satisfying the summation condition. By using the integral representation of the Calderón-Zygmund operator $T$, we write
\[ a_{ij}^{ij}(x', x) = \int_{\mathbb{R}^d} 1_{I_{K,x}}(y) 1_{J_{K,x'}}(y') k(y', y) |y - y'|d^d \times [K] \frac{1}{|y - y'|^d} h_{K,x}(y) h_{J_{K,x'}}(y') h_{J_{K,x}}(x) h_{J,K,x}(x') dy dy'. \]
By the Rademacher standard estimates,
\[ \mathcal{R}(\{L_{K,x,x'}(y, y') : x \in K, x' \in K \text{ and } y, y' \in \mathbb{R}\}) \leq \mathcal{R}_{CZ_0}. \]
We next check that
\[ \sup \{ \int_{\mathbb{R}^d} |\lambda_{K,x,x'}(y, y')| dydy : x \in K, x' \in K \} \leq r_{c_0} 1, \]
which, by Theorem 25, completes the proof.

For the remaining, we suppress the dependence on the triplet $(K, x, x')$ in the notation. Since $\ell(K) \leq 2^{r+1} \ell(I)$ and $K \supset I$, we have $2^{r+1} I \supset K$, and since $K \supset I$ and $I \cap J = \emptyset$, we have $(K \setminus J) \supset J$; hence $(2^{r+1} I \setminus I) \supset J$. Since $\ell(I) \leq \ell(J)$, $J \supset K$ (and hence $\ell(J) \leq \ell(K)$), and $\ell(K) \leq 2^{r+1} \ell(I)$, we have $|I| \approx_r |J| \approx_r |K|$. Therefore
\[ \int_{\mathbb{R}^d} |\lambda(y, y')| dy dy \leq |K||h_I|^{2}_{\infty} |h_J|^{2}_{\infty} \int_{I} \frac{1}{|y - y'|^d} dy dy' \leq |K||h_I|^{2}_{\infty} |h_J|^{2}_{\infty} \int_{I} \frac{1}{|y - y'|^d} dy dy' \approx r_{c_0} |K||h_I|^{2}_{\infty} |h_J|^{2}_{\infty} |I| \approx_r 1. \]

**Lemma 6.6** (Case $I = J = K$). Let
\[ a_{ij}^{00}(x', x) := |I| (h_I, T h_I) h_{I,x}(x) h_{I,x}(x'). \]
By the Rademacher weak boundedness property, we have

\[
\mathcal{R}(\{a_I(x', x) : I \in D, x \in I \text{ and } x' \in I\}) \lesssim_{d} \mathcal{R}_{\text{CZ}} + \mathcal{R}_{\text{WBP}}.
\]

**Proof.** Let \( I_i \) (where \( i = 1, \ldots, 2^d \)) denote the dyadic children of \( I \). By decomposing \( 1_I = \sum I_i \), and using the integral representation of the Calderón–Zygmund operator \( T \), we write

\[
a_{I}^{00}(x', x) = \sum_{I_i} |I|h_I(x)h_I(x')(h_I)_I, h_I, (1_{I_i}, T1_{I_i})
\]

\[
= \sum_{I_i} |I||I_i|h_I(x)h_I(x')(h_I)_I, h_I, \left(\frac{1_{I_1}T1_{I_i}}{|I_i|}\right)
\]

\[
+ \sum_{I_i \neq I_j} |I||I_i|h_I(x)h_I(x')(h_I)_I, h_I, \int_{\mathbb{R}^d \times \mathbb{R}^d} 1_{I_i}(y)1_{I_j}(y')k(y, y')|y - y'|^d
\]

\[
\times \frac{1}{|I_i||y - y'|^d} 1_{I_i}(y)1_{I_j}(y')dydy'
\]

\[
= \sum_{I_i} \left(\frac{1_{I_1}T1_{I_i}}{|I_i|}\right) + \sum_{I_i \neq I_j} \int_{\mathbb{R}^d \times \mathbb{R}^d} L_{I_i, I_j}(y, y') \times \lambda_{I_i, I_j}(y, y')dydy'.
\]

By the Rademacher standard estimates, we have

\[
\mathcal{R}(\{L_{I_i, I_j}(y, y') : I \in D, I_i \neq I_j, y \in I_i, y' \in I_j, \}) \lesssim_{d} \mathcal{R}_{\text{CZ}}.
\]

Moreover, we have

\[
\sup(I_i \in D, I_i \neq I_j, y \in I_i, y' \in I_j, \) \lesssim_{d} \mathcal{R}_{\text{WBP}}.
\]

The proof is completed by using Theorem [2.2] and Proposition [2.3].

**Lemma 6.7** (Case \( \ell(I) < 2^{-r}\ell(J), I \not\subset J \)). Suppose that \( i \) is a nonnegative integer such that \( i > r \). Let

\[
a_{I_j}^0(x', x) := |I| \sum_J (1_{J_j}h_J - (h_J)_{J_j}T h_J) h_I(x)h_J(x'),
\]

where \( J_j \) is the dyadic child of \( J \) that contains \( I \) and the summation is over all the dyadic cubes \( I \) such that \( I \not\subset J \), \( \ell(I) = 2^{-i}\ell(J) \) and \( I \) is good with threshold \( r \) and exponent \( \gamma \). Then

\[
\mathcal{R}(\{a_{I_j}^0(x', x) : J \in D, x \in J \text{ and } x' \in J\}) \lesssim_{d} \mathcal{R}_{\text{CZ}}2^{-i(1-\gamma)}.
\]

**Proof.** We observe that for each triplet \((J, x, x')\) either the sum is empty or there is a unique \( I_{J, x} \) satisfying the summation condition. Let \( y_{I_{J, x}} \) denote the center of the dyadic interval \( I_{J, x} \). By using the integral representation of the Calderón–Zygmund
operator $T$ and by using the cancellation of the Haar functions, we have

$$a_{ij}^y(x', x) = \int_{\mathbb{R}^d} (k(y', y) - k(y', y_{J_i}j_j)) \left( \frac{|y-y'|}{|y-y_{J_i}j_j|} \right)^\alpha |y-y'|^d$$

$$\times |J_i\left( \frac{|y-y_{J_i}j_j|}{|y-y'|} \right)^\alpha \frac{1}{|y-y'|^d} h_{J_i,x_i}(y) 1_{J_i}(y')(h_J(y') - \langle h_J \rangle_{J_i})$$

$$\cdot h_{J_{K, x_i}}(x) h_J(x') \, dy'$$

$$=: \int_{\mathbb{R}^d} L_{J_i,x_i} J(x', y') \, dy'.$$

Under the assumptions, we have $|y_{J_i}j_j - y| \leq \frac{1}{2}|y-y'|$, which is checked in next paragraph. Hence, by the Rademacher standard estimates, we have

$$\mathcal{R}(\{ (J_i, x_i')(y, y') : x \in J, x' \in J \text{ and } y \in \mathbb{R}^d, y' \in \mathbb{R} \}) \leq \mathcal{R}_{CZ_a}.$$

Next, we show that

$$\sup\{ \int_{\mathbb{R}^d} |\lambda_{J_i,x_i}(y, y')| \, dy' \, dy : x \in J, x' \in J \} \leq \gamma 2^{-i(1-\gamma)},$$

which, by Theorem 2.2, completes the proof.

For the remaining, we suppress the dependence on the triplet $(J, x, x')$ in the notation. Since $\text{dist}(I, J_i) > \ell(I)$ (which follows from the facts that $I$ is good and $\ell(J_i) \geq 2^r \ell(I)$), since $y \in I$, and since $y' \in J_i$, we have that $|y_i - y| \leq \frac{1}{2}|y-y'|$. Hence, by the triangle inequality, $|y - y'| \geq \frac{3}{2}|y' - y|$. Therefore

$$\left( \frac{|y-y'|}{|y-y'|} \right)^\alpha \frac{1}{|y-y'|^d} \leq |y-y'|^\alpha \frac{1}{|y-y'|^{\alpha+d}}.$$

Therefore

$$\int_{\mathbb{R}^d} |\lambda(y, y')| \, dy' \, dy$$

$$\leq \frac{\alpha}{\ell(I)} \int_{J_i} \frac{1}{|y-y'|^{\alpha+d}} dy'$$

$$\leq \left( \frac{\ell(I)}{\text{dist}(I, J_i)} \right)^\alpha \left( \frac{\ell(J_i)}{\ell(I)} \right)^\alpha.$$

Since $I$ is good and $\ell(J_i) \geq 2^r \ell(I)$, we have that

$$\text{dist}(I, J_i) > \ell(J_i) \left( \frac{\ell(I)}{\ell(J_i)} \right)^\gamma,$$

which concludes the proof.

\begin{lemma}[Case $\ell(I) \geq 2^{-r} \ell(J), I \not\subset J'$] Suppose that $i$ is a nonnegative integer such that $1 \leq i \leq r$. Let

$$a_{ij}^y(x', x) := |J| \sum_{J_i} 1_{J_i}(h_J - \langle h_J \rangle_{J_i}) T h_{J_i} h(x) h_J(x'),$$

where $J_i$ is the dyadic child of $J$ that contains $I$ and the summation is over all the dyadic cubes $I$ such that $I \subset J$, $\ell(I) = 2^{-i} \ell(J)$ and $I$ is good with threshold $r$ and exponent $\gamma$. Then

$$\mathcal{R}(\{ a_J(x', x) : J \in \mathcal{D}, x \in J \text{ and } x' \in J \}) \leq \mathcal{R}_{CZ_a} + \mathcal{R}_{CZ_0}.$$
We observe that for each triplet \((J, x, x')\) either the sum is empty or there is a unique \(I_{J,x}\) satisfying the summation condition. Let \(y_{I_{J,x}}\) denote the center of the dyadic interval \(I_{J,x}\). We split \(1_{J} = 1_{J_{J}^c(3I)} + 1_{J_{J}^c(3I)^c}\). By using the integral representation of the Calderón–Zygmund operator \(T\), and by using the cancellation of the Haar functions, we have
\[
a_{J}^{0}(x', x) = \int_{\mathbb{R}^{d} \times \mathbb{R}^{d}} \frac{1}{|y - y'|^d} h_{J_{J}(y)}(y') \left( h_{J}(y') - (h_{J})_{J_{J}} h_{I_{K,x}}(x) h_{J}(x') \right) dy dy' + \int_{\mathbb{R}^{d} \times \mathbb{R}^{d}} k(y', y) \left| y - y' \right|^d dy dy',
\]
and
\[
\mathcal{R}(\{L_{J,x,x'}(y, y') : x \in J, x' \in J \text{ and } y \in \mathbb{R}^{d}, y' \in \mathbb{R}^{d}\}) \leq \mathcal{R}_{CZ}.\]

The same calculation as in \([5, 2]\) yields
\[
\sup \left\{ \int_{\mathbb{R}^{d} \times \mathbb{R}^{d}} |\lambda_{J,x,x'(y, y')}| dy' dy : x \in J, x' \in J \right\} \leq_{d, \alpha} 1.
\]

Moreover, we have
\[
\sup \left\{ \int_{\mathbb{R}^{d} \times \mathbb{R}^{d}} |\mu_{J,x,x'}(y, y')| dy' dy : x \in J, x' \in J \right\} \leq |J| h_{\infty} \|h_{J}\|_{\infty} \|h_{J} - (h_{J})_{J_{J}}\|_{\infty} \|h_{J}\|_{\infty} \int_{I} \int_{(3I)^{c}} \frac{1}{|y - y'|^d} dy dy' \leq_{d} 1.
\]

By Theorem \([2, 2]\) and Proposition \([2, 3]\) the proof is completed.
[4] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[5] T. Figiel and P. Wojtaszczyk. Special bases in function spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 561–597. North-Holland, Amsterdam, 2001.

[6] Tadeusz Figiel. Singular integral operators: a martingale approach. In *Geometry of Banach spaces (Strobl, 1989)*, volume 158 of *London Math. Soc. Lecture Note Ser.*, pages 95–110. Cambridge Univ. Press, Cambridge, 1990.

[7] Tuomas Hytönen and Lutz Weis. A $T_1$ theorem for integral transformations with operator-valued kernel. *J. Reine Angew. Math.*, 599:155–200, 2006.

[8] Tuomas P. Hytönen. Representation of singular integrals by dyadic operators, and the $A_2$ theorem. Lecture notes. 2011. arXiv:1108.5119 [math.CA].

[9] Tuomas P. Hytönen. The sharp weighted bound for general Calderón-Zygmund operators. *Ann. of Math. (2)*, 175(3):1473–1506, 2012.

[10] Tuomas P. Hytönen. The vector-valued nonhomogeneous Tb theorem. *Int. Math. Res. Not. IMRN*, (2):451–511, 2014.

[11] Nets Hawk Katz and María Cristina Pereyra. Haar multipliers, paraproducts, and weighted inequalities. In *Analysis of divergence (Orono, ME, 1997)*, Appl. Numer. Harmon. Anal., pages 145–170. Birkhäuser Boston, Boston, MA, 1999.

[12] Michael T. Lacey, Stefanie Petermichl, and Maria Carmen Reguera. Sharp $A_2$ inequality for Haar shift operators. *Math. Ann.*, 348(1):127–141, 2010.

[13] Terry R. McConnell. Decoupling and stochastic integration in UMD Banach spaces. *Probab. Math. Statist.*, 10(2):283–295, 1989.

[14] Stefanie Petermichl. Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(6):455–460, 2000.

[15] Sandra Pott and Andrei Stoica. Linear bounds for Calderón-Zygmund operators with even kernel on UMD spaces. *J. Funct. Anal.*, 266(5):3303–3319, 2014.

[16] J.M.A.M. van Neerven. Stochastic Evolution Equations. Lecture notes of the Internet Seminar 2007-2008. http://fa.its.tudelft.nl/~neerven/publications/papers/ISEM.pdf.

[17] Lutz Weis. Operator-valued Fourier multiplier theorems and maximal $L_p$-regularity. *Math. Ann.*, 319(4):735–758, 2001.

Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, FI-00014 HELSINKI, FINLAND

E-mail address: timo.s.hanninen@helsinki.fi

Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, FI-00014 HELSINKI, FINLAND

E-mail address: tuomas.hytonen@helsinki.fi