Rev. Mat. Iberoam. **28** (2012), no. 4, 961–998
DOI 10.4171/RMI/699

Vector-valued non-homogeneous *Tb* theorem on metric measure spaces

Henri Martikainen

Abstract. We prove a vector-valued non-homogeneous Tb theorem on certain quasimetric spaces equipped with what we call an upper doubling measure. Essentially, we merge recent techniques from the domain and range side of things, achieving a Tb theorem which is quite general with respect to both of them.

1. Introduction

In the seminal paper [13] by Nazarov, Treil and Volberg, it was already indicated that it should be possible to prove some version of their (Euclidean) non-homogeneous Tb theorem also in a more abstract metric space setting, just like the well-established homogeneous theory in this generality [3], [2]. A recent paper [6] by the author and Tuomas Hytönen shows that this is indeed the case: a non-homogeneous Tb theorem in the general framework of quasimetric spaces equipped with an upper doubling measure (this is a class of measures that encompasses both the power bounded measures, and also, the more classical doubling measures) was proved. See also [15].

It is natural to seek to extend the generality in the range too (instead of considering only scalar valued operators). These type of developments, just like the regular scalar valued Tb theorems, have a long history (for a discussion of the origins of the vector-valued Tb theory consult e.g. [9]). In the very recent work [10], a UMD-valued $T1$ theorem is established in metric spaces – however, only with Ahlfors-regular measures μ (i.e., $\mu(B(x,r)) \sim r^m$). This assumption seems to be necessary for their method of proof based on rearrangements of dyadic cubes. In [9] a vector-valued non-homogeneous Tb theorem is proved in the case of the domain being $\mathbb{R}^n$ and the relevant measure μ being power bounded (that is, $\mu(B(x,r)) \le Cr^m$).

Mathematics Subject Classification (2010): Primary 42B20; Secondary 30L99, 46B09, 46E40, 60D05, 60G46.
Keywords: Calderón–Zygmund operator, non-doubling measure, probabilistic constructions in metric spaces, martingale difference, paraproduct.

The methods of [9] are already less dependent on the structure of $\mathbb{R}^n$ than much of the earlier vector-valued work, thus foreshadowing the possibility of extending to more general domains. The goal here is to carefully combine key techniques from the recent developments [6] and [9] and obtain a proof of a non-homogeneous Tb theorem, which is simultaneously general with respect to the domain (a metric space), the measure (an upper doubling measure) and the range (a UMD Banach space).

2. Preliminaries and the main result

2.1. Geometrically doubling quasimetric spaces

A quasimetric space (X, ρ) is geometrically doubling if every open ball $B(x, r) = \{y \in X : \rho(y, x) < r\}$ can be covered by at most N balls of radius $r/2$. Our proof requires that we impose this geometric condition. A basic observation is that in a geometrically doubling quasimetric space, a ball $B(x, r)$ can contain the centers x_i of at most $N\alpha^{-n}$ disjoint balls $B(x_i, \alpha r)$ for $\alpha \in (0, 1]$.

For many purposes, quasimetrics are just as good as metrics, only somewhat more technical to deal with. However, some of the more delicate estimates in [6] require the following regularity condition: for every $\epsilon > 0$ there exists $A(\epsilon) < \infty$ so that

$$\rho(x, y) \leq (1 + \epsilon)\rho(x, z) + A(\epsilon)\rho(z, y).$$

Notice that this property is in particular satisfied by all positive powers of an honest metric, and every quasimetric is equivalent to one of that form by a result of Macías and Segovia [11].

We want to reduce the proof of our main theorem to the case of metric spaces. All of our assumptions, except possibly for the weak boundedness property (for the definition, see §2.5 below), are stable under the change to d if ρ is equivalent with d^β −see §3 of [6]. As noticed in [6], the main problem with the reduction is that the weak boundedness property is formulated using balls defined by the given quasimetric ρ. Specifically, the weak boundedness property seems difficult to transfer for any other type of sets than those for which it is assumed (even for d-balls). This problem was circumvented in [6] by explicitly constructing a certain random covering using ρ-balls instead of d-balls (the details of this construction require the regularity of ρ). Naturally we need to assume this regularity condition also in the present paper as it is already needed in the simpler scalar case.

2.2. Upper doubling measures

A Borel measure μ in some quasimetric space (X, ρ) is called upper doubling if there exists a dominating function $\lambda\colon X \times (0, \infty) \to (0, \infty)$ so that $r \mapsto \lambda(x, r)$ is non-decreasing, $\lambda(x, 2r) \leq C_\lambda \lambda(x, r)$ and $\mu(B(x, r)) \leq \lambda(x, r)$ for all $x \in X$ and $r > 0$. The number $d := \log_2 C_\lambda$ can be thought of as (an upper bound for) a dimension of the measure μ, and it will play a similar role as the quantity denoted by the same symbol in [13].

2.3. Standard kernels and Calderón–Zygmund operators

Define $\Delta = \{(x,x) : x \in X\}$. A standard kernel is a mapping $K\colon X^2 \setminus \Delta \to \mathbb{C}$ for which we have for some $\alpha > 0$ and $B, C < \infty$ that

$$|K(x,y)| \le B \min\Big(\frac{1}{\lambda(x,\rho(x,y))}, \frac{1}{\lambda(y,\rho(x,y))}\Big), \quad x \ne y,$$

$$|K(x,y) - K(x',y)| \le B\, \frac{\rho(x,x')^\alpha}{\rho(x,y)^\alpha \lambda(x,\rho(x,y))}, \quad \rho(x,y) \ge C\rho(x,x'),$$

and

$$|K(x,y) - K(x,y')| \le B\, \frac{\rho(y,y')^\alpha}{\rho(x,y)^\alpha \lambda(y,\rho(x,y))}, \quad \rho(x,y) \ge C\rho(y,y').$$

The smallest admissible B will be denoted by $\|K\|_{CZ_\alpha}$; it is understood that the parameter C has been fixed, and it will not be indicated explicitly in this notation.

Let $T\colon f \mapsto Tf$ be a bounded linear operator $L^2(X) \to L^2(X)$. It is called a Calderón–Zygmund operator with kernel K if

$$Tf(x) = \int_X K(x,y) f(y)\, d\mu(y)$$

for x outside the support of f.

2.4. Accretivity

A function $b \in L^\infty(\mu)$ is called accretive if $|\int_A b\, d\mu| \ge a\mu(A)$ for all Borel sets A which satisfy the condition that $B \subset A \subset CB$ for some ball $B = B(A)$, where C is some large constant which depends on the quasimetric ρ. (We note that, e.g., $C = 500$ will do in the case that ρ is a metric. Otherwise, one defines β via the equation $2^\beta = 3A_0^2$, where A_0 is the constant from the triangle inequality of ρ, and then $C = 4000^\beta$ suffices. These details are in §3 of [6].)

The point is to have the above estimate whenever A is a ball or one of the metric dyadic cubes (even after switching to an equivalent metric), but there is no easy explicit description of what kind of sets they actually are. Because of this, our formulation of accretivity is technical. However, notice that at least the classical condition $\mathrm{Re}\, b \ge a > 0$ implies it (with any set A).

2.5. Weak boundedness property

Let $\Lambda > 1$. Suppose that for every ball B and every $\epsilon \in (0,1]$ there is a function $\tilde{\chi}_{B,\epsilon}$ such that $\chi_B \le \tilde{\chi}_{B,\epsilon} \le \chi_{(1+\epsilon)B}$, and we have the estimate $|\langle T\tilde{\chi}_{B,\epsilon}, \tilde{\chi}_{B,\epsilon}\rangle| \le S(\epsilon)A\mu(\Lambda B)$ with some $S(\epsilon) < \infty$ independent of the other quantities. Here $\langle \cdot\,, \cdot\rangle$ is the bilinear duality $\langle f, g\rangle = \int fg\, d\mu$. We denote the smallest admissible A by $\|T\|_{WBP_{\Lambda,S}}$. Note that this notion of the weak boundedness property simply asks the above inequality for some set of functions $\tilde{\chi}_{B,\epsilon}$, regular or not. Depending on the structure of the underlying space X, one may find such functions with different degrees of regularity.

Note that if T satisfies $|\langle T\chi_B, \chi_B\rangle| \le \|T\|_{WBP_\Lambda}\mu(\Lambda B)$ for all balls B (where $\|T\|_{WBP_\Lambda}$ is the best possible constant for that inequality), then $\|T\|_{WBP_{\Lambda,1}} \le \|T\|_{WBP_\Lambda}$. This is because we can take $\tilde{\chi}_{B,\epsilon} := \chi_B$ and $S(\epsilon) := 1$ for all $\epsilon > 0$.

In the Tb theorem, the weak boundedness property is demanded from the operator $M_{b_2}TM_{b_1}$, where b_1 and b_2 are accretive functions and $M_b\colon f \mapsto bf$.

2.6. BMO and RBMO

We say that $f \in L^1_{\mathrm{loc}}(\mu)$ belongs to $\mathrm{BMO}^p_\kappa(\mu)$, if for any ball $B \subset X$ there exists a constant f_B such that

$$\Big(\int_B |f - f_B|^p \, d\mu\Big)^{1/p} \le L\mu(\kappa B)^{1/p},$$

where the constant L does not depend on B.

For $b \in L^\infty(\mu)$ one can define

$$\langle Tb, f\rangle = \langle T(\chi_{2B}b), f\rangle + \int_{(2B)^c} b(x)T^*f(x)\, d\mu(x)$$

say for every essentially bounded f which is supported in a ball B and satisfies $\int f\, d\mu = 0$. The integral over $(2B)^c$ converges by the kernel estimates. The pairing $\langle T(\chi_{2B}b), f\rangle$ makes sense, since $T\colon L^2(X) \to L^2(X)$. Now the condition $Tb \in \mathrm{BMO}^p_\kappa(\mu)$ is defined to mean that $|\langle Tb, f\rangle| \le L\|f\|_{L^{p'}(\mu)}\mu(\kappa B)^{1/p}$ for every f like before.

Let $\varrho > 1$. A function $f \in L^1_{\mathrm{loc}}(\mu)$ belongs to $\mathrm{RBMO}(\mu)$ if there exists a constant L, and for every ball B, a constant f_B, such that one has

$$\int_B |f - f_B|\, d\mu \le L\mu(\varrho B),$$

and, whenever $B \subset B_1$ are two balls,

$$|f_B - f_{B_1}| \le L\Big(1 + \int_{2B_1\setminus B} \frac{1}{\lambda(c_B, \rho(x, c_B))}\, d\mu(x)\Big).$$

We do not demand that f_B be the average $\langle f\rangle_B = \frac{1}{\mu(B)}\int_B f\, d\mu$, and this is actually important in the $\mathrm{RBMO}(\mu)$-condition. The useful thing here is that the space $\mathrm{RBMO}(\mu)$ is independent of the choice of parameter $\varrho > 1$ and satisfies the John–Nirenberg inequality. For these results in our setting, see [8]. The norms in these spaces are defined in the obvious way as the best constant L.

2.7. UMD Banach spaces

A Banach space Y is said to satisfy the UMD property if there holds that

$$\Big\|\sum_{k=1}^n \epsilon_k d_k\Big\|_{L^p(\Omega, Y)} \le C\Big\|\sum_{k=1}^n d_k\Big\|_{L^p(\Omega, Y)}$$

whenever $(d_k)_{k=1}^n$ is a martingale difference sequence in $L^p(\Omega, Y)$ and $\epsilon_k = \pm 1$ are constants. This property does not depend on the parameter $1 < p < \infty$ in any way. Moreover, it is standard knowledge that the dual space Y^* of a UMD space Y is also UMD.

2.8. Vinogradov notation and implicit constants

The notation $f \lesssim g$ is used synonymously with $f \le Gg$ for some constant G. We also use $f \sim g$ if $f \lesssim g \lesssim f$. The dependence on the various parameters should be somewhat clear, but basically G may depend on the various constants of the avove definitions, and on an auxiliary parameter r (which is eventually fixed to depend on the above parameters only).

We now state our main theorem. We will discuss the roles of some of the assumptions in Remark 2.2 below. There we will comment on the usage of different BMO spaces and weak boundedness assumptions and the role of the a priori boundedness assumption.

Theorem 2.1. *Let (X, ρ) be a geometrically doubling regular quasimetric space which is equipped with an upper doubling measure μ. Let Y be a* UMD *space and $1 < p < \infty$. Let T be an $L^p(X)$-bounded Calderón–Zygmund operator with a standard kernel K, let b_1 and b_2 be two essentially bounded accretive functions, let $\alpha > 0$ and $\kappa, \Lambda > 1$ be constant, and let $S\colon (0,1] \to (0,\infty)$ be a function. Then there holds*

$$\|T\| \lesssim \|Tb_1\|_{\mathrm{BMO}^2_\kappa(\mu)} + \|T^*b_2\|_{\mathrm{BMO}^2_\kappa(\mu)} + \|M_{b_2}TM_{b_1}\|_{\mathrm{WBP}_{\Lambda,\mathrm{S}}} + \|K\|_{CZ_\alpha},$$

where $\|T\| = \|T\|_{L^p(X,Y)\to L^p(X,Y)}$.

If we in addition assume that $\rho = d^\beta$ for some metric d and $\beta \ge 1$, then

$$\|T\| \lesssim \|Tb_1\|_{\mathrm{BMO}^1_\kappa(\mu)} + \|T^*b_2\|_{\mathrm{BMO}^1_\kappa(\mu)} + \|M_{b_2}TM_{b_1}\|_{\mathrm{WBP}_{\Lambda}} + \|K\|_{CZ_\alpha}.$$

Remark 2.2. In both of the scenarios we want to reduce to the case of metric spaces, update our weak boundedness property and enhance our BMO assumptions. In the first part of this remark we shall discuss this. The order of the reductions is a bit different depending on the case, but after they are done, the proofs coincide.

Note that in the case of regular quasimetric spaces the Theorem 2.1 is formulated using the space $\mathrm{BMO}^2_\kappa(\mu)$. This is the case also for the scalar valued analog Theorem 2.10 in [6]. Indeed, we may use Theorem 2.10 of [6] to conclude that under the assumptions of Theorem 2.1 we have the following quantitative scalar-valued L^2 operator norm bound:

$$\|T\|_{L^2(X)\to L^2(X)} \lesssim \|Tb_1\|_{\mathrm{BMO}^2_\kappa(\mu)} + \|T^*b_2\|_{\mathrm{BMO}^2_\kappa(\mu)} + \|M_{b_2}TM_{b_1}\|_{\mathrm{WBP}_{\Lambda,\mathrm{S}}} + \|K\|_{CZ_\alpha}.$$

Using this we may immediately strengthen our weak boundedness property. In fact, we have $|\langle T(\chi_A b_1), \chi_A b_2\rangle| \lesssim \mu(A)$ with any Borel set $A \subset X$. In particular, we may then completely reduce to the case of metric spaces ($\rho = d$ for some metric d). This reduction is done in §3 of [6]. Note that the original weak boundedness property would not need to transfer to d-balls. However, we just noted that the work done in the scalar case gives us the much stronger weak boundedness property.

Let us explain the point of the assumption $\rho = d^\beta$, and why this allows the usage of the larger space $\mathrm{BMO}^1_\kappa(\mu)$. Notice however the usage of the stronger notion of the weak boundedness property $\|M_{b_2}TM_{b_1}\|_{WBP_\Lambda} < \infty$ in this case. Since $\rho = d^\beta$, we have that d-balls are ρ-balls. Thus, we have that $|\langle T(\chi_B b_1), \chi_B b_2)\rangle| \lesssim \mu(\Lambda B)$ for every d-ball B. After this observation one may assume that $\rho = d$, since every other assumption always transfers. In *metric* spaces (Theorem 3.3) we are able to show that

$$\|Tb_1\|_{\mathrm{BMO}^q_\kappa(\mu)} \lesssim \|K\|_{CZ_\alpha} + \|Tb_1\|_{\mathrm{BMO}^1_\kappa(\mu)} + \|M_{b_2}TM_{b_1}\|_{\mathrm{WBP}_\Lambda}, \quad 1 \le q < \infty.$$

In particular, $Tb_1, T^*b_2 \in \mathrm{BMO}^2_\kappa(\mu)$ so we can even in this case infer from Theorem 2.10 in [6] that $|\langle T(\chi_A b_1), \chi_A b_2\rangle| \lesssim \mu(A)$ with any Borel set $A \subset X$.

Therefore, with either set of assumptions we can eventually work in a metric space with the weak boundedness property enhanced to $|\langle T(\chi_A b_1), \chi_A b_2\rangle| \lesssim \mu(A)$ for any Borel set $A \subset X$, and the BMO assumptions enhanced to $Tb_1, T^*b_2 \in \mathrm{BMO}^q_\kappa(\mu)$ with any $1 \le q < \infty$. We consider this done after §3, and then prove everything only in this context.

Let us now also comment on the a priori boundedness assumption. Notice that we only assume the scalar boundedness $T\colon L^p(X) \to L^p(X)$. However, in the proof we want to actually be able to assume that $\|T\| = \|T\|_{L^p(X,Y)\to L^p(X,Y)} < \infty$. Let us, for this argument, assume that we have shown Theorem 2.1 with this extra assumption. Consider then any simple function $f = \sum_{n=1}^N \chi_{A_n} y_n$, where $y_n \in Y$ and $A_n \subset X$ with $\mu(A_n) < \infty$. It follows from the scalar boundedness that $T\colon L^p(X,E) \to L^p(X,E)$ boundedly for every finite-dimensional subspace $E \subset Y$. Certainly this trivial bound depends on the dimension, but it has no relevance since this information is used purely qualitatively. Note that subspaces $E \subset Y$ are UMD with their UMD constants uniformly bounded by the UMD constant of Y. Therefore, we may conclude that

$$\|Tf\|_{L^p(X,Y)} = \|Tf\|_{L^p(X,E)} \lesssim \|f\|_{L^p(X,E)} = \|Tf\|_{L^p(X,Y)}$$

for $E = \mathrm{span}\{y_n : n = 1 \ldots, N\}$. Since simple functions are dense in $L^p(X,Y)$, Theorem 2.1 follows if we have proven it under this extra assumption.

The scalar boundedness is a separate issue. Certainly most of the time the point of vector-valued Tb or $T1$ theorems is just to ensure that bounded operators extend to bounded vector-valued operators (see the basic usage in Example 2.3 below). Indeed, the point is that one does not know how to use the scalar-valued boundedness directly to establish the vector-valued boundedness. Instead, the idea is to rely on the characterization given by Tb theorems: one proves that the hypothesis of Tb theorems are enough to actually guarantee the UMD-valued boundedness. Another important point is the established quantitative bound.

We still note that one of the standard ways to make operators automatically bounded in all L^p spaces is by suitably truncating them. Furthermore, we point out the reference [5], where Hytönen, Liu, Yang and Yang, generalizing the results of Nazarov, Treil and Volberg [12] to the upper doubling setting, deal with many standard aspects of this problem.

We give an example before proceeding with the proof of the theorem.

Example 2.3. In Chapter 12 of [6] we gave an example related to the paper [16], and there the application was in a situation where the measure in question was genuinely upper doubling (the doubling theory or the theory of power bounded measures would not have sufficed), and the space was a quasimetric one (so it really was non-homogeneous theory on metric spaces).

Now we give an example which is actually in the homogeneous situation, but as the domain is a metric space and the range is a general UMD space, this seems not to follow from the previous works. Also, it goes to show that it is convenient to get this doubling theory as a byproduct of the upper doubling theory.

The example we have in mind is the boundedness of the classical Cauchy–Szegö projection as a UMD-valued operator (this question was asked by Tao Mei through a private communication with Tuomas Hytönen, and Mei had solved this question in the special case when the range space Y is a so-called non-commutative L^p space). The setting is the Heisenberg group $\mathbb{H}^n$, which is identified with $\mathbb{R}^{2n+1}$, and is a non-abelian group where the group operation is given by

$$x \cdot y = (x_1 + y_1, \ldots, x_{2n} + y_{2n}, x_{2n+1} + y_{2n+1} - 2\sum_{j=1}^{n}(x_j y_{j+n} - x_{j+n} y_n)).$$

The metric is given by

$$d(x,y) = \|x^{-1} \cdot y\|,$$

where

$$\|x\| = (\|(x_1, \ldots, x_{2n})\|^4_{\mathbb{R}^{2n}} + x^2_{2n+1})^{1/4}.$$

One can also write $x = [\xi, t] \in \mathbb{H}^n = \mathbb{C}^n \times \mathbb{R}$. We use the Haar measure for $\mathbb{H}^n$ (this is just the Euclidean Lebesgue measure $d\xi dt$ on $\mathbb{C}^n \times \mathbb{R}$). Now $\lambda(x,r) = Cr^{2n+2}$ for some appropriate constant C.

Using the above notation $x = [\xi, t]$, let $K(x) = C(t + i|\xi|)^{-n-1}$. Set $K(x,y) = K(y^{-1} \cdot x)$ for $x \neq y$ (i.e. $y^{-1} \cdot x \neq 0$). The Cauchy–Szegö projection C is an L^2-bounded operator of the form

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

See e.g. [14] for a more exhaustive treatment of the Cauchy–Szegö projection.

Clearly the standard kernel estimates known for K are precisely the same as demanded by our theory with our chosen λ. Thus, as C is a Calderón–Zygmund operator which is bounded as a scalar-valued operator (and thus satisfies the BMO conditions with e.g. $b_1 = b_2 = 1$ and the weak boundedness property), we have by our above Tb (or $T1$ in this case) theorem that T is a bounded operator $L^p(\mathbb{H}^n, Y) \to L^p(\mathbb{H}^n, Y)$ for every UMD space Y and for every index $p \in (1, \infty)$.

3. John–Nirenberg theorem for Tb_1

In this section we work on a geometrically doubling metric space (X,d) with an upper doubling measure μ. We are given a Calderón–Zygmund operator T with a standard kernel K and the constants $\alpha > 0$, $\kappa, \Lambda > 1$. Moreover, we have the essentially bounded accretive functions b_1 and b_2. We will show that

$$\|Tb_1\|_{\mathrm{BMO}^q_\kappa(\mu)} \lesssim \|K\|_{CZ_\alpha} + \|Tb_1\|_{\mathrm{BMO}^1_\kappa(\mu)} + \|M_{b_2}TM_{b_1}\|_{\mathrm{WBP}_\Lambda}, \quad 1 \le q < \infty. \tag{3.1}$$

Of course, the analogous result holds with T^*b_2.

This reduction is known in the Euclidean setting with a power bounded measure (see [13]). We now work out the details in our setting.

Lemma 3.1. *Consider a fixed ball $B = B(c_B, r_B)$. There exists $R_B \in [r_B, 1.2r_B]$ so that, for all $s \in [0, 3/2]$,*

$$\mu(\{x \in X : R_B - r_B s < d(x, c_B) < R_B + r_B s\}) \lesssim s\, \mu(B(c_B, 3r_B))\,.$$

Proof. See page 184 in [13]. □

Lemma 3.2. *If $B = B(c_B, r_B)$ is a ball and R_B is a related regularized radius as in the previous lemma, then it holds that*

$$\begin{aligned}\int_{B(c_B,R_B)} \int_{B(c_B,3r_B)\setminus B(c_B,R_B)} &|K(x,y)| d\mu(y)\, d\mu(x) \\ &\lesssim \mu(B(c_B,R_B))^{1/2} \mu(B(c_B,3r_B))^{1/2} \\ &\le \mu(B(c_B,3r_B)).\end{aligned}$$

Proof. Consider $f(x) = \int_{B(c_B,3r_B)\setminus B(c_B,R_B)} |K(x,y)|\, d\mu(y)$, $x \in B(c_B,R_B)$. Fix $x \in B(c_B,R_B)$ for the moment and note that we have for all $y \in B(c_B,3r_B) \setminus B(c_B,R_B)$ that $d(x,y) \le R_B + 3r_B \le 4.2r_B < 5r_B$ and $d(x,y) \ge d(y,c_B) - d(x,c_B) \ge R_B - d(x,c_B)$. We temporarily set $h = R_B - d(x,c_B)$ for this fixed x and estimate

$$\begin{aligned}f(x) &\lesssim \int_{h \le d(x,y) < 5r_B} \frac{d\mu(y)}{\lambda(x, d(x,y))} \le \sum_{1 \le j < \log_2(10r_B/h)} \int_{2^{j-1}h \le d(x,y) < 2^j h} \frac{d\mu(y)}{\lambda(x,d(x,y))} \\ &\le \sum_{1 \le j < \log_2(10r_B/h)} \frac{\mu(B(x,2^j h))}{\lambda(x, 2^{j-1}h)} \lesssim \log(10r_B/h) = \log\Big(\frac{10r_B}{R_B - d(x,c_B)}\Big).\end{aligned}$$

This implies through Hölder's inequality that

$$\begin{aligned}\int_{B(c_B,R_B)} & f(x)\, d\mu(x) \\ &\lesssim \mu(B(c_B,R_B))^{1/2} \Big(\int_{B(c_B,R_B)} \Big[\log\Big(\frac{10r_B}{R_B - d(x,c_B)}\Big)\Big]^2 d\mu(x)\Big)^{1/2}.\end{aligned}$$

We then continue to note that

$$\int_{B(c_B,R_B)} \Big[\log\Big(\frac{10r_B}{R_B - d(x,c_B)}\Big)\Big]^2 \, d\mu(x)$$

equals

$$\int_0^\infty \mu\Big(\Big\{x \in B(c_B,R_B) : \Big[\log\Big(\frac{10r_B}{R_B - d(x,c_B)}\Big)\Big]^2 > t\Big\}\Big)\, dt,$$

which in turn equals

$$\int_0^\infty \mu\big(\{x : R_B - 10r_B e^{-\sqrt{t}} < d(x,c_B) < R_B\}\big)\, dt = \Big[\log\Big(\frac{10r_B}{R_B}\Big)\Big]^2 \mu(B(c_B,R_B)) + \int_{[\log(10r_B/R_B)]^2}^\infty \mu\big(\{x : R_B - 10r_B e^{-\sqrt{t}} < d(x,c_B) < R_B\}\big)\, dt.$$

Note that $\int_0^\infty e^{-\sqrt{t}}\, dt = 2$ and use the previous lemma with $s = 10e^{-\sqrt{t}} \leq R_B/r_B \leq 1.2 < 1.5$ for $t \geq [\log(10r_B/R_B)]^2$ to get that

$$\int_{[\log(10r_B/R_B)]^2}^\infty \mu(\{x : R_B - 10r_B e^{-\sqrt{t}} < d(x,c_B) < R_B\})\, dt \lesssim \mu(B(c_B,3r_B)).$$

This yields the claim. $\square$

Theorem 3.3. *Under the assumptions stated at the beginning of this section, the estimate* (3.1) *holds.*

Proof. We will begin by proving that $Tb_1 \in \mathrm{RBMO}(\mu)$ with the following interpretation. If $d(X) < \infty$, we set $B_0 = X$. If $d(X) = \infty$, we write $X = \bigcup_{i=1}^\infty B_i$ so that $B_1 \subset B_2 \subset \cdots$, and $1.1B_i \neq B_i$ for every $i = 1, 2, \ldots$ We consider some fixed i and set $B_0 = B_i$. We need to show that the function $T(\chi_{100B_0} b_1)$ satisfies the defining properties of the $\mathrm{RBMO}(\mu)$ space for all the balls that are subset of B_0, and in such a way that the $\mathrm{RBMO}(\mu)$ norm does not depend on B_0.

Notice that $1.1B_0 \neq B_0$ implies that if $B \subset B_0$ is a ball, then $aB \subset (2.1a+1)B_0$. We define $f_B = T(b_1\chi_{100B_0\setminus 4B})(c_B)$ if $B \subset B_0$ is a ball.

We note that

$$|T(b_1\chi_{100B_0\setminus 4B})(x) - T(b_1\chi_{100B_0\setminus 4B})(c_B)| \lesssim 1,$$

if $B \subset B_0$ is a ball and $x \in B$. Indeed, we have

$$\begin{aligned} &|T(b_1\chi_{100B_0\setminus 4B})(x) - T(b_1\chi_{100B_0\setminus 4B})(c_B)| \\ &\lesssim \int_{100B_0\setminus 4B} |K(x,y) - K(c_B,y)|\, d\mu(y) \lesssim r_B^\alpha \int_{X\setminus B} \frac{d(c_B,y)^{-\alpha}}{\lambda(c_B,(d(c_B,y))}\, d\mu(y) \lesssim 1. \end{aligned}$$

The last estimate follows from Lemma 2.2 of [6].

Let $B \subset B_1 \subset B_0$ be balls. Notice that

$$\int_B |T(b_1\chi_{100B_0}) - f_B|\, d\mu \lesssim \int_B |T(b_1\chi_{4B})|\, d\mu + \mu(B),$$

since, like noted above, there holds

$$|T(b_1\chi_{100B_0\setminus 4B})(x) - f_B| = |T(b_1\chi_{100B_0\setminus 4B})(x) - T(b_1\chi_{100B_0\setminus 4B})(c_B)| \lesssim 1$$

for every $x \in B$. Moreover, we have that $|f_B - f_{B_1}| = |T(b_1\chi_{100B_0\setminus 4B})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_{B_1})|$, and this can be dominated by

$$|T(b_1\chi_{100B_0\setminus 4B})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_B)| \\ + |T(b_1\chi_{100B_0\setminus 4B_1})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_{B_1})|.$$

Notice that again $|T(b_1\chi_{100B_0\setminus 4B_1})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_{B_1})| \lesssim 1$.

One has to be careful when estimating

$$|T(b_1\chi_{100B_0\setminus 4B})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_B)|.$$

Let us first study the case $B = B_1$ (as sets, that is, a ball is defined by a pair (c, r)). This does not necessarily imply anything particular about c_B, c_{B_1} and r_B, r_{B_1}. We split this into two cases: $4B_1 \subset 4B$ or $4B_1 \not\subset 4B$. If $4B_1 \not\subset 4B$, then $4B \subset 6B_1 \subset 9B_1$. If $B \subsetneq B_1$, then also necessarily $4B \subset 9B_1$. We deduce that if $B \subset B_1$, then at least one of the inclusions $4B_1 \subset 4B$ or $4B \subset 9B_1$ is true.

Let us first estimate $|T(b_1\chi_{100B_0\setminus 4B})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_B)|$ when we know that $B \subset B_1$ and $4B_1 \subset 4B$. We write

$$\begin{aligned}
&|T(b_1\chi_{100B_0\setminus 4B})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_B)| \\
&= \Big| \int_{100B_0\setminus 4B} K(c_B, y) b_1(y)\, d\mu(y) - \int_{100B_0\setminus 4B_1} K(c_B, y) b_1(y)\, d\mu(y) \Big| \\
&= \Big| \int_{4B\setminus 4B_1} K(c_B, y) b_1(y)\, d\mu(y) \Big| \\
&= \Big| \int_{4B_1\setminus B} K(c_B, y) b_1(y)\, d\mu(y) - \int_{4B\setminus B} K(c_B, y) b_1(y)\, d\mu(y) \Big| \\
&\lesssim \int_{4B_1\setminus B} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))} + \int_{4B\setminus B} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))} \\
&= \int_{2B_1\setminus B} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))} + \int_{4B_1\setminus 2B_1} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))} + \int_{4B\setminus B} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))}.
\end{aligned}$$

Then we note that

$$\int_{4B_1\setminus 2B_1} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))} \le \frac{\mu(4B_1)}{\lambda(c_B, r_{B_1})} \le \frac{\mu(B(c_B, 5r_{B_1}))}{\lambda(c_B, r_{B_1})} \lesssim 1$$

and

$$\int_{4B\setminus B} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))} \le \frac{\mu(4B)}{\lambda(c_B, r_B)} \lesssim 1.$$

We have shown that

$$|f_B - f_{B_1}| \lesssim 1 + \int_{2B_1\setminus B} \frac{d\mu(y)}{\lambda(c_B, d(c_B, y))}$$

if $B \subset B_1$ and $4B_1 \subset 4B$.

Let us then study the more natural case $B \subset B_1$ and $4B \subset 9B_1$. Now we have that

$$\begin{aligned}
&|T(b_1\chi_{100B_0\setminus 4B})(c_B) - T(b_1\chi_{100B_0\setminus 4B_1})(c_B)| \\
&\quad = \Big| \int_{9B_1\setminus 4B} K(c_B,y)b_1(y)\,d\mu(y) - \int_{9B_1\setminus 4B_1} K(c_B,y)b_1(y)\,d\mu(y)\Big| \\
&\quad \lesssim \int_{2B_1\setminus B} \frac{d\mu(y)}{\lambda(c_B,d(c_B,y))} + \int_{9B_1\setminus 2B_1} \frac{d\mu(y)}{\lambda(c_B,d(c_B,y))},
\end{aligned}$$

and

$$\int_{9B_1\setminus 2B_1} \frac{d\mu(y)}{\lambda(c_B,d(c_B,y))} \le \frac{\mu(9B_1)}{\lambda(c_B,r_{B_1})} \le \frac{\mu(B(c_B,10r_{B_1}))}{\lambda(c_B,r_{B_1})} \lesssim 1.$$

We have now established the right bound for $|f_B - f_{B_1}|$ for every ball $B \subset B_1 \subset B_0$.

Recalling the definition of $\mathrm{RBMO}(\mu)$ we notice that we have reduced to controlling $\int_B |T(b_1\chi_{4B})|\,d\mu$ for every ball $B \subset X$. We shall prove that

$$\int_B |T(\chi_{4B}b_1)|\,d\mu \lesssim \mu(\eta B)$$

for $\eta = \max(\kappa, 3, 2\Lambda)$. Given a function g so that $\|g\|_{L^\infty(\mu)} \le 1$ and $\operatorname{spt} g \subset B$ it suffices to show that $|\langle T(\chi_{4B}b_1), gb_2\rangle| \le \mu(\eta B)$. Define c by setting

$$c = \frac{\int_B gb_2\,d\mu}{\int_{\tilde B} b_2\,d\mu},$$

where $\tilde B = B(c_B, R_B)$ and R_B is a regularized radius given by Lemma 3.1. We have $|c| \lesssim \mu(B)/\mu(\tilde B) \le 1$ by the accretivity of b_2. The definition of c precisely means that $\int (gb_2 - cb_2\chi_{\tilde B})\,d\mu = 0$. Moreover, there holds $\|gb_2 - cb_2\chi_{\tilde B}\|_{L^\infty(\mu)} \lesssim 1$ and $\operatorname{spt}(gb_2 - cb_2\chi_{\tilde B}) \subset \tilde B \subset 2B$.

We now split

$$\begin{aligned}
\langle T(b_1\chi_{4B}), gb_2\rangle = \langle T(b_1\chi_{4B}), gb_2 - cb_2\chi_{\tilde B}\rangle + c\,\big(\langle T(b_1\chi_{4B\setminus\tilde B}), b_2\chi_{\tilde B}\rangle \\
+ \langle T(b_1\chi_{\tilde B}), b_2\chi_{\tilde B}\rangle\big).
\end{aligned}$$

Write $h = gb_2 - cb_2\chi_{\tilde B}$. We have that

$$\langle T(b_1\chi_{4B}), h\rangle = \langle Tb_1, h\rangle - \int_{(4B)^c} b_1(x)T^*h(x)\,d\mu(x)\,,$$

from which it follows that

$$|\langle T(b_1\chi_{4B}), h\rangle| \lesssim \mu(\kappa B) + \mu(2B) \lesssim \mu(\eta B).$$

Here it was used that $Tb_1 \in \mathrm{BMO}^1_\kappa(\mu)$. The bound $\mu(2B)$ comes from the second term via kernel estimates. Lemma 3.2 yields that $|\langle T(b_1\chi_{4B\setminus\tilde B}), b_2\chi_{\tilde B}\rangle| \lesssim \mu(3B)$.

The fact that here we have $4B \setminus \tilde{B}$ instead of $3B \setminus \tilde{B}$ makes absolutely no difference, since

$$\int_{\tilde{B}} \int_{4B\setminus 3B} \frac{1}{\lambda(x, d(x,y))}\, d\mu(y)\, d\mu(x) \le \int_{\tilde{B}} \frac{\mu(B(x, 6r_B))}{\lambda(x, r_B)}\, d\mu(x) \lesssim \mu(2B).$$

Finally, the property $\|M_{b_2} T M_{b_1}\|_{WBP_\Lambda} < \infty$ yields $|\langle T(b_1\chi_{\tilde{B}}), b_2\chi_{\tilde{B}}\rangle| \lesssim \mu(2\Lambda B)$.

We have established that $Tb_1 \in \mathrm{RBMO}(\mu)$. We have by Corollary 6.3 of [8] that the $\mathrm{BMO}^q_\rho(\mu)$ norm can be dominated by the $\mathrm{RBMO}(\mu)$ norm for every $q \in [1, \infty)$ and for every $\rho > 1$ (note that the corollary is, indeed, proven in the context of a geometrically doubling metric space equipped with an upper doubling measure). This, in particular, implies the estimate (3.1). □

4. Random dyadic systems and good/bad cubes

One feature of the proof in [9] is that one basically takes all the cubes to be good in the various summations – this is in contrast with the proof in [6] where things were usually summed so that the bigger cubes are arbitrary but the smaller cubes from the other grid were assumed to be good. This modification seems to be particularly useful when dealing with certain paraproducts in these general UMD spaces.

This leads us to fiddle with our randomization from [6]. We shall make the randomization so that there is no removal procedure involved (unlike in [6]) – then a certain index set may serve as a fixed reference set more conveniently. Such a modification is also used in the paper [4] by T. Hytönen and A. Kairema, where the authors, among other things, provide a streamlined version of the dyadic constructions presented in [6].

Furthermore, we will change the definition of a good cube to be such that given a cube Q its change to be good does not depend on the smaller cubes R with $\ell(R) \le \ell(Q)$. Related to this we shall also make a minor tweak to our half-open cubes from [6] (to get a better dependence on the randomized dyadic points). This is also spelled out in [4]. Finally, we add a layer of artificial badness so that $\mathbb{P}(Q \text{ is good})$ does not depend on the particular choice of the cube Q.

Remark 4.1. The basic source of these randomization techniques in metric spaces is [6]. However, as stated, we need some modifications for the purposes of this paper. When writing the first version of this article, the paper [4] was not yet available. However, I learned about the trick of avoiding the removal procedure from the authors through a private communication. Moreover, the paper [1] by P. Auscher and T. Hytönen has also become available. Striving for the most simple and state of the art approach, I will also borrow, in this revision, a few nice simplifying and expository details from [1]. Certainly we will not repeat every detail here, but we do present the basic aspects of the construction so that one may clearly follow the metric probabilistic arguments, some of which are completely new in this paper (like the pseudogoodness).

4.1. Construction of the random dyadic cubes

Let $\delta = 1/1000$. One starts by constructing a collection of points $(z_\alpha^k)_{k,\alpha}$ so that for every $k \in \mathbb{Z}$:

(i) $\{z_\alpha^k\}_\alpha \subset \{z_\beta^{k+1}\}_\beta$;

(ii) $\min_\alpha d(x, z_\alpha^k) < 2\delta^k$ for every $x \in X$;

(iii) $d(z_\alpha^k, z_\beta^k) \geq \delta^k$ for $\alpha \neq \beta$.

In page 4 of [1] it was noted that one may easily arrange the extra property (i) by allowing the harmless factor 2 in (ii). Before this, it was standard to just use some maximal δ^k-separated sets for every $k \in \mathbb{Z}$, and not to necessarily have the property (i). The property (i) is just an added convenience for us: every point z_α^k is also of the form z_β^{k+1}.

Note that the above set of points z_α^k and indices (k, α) are now fixed once and for all. Now we will further fix a transitive relation $\leq$ among the labels (k, α) as follows: each $(k+1, \beta)$ satisfies $(k+1, \beta) \leq (k, \alpha)$ for exactly one (k, α), and we have that if $d(z_\beta^{k+1}, z_\alpha^k) < \delta^k/2$, then necessarily $(k+1, \beta) \leq (k, \alpha)$, and that $(k+1, \beta) \leq (k, \alpha)$ always implies that $d(z_\beta^{k+1}, z_\alpha^k) < 2\delta^k$.

We call pairs of same generation (k, α) and (k, β) neighbours, if they have such children $(k+1, \gamma) \leq (k, \alpha)$ and $(k+1, \eta) \leq (k, \beta)$ that $d(z_\gamma^{k+1}, z_\eta^{k+1}) < \delta^k/2$. The whole idea of the randomization procedure is to replace, according to some rule, each z_α^k by some z_β^{k+1}, $(k+1, \beta) \leq (k, \alpha)$. However, we cannot allow the new dyadic points to end up arbitrarily close to each other. Hence neighbours form an obstruction, which has to be circumvented. In [6] we used a certain removal procedure. However, we will this time circumvent this particular problem by using the idea of double labels from [4] together with some simplification from [1].

Let L be the maximal number of neighbours and M be the maximal number of children a pair (k, α) can have. It follows from the geometric doubling condition that $L \lesssim 1$ and $M \lesssim 1$. We will now equip each pair (k, α) with two labels $L_1(k, \alpha)$ and $L_2(k, \alpha)$. The label $L_1(k, \alpha) \in \{0, 1, \ldots, L\}$ is chosen in a way that any two neighbours have a different label. The label $L_2(k, \alpha) \in \{0, 1 \ldots, M\}$ is chosen in a way that no two children of the same parent have the same label.

We let $\Upsilon = (\{0, 1, \ldots, L\} \times \{0, 1 \ldots, M\})^{\mathbb{Z}}$ be the underlying probability space equipped with the natural product probability measure. Given $\upsilon = (\upsilon_k)_{k \in \mathbb{Z}} \in \Upsilon$, $\upsilon_k = (\ell_{1,k}, \ell_{2,k})$, and a pair (k, α), define

$$x_\alpha^k(\upsilon) = x_\alpha^k(\upsilon_k) = \begin{cases} z_\beta^{k+1}, & \text{if } L_1(k,\alpha) = \ell_{1,k} \text{ and } (k+1,\beta) \leq (k,\alpha) \\ & \text{with } L_2(k+1,\beta) = \ell_{2,k}, \\ z_\alpha^k, & \text{if } L_1(k,\alpha) \neq \ell_{1,k} \text{ or there is no } (k+1,\beta) \leq (k,\alpha) \\ & \text{with } L_2(k+1,\beta) = \ell_{2,k}. \end{cases}$$

Notice carefully that the new dyadic points of generation k depend only on υ_k. This means by the product probability structure that the new dyadic points of different generations are independently chosen. The following list gives the other crucial, but almost immediate, properties:

1. We have for every $\upsilon \in \Upsilon$ that $x_\alpha^k(\upsilon) = z_\beta^{k+1}$ for some $(k+1,\beta) \le (k,\alpha)$.

2. Given (k,α) and a children $(k+1,\beta) \le (k,\alpha)$ there holds that $\mathbb{P}(\{\upsilon \in \Upsilon : x_\alpha^k(\upsilon) = z_\beta^{k+1}\}) \ge \pi_0$ for some absolute $\pi_0 > 0$ depending only on the geometric doubling condition.

3. For every $\upsilon \in \Upsilon$ we have $\min_\alpha d(x, x_\alpha^k(\upsilon)) < 4\delta^k$ for every $k \in \mathbb{Z}$ and $x \in X$. Moreover, there holds $d(x_\alpha^k(\upsilon), x_\beta^k(\upsilon)) \ge \delta^k/2$ if $\alpha \ne \beta$.

Fix some $k_0 \in \mathbb{Z}$. This preassigned index is used purely for technical reasons as will become clear. Given new dyadic points $x_\beta^\ell = x_\beta^\ell(\upsilon_\ell)$, we want to construct certain sets $Q_\alpha^k = Q_\alpha^k(\upsilon)$, which are called metric dyadic cubes. The original deterministic construction of sets of such type is by M. Christ [2]. Moreover, we want that if $k \le k_0$, the cube Q_α^k depends only on υ_ℓ for $\ell \ge k$.

We will now indicate the construction of the cubes. For this, we need a new relation $\le_\upsilon$. It would be possible to construct this like the relation $\le$ was constructed (just using the new dyadic points and a bit different constants). We note that the construction has some degrees of freedom, but any one way to do it is ok. Then the truth or falsity of the relation $(k+1,\beta) \le_\upsilon (k,\alpha)$ depends on υ_k and υ_{k+1}. However, we shall use the following explicit definition of $\le_\upsilon$ given in [1]: $(k+1,\beta) \le_\upsilon (k,\alpha)$ if and only if

$$d(z_\beta^{k+1}, x_\alpha^k(\upsilon_k)) < \delta^k/4$$

or

$$(k+1,\beta) \le (k,\alpha) \text{ and there is no such } \gamma \text{ that } d(z_\beta^{k+1}, x_\gamma^k(\upsilon_k)) < \delta^k/4.$$

Defined like this the truth or falsity of the relation $(k+1,\beta) \le_\upsilon (k,\alpha)$ depends only on υ_k −a fact that is by no means crucial for us, but we prefer it anyway. What is important is that this definition still implies the following:

1. If $d(x_\beta^{k+1}(\upsilon_{k+1}), x_\alpha^k(\upsilon_k)) < \delta^k/5$, then $(k+1,\beta) \le_\upsilon (k,\alpha)$.

2. If $(k+1,\beta) \le_\upsilon (k,\alpha)$, then $d(x_\beta^{k+1}(\upsilon_{k+1}), x_\alpha^k(\upsilon_k)) < 5\delta^k$.

This means that it behaves analogously to $\le$ despite the bit different definition. The property (2) is easily iterated to yield that if $(\ell,\beta) \le_\upsilon (k,\alpha)$, then

$$d(x_\beta^\ell(\upsilon_\ell), x_\alpha^k(\upsilon_k)) < 6\delta^k.$$

We may now define various dyadic cube type objects:

$$\begin{aligned}
\widehat{Q}_\alpha^k(\upsilon) &= \{x_\beta^\ell(\upsilon_\ell) : (\ell,\beta) \le_\upsilon (k,\alpha)\},\\
\bar{Q}_\alpha^k(\upsilon) &= \overline{\widehat{Q}_\alpha^k(\upsilon)},\\
\tilde{Q}_\alpha^k(\upsilon) &= \operatorname{int} \bar{Q}_\alpha^k(\upsilon),
\end{aligned}$$

where the overline means closure and int means the interior points. Notice that all of these sets depend only on υ_ℓ for $\ell \ge k$.

Just like in $\mathbb{R}^n$, instead of these closed or open dyadic cubes, we prefer to use what can be understood as half-open dyadic cubes (these will be finally denoted by $Q^k_\alpha(\upsilon)$). With this we mean that the cubes of the same generation will cover exactly the whole X and two different cubes of the same generation will be disjoint. It is true that $X = \bigcup_\alpha \bar{Q}^k_\alpha(\upsilon)$ with any $k \in \mathbb{Z}$, but there may be overlap in the boundaries. Indeed, one only knows that $\bar{Q}^k_\alpha(\upsilon) \cap \tilde{Q}^k_\beta(\upsilon) = \emptyset$ for $\alpha \neq \beta$.

This further minor tuning of the existing sets will worsen the dependence on υ. However, this problem is not too big, since one is free to choose the k_0, and one may perform the tuning so that the dependence remains the same for $k \leq k_0$. Assume (by enumeration) that the pairs (k, α) are parametrized by $\alpha \in \mathbb{N}$ for each $k \in \mathbb{Z}$. We set

$$Q^{k_0}_0(\upsilon) = \bar{Q}^{k_0}_0(\upsilon), \quad Q^{k_0}_\alpha(\upsilon) = \bar{Q}^{k_0}_\alpha(\upsilon) \setminus \bigcup_{\beta=0}^{\alpha-1} Q^{k_0}_\beta(\upsilon), \alpha \geq 1.$$

For $k < k_0$ we set

$$Q^k_\alpha(\upsilon) = \bigcup_{\beta\colon (k_0,\beta) \leq_\upsilon (k,\alpha)} Q^{k_0}_\beta(\upsilon).$$

Notice that, indeed, these sets still depend only on υ_ℓ for $\ell \geq k$. The way to define $Q^k_\alpha(\upsilon)$ for $k > k_0$ is by induction. The easy details are spelled out in Theorem 4.4 of [6] for $k_0 = 0$ and in Lemma 2.18 of [4] with general $k_0 \in \mathbb{Z}$ (the difference being absolutely trivial). We shall not repeat this, since what we have said is enough to thoroughly understand the dependence on υ_ℓ at least for all the cubes of generation $k \leq k_0$, and this will be enough for us (we do not actually use arbitrarily small cubes at all). Let us formulate the dyadic structure of these sets as a proposition. For a verification that our definitions actually yield these properties, the most thorough reference by now is [4].

Proposition 4.2. *For any fixed υ the cubes $Q^k_\alpha(\upsilon)$ satisfy: for every $k \in \mathbb{Z}$ we have*

$$X = \bigcup_\alpha Q^k_\alpha(\upsilon);$$

for every $k \in \mathbb{Z}$ and $\ell \geq k$ there holds that either $Q^k_\alpha(\upsilon) \cap Q^\ell_\beta(\upsilon) = \emptyset$ or $Q^\ell_\beta(\upsilon) \subset Q^k_\alpha(\upsilon)$, and for every $\ell \geq k$ we have

$$Q^k_\alpha(\upsilon) = \bigcup_{\beta:(\ell,\beta) \leq_\upsilon (k,\alpha)} Q^\ell_\beta(\upsilon).$$

We also have that $\tilde{Q}^k_\alpha(\upsilon) \subset Q^k_\alpha(\upsilon) \subset \bar{Q}^k_\alpha(\upsilon)$. Actually, there holds $\operatorname{int} Q^k_\alpha(\upsilon) = \tilde{Q}^k_\alpha(\upsilon)$ and $\overline{Q^k_\alpha(\upsilon)} = \bar{Q}^k_\alpha(\upsilon)$. Finally, we note that $B(x^k_\alpha(\upsilon), C_1\delta^k) \subset Q^k_\alpha(\upsilon)$ and the diameter $d(Q^k_\alpha(\upsilon)) < C_0\delta^k$ for $C_0 = 10$ and $C_1 = 1/10$, say. We now define $\ell(Q^k_\alpha(\upsilon)) = \delta^k$ – a constant that depends only on the generation $\operatorname{gen}(Q^k_\alpha(\upsilon)) = k$. Therefore, many things can be formulated completely equivalently using either the "sidelength" $\ell(Q^k_\alpha(\upsilon))$ or generation $\operatorname{gen}(Q^k_\alpha(\upsilon))$.

4.2. Probabilistic notions: geometric goodness, pseudogoodness and goodness

The following crucial lemma, which states the small boundary layer property, is based on the independence of different υ_k and the following: given (k,α) and $(k+1,\beta) \le (k,\alpha)$, there holds that $\mathbb{P}(x_\alpha^k = z_\beta^{k+1}) \ge \pi_0 > 0$. Recall that this property was already stated before. Now the same proof as in Lemma 10.1 of [6] also gives us the same result with this modified randomization. That is, we have:

Lemma 4.3. *For any fixed $x \in X$ and $k \in \mathbb{Z}$, there holds*

$$\mathbb{P}\Big(x \in \bigcup_\alpha \delta_{Q_\alpha^k}\Big) \lesssim \epsilon^\eta$$

for some $\eta > 0$. Here $\delta_{Q_\alpha^k} = \{y : d(y, Q_\alpha^k) \le \epsilon\ell(Q_\alpha^k)$ and $d(y, X \setminus Q_\alpha^k) \le \epsilon\ell(Q_\alpha^k)\}$.

Proof. We point out a different source for the proof than Lemma 10.1 of [6]. It is page 7 of [1], and the proof there is particularly nice. It is formulated in a slightly different way. To see that it can be applied, one only has to notice that $\bigcup_\alpha \delta_{Q_\alpha^k(\upsilon)} \subset \bigcup_\alpha \delta_{k,\alpha}(\upsilon)$, where

$$\delta_{k,\alpha}(\upsilon) = \{y \in \bar{Q}_\alpha^k(\upsilon) : d(y, X \setminus \tilde{Q}_\alpha^k(\upsilon)) \le \epsilon\delta^k\}.$$

Therefore, it suffices to show that $\mathbb{P}(x \in \cup_\alpha \delta_{k,\alpha}(\upsilon)) \lesssim \epsilon^\eta$ for any fixed $x \in X$ and $k \in \mathbb{Z}$. This is what is shown in page 7 of [1]. □

We are now given two independent copies Υ and Υ' of the probability space $(\{0,1,\dots,L\} \times \{0,1\dots,M\})^{\mathbb{Z}}$, and we denote $\upsilon \in \Upsilon$ and $\upsilon' \in \Upsilon'$. This generates two independent dyadic systems $\mathcal{D} = \mathcal{D}_\upsilon = \{Q_\alpha^k(\upsilon)\} = \{Q_\alpha^k\}$ and $\mathcal{D}' = \mathcal{D}_{\upsilon'} = \{R_\alpha^k(\upsilon')\} = \{R_\alpha^k\}$ (we use the notation R_α^k just to distinguish the grids more easily). We denote the related points $x_\alpha^k = x_\alpha^k(\upsilon_k)$ and $y_\alpha^k = y_\alpha^k(\upsilon_k')$.

We set

$$\gamma := \frac{\alpha}{2(\alpha+d)},$$

where we recall that $d := \log_2 C_\lambda$ in our setting. We now want to define what we mean by geometric goodness. We want that if $Q \in \mathcal{D}$ is geometrically $\mathcal{D}'$-good, then for every $R \in \mathcal{D}'$ for which $\ell(Q) \le \delta^r\ell(R)$ we have either $d(Q,R) \gtrsim \ell(Q)^\gamma\ell(R)^{1-\gamma}$ or $d(Q, X\setminus R) \gtrsim \ell(Q)^\gamma\ell(R)^{1-\gamma}$. However, if we define it like this, then for every Q the condition in particular depends on υ'_ℓ for every ℓ. However, for technical reasons, we need it to only depend on υ'_ℓ for $\ell < \mathrm{gen}(Q)$. We now give a technical definition, which only has the aforementioned dependence. Then we show that it implies the much more natural geometric condition – the proof of this fact should be the key to understanding the sufficiency of the technical definition.

Definition 4.4. We say that $Q_\alpha^k \in \mathcal{D}$ is geometrically $\mathcal{D}'$-bad, if there exists $(k-s,\beta) \ne (k-s,\gamma)$ for some $s \ge r$ so that for some $(k-1,\eta) \le_{\upsilon'} (k-s,\beta)$ and $(k-1,\xi) \le_{\upsilon'} (k-s,\gamma)$ we have $d(x_\alpha^k, y_\eta^{k-1}) \le \delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$ and $d(x_\alpha^k, y_\xi^{k-1}) \le \delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$. Otherwise Q_α^k is geometrically $\mathcal{D}'$-good.

Remark 4.5. Note that the geometric goodness is actually a property of the center $x^k_\alpha(\upsilon_k)$ and not of the cube. This notion depends on υ_k and υ'_ℓ for $\ell < k$.

The usage of this lesser dependence will be as follows. Suppose $\mathrm{gen}(Q) \leq \mathrm{gen}(R) \leq k_0$, that is $\delta^{k_0} \leq \ell(R) \leq \ell(Q)$ for $Q \in \mathcal{D}$ and $R \in \mathcal{D}'$. The cube R as a set depend on υ'_ℓ for $\ell \geq \mathrm{gen}(R)$. The geometric goodness of Q depends on υ'_ℓ for $\ell < \mathrm{gen}(Q) \leq \mathrm{gen}(R)$. So in view of the probability $\mathbb{P}_{\upsilon'}$, these are independent.

Let us then explain why this is still pretty close to the definition given in [6]. That is, why it implies the natural geometric condition.

Note that $\delta^k = \delta^{(1-\gamma)s} \cdot \delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$ and $\delta^{(1-\gamma)s} \leq \delta^{(1-\gamma)r} < 10^{-5}$ (as r is fixed to be big enough). Suppose Q^k_α is geometrically $\mathcal{D}'$-good and $s \geq r$. We have that $x^k_\alpha \in R^{k-1}_\eta \subset R^{k-s}_\beta$ for some unique $(k-1,\eta) \leq_{\upsilon'} (k-s,\beta)$. Now $d(x^k_\alpha, y^{k-1}_\eta) < 10\delta^{k-1} = 10^4\delta^k < \delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$. Suppose (aiming for a contradiction) that we would have $d(x^k_\alpha, X \setminus R^{k-s}_\beta) < (3/4)\delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$. Then we would have for some $z \in X \setminus R^{k-s}_\beta$ that $d(x^k_\alpha, z) \leq (3/4)\delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$. But then $z \in R^{k-1}_\xi \subset R^{k-s}_\gamma$ for some $(k-1,\xi) \leq_{\upsilon'} (k-s,\gamma) \neq (k-s,\beta)$, and

$$d(x^k_\alpha, y^{k-1}_\xi) \leq d(x^k_\alpha, z) + d(z, y^{k-1}_\xi) \leq [3/4 + 10^{-1}]\delta^{\gamma k}\delta^{(1-\gamma)(k-s)} < \delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$$

contradicting the goodness of Q^k_α. So we must have

$$\begin{aligned} d(Q^k_\alpha, X \setminus R^{k-s}_\beta) &\geq d(x^k_\alpha, X \setminus R^{k-s}_\beta) - 10\delta^k \\ &\geq [3/4 - 10^{-4}]\delta^{\gamma k}\delta^{(1-\gamma)(k-s)} \geq 2^{-1}\delta^{\gamma k}\delta^{(1-\gamma)(k-s)}. \end{aligned}$$

Thus also $d(Q^k_\alpha, R^{k-s}_\gamma) \geq 2^{-1}\delta^{\gamma k}\delta^{(1-\gamma)(k-s)}$ for every $\gamma \neq \beta$. We record these easy observations as a lemma.

Lemma 4.6. *If $Q \in \mathcal{D}$ is geometrically $\mathcal{D}'$-good, then for every $R \in \mathcal{D}'$ for which $\ell(Q) \leq \delta^r \ell(R)$ we have either $d(Q,R) \gtrsim \ell(Q)^\gamma \ell(R)^{1-\gamma}$ or $d(Q, X \setminus R) \gtrsim \ell(Q)^\gamma \ell(R)^{1-\gamma}$.*

If Q^k_α is geometrically $\mathcal{D}'$-bad, then the definition demands that for some $s \geq r$ we have that $x^k_\alpha \in R^{k-s} \in \mathcal{D}'$ so that $d(x^k_\alpha, X \setminus R^{k-s}) \leq \delta^{\gamma k}\delta^{(1-\gamma)(k-s)} = \delta^{\gamma s}\delta^{k-s} = \delta^{\gamma s}\ell(R^{k-s})$. Lemma 4.3 with $\epsilon = \delta^{\gamma s}$ then yields that

$$\mathbb{P}(Q^k_\alpha \text{ is geometrically } \mathcal{D}'\text{-bad}) \lesssim \sum_{s=r}^{\infty} (\delta^{\gamma\eta})^s \lesssim \delta^{r\gamma\eta}.$$

We have proved the following:

Lemma 4.7. *For a fixed $Q \in \mathcal{D}$ we have under the random choice of the $\mathcal{D}'$-grid that*

$$\mathbb{P}(Q \textit{ is geometrically } \mathcal{D}'\textit{-bad}) \lesssim \delta^{r\gamma\eta}.$$

We still need to achieve the effect that $\mathbb{P}(Q \text{ is good})$ would not depend on the particular choice of the cube Q (in $\mathbb{R}^n$ this followed from symmetry, see [9]). That is, we want to make it a constant independent of (k,α) and υ. There seems to be no obvious reason why this should be the case already, so we will force this by understanding goodness in a stronger sense: a cube is good if it is geometrically good and pseudogood (an extra condition that scales the probability down).

Define $\pi_{k,\alpha}(\upsilon_k) = \mathbb{P}_{\upsilon'}(x^k_\alpha(\upsilon_k)$ is geometrically good) (recall that the geometric goodness is actually a property of the center and not of the cube). Set $\pi_{\text{good}} = 1 - C\delta^{r\gamma\eta}$ so that always $\pi_{k,\alpha}(\upsilon_k) \geq \pi_{\text{good}}$. For every (k,α) we take an independent uniformly on $[0,1]$ distributed random variable t^k_α. We say that $Q^k_\alpha(\upsilon)$ (or rather the triple (k,α,υ_k)) is pseudogood if $t^k_\alpha \in [0, \pi_{\text{good}}/\pi_{k,\alpha}(\upsilon_k)]$. Then we define that $Q^k_\alpha(\upsilon)$ is $\mathcal{D}'$-good if it is geometrically $\mathcal{D}'$-good and pseudogood. Consider the grid $\mathcal{D}$ fixed – that is, consider υ fixed. There holds $\mathbb{P}(Q^k_\alpha$ is $\mathcal{D}'$-good$)$ $=$ $\mathbb{P}_{\upsilon'}(Q^k_\alpha$ is geometrically $\mathcal{D}'$-good$)\mathbb{P}(t^k_\alpha \in [0, \pi_{\text{good}}/\pi_{k,\alpha}(\upsilon_k)]) = \pi_{\text{good}}$ (notice that with fixed υ the geometric goodness depends only on υ' and the pseudogoodness on t^k_α). We use independent uniformly on $[0,1]$ distributed random variables u^k_α to define pseudogoodness (and then $\mathcal{D}$-goodness) for pairs (k,α,υ'_k) (that is for the grid $\mathcal{D}'$).

Basically all these modification were done to prove the following analogue of Lemma 5.2 of [9] with our randomized systems of metric dyadic cubes. This enables us to later establish that a certain paraproduct is bounded following the strategy used in [9].

First a few comments. In the following section we shall introduce two fixed functions f and g, and their martingale difference decompositions using Haar functions. The aim is then to control a certain average (5.1). The details of this are not important for the next lemma, except for the fact that looking at that particular sum one sees that it is enough to sum over some fixed finite index set (k,α) (because the functions have bounded support, the space is geometrically doubling, and cubes of only finitely many generations are needed). Thus, we assume that such is the case in the next lemma also. This enables us to move $\mathbb{E}$ in and out the summation freely (see the proof). Also, $\varphi(Q,R)$ is an L^1-function of cubes Q and R and their children – basically in the only application of this lemma we take $\varphi(Q,R) = \langle g, \psi_R\rangle\langle b_2, T(b_1\varphi_Q)\rangle\langle\psi_R\rangle_Q\langle\varphi_Q, f\rangle$ (see Sections 5 and 8).

Lemma 4.8. *We have that*

$$(1 - C\delta^{r\gamma\eta})\,\mathbb{E}\sum_{R\in\mathcal{D}'}\sum_{\substack{Q\in\mathcal{D}_{\text{good}}\\ \delta^{k_0}<\ell(Q)\leq\ell(R)}}\varphi(Q,R) = \mathbb{E}\sum_{R\in\mathcal{D}'_{\text{good}}}\sum_{\substack{Q\in\mathcal{D}_{\text{good}}\\ \delta^{k_0}<\ell(Q)\leq\ell(R)}}\varphi(Q,R),$$

where the grid $\mathcal{D}'$ is fixed (so υ' is fixed) and we average over every other random quantity υ, t^k_α, u^k_α.

Proof. We start by recalling the dependencies (remember that the υ' is fixed). The goodness of $R^m_\gamma(\upsilon')$ (or $\chi_{\text{good}}(R^m_\gamma(\upsilon'))$) depends on u^m_γ and υ_ℓ for $\ell < m$. The goodness of $Q^k_\alpha(\upsilon)$ depends on υ_k and t^k_α. The $Q^k_\alpha(\upsilon)$ and its children as sets are determined by υ_ℓ, $\ell \geq k$ (this uses the fact that $k+1 \leq k_0$). Therefore, $\chi_{\text{good}}(Q^k_\alpha(\upsilon))\varphi(Q^k_\alpha(\upsilon), R^m_\gamma(\upsilon'))$ depends on υ_ℓ, $\ell \geq k$, and on and t^k_α. This means that $\chi_{\text{good}}(R^m_\gamma(\upsilon'))$ and $\chi_{\text{good}}(Q^k_\alpha(\upsilon))\varphi(Q^k_\alpha(\upsilon), R^m_\gamma(\upsilon'))$ are independent for $m \leq k < k_0$. Moreover, $1 - C\delta^{r\gamma\eta} = \pi_{\text{good}} = \mathbb{P}(R^m_\gamma(\upsilon') \in \mathcal{D}'_{\text{good}}) = \mathbb{E}(\chi_{\text{good}}(R^m_\gamma(\upsilon')))$.

Using this information we may now calculate

$$\begin{aligned}
&\pi_{\text{good}}\,\mathbb{E}\sum_{R\in\mathcal{D}'}\sum_{\substack{Q\in\mathcal{D}_{\text{good}}\\ \delta^{k_0}<\ell(Q)\le\ell(R)}}\varphi(Q,R)\\
&\quad=\pi_{\text{good}}\,\mathbb{E}\sum_{(m,\gamma)}\sum_{\substack{(k,\alpha)\\ m\le k<k_0}}\chi_{\text{good}}(Q^k_\alpha)\varphi(Q^k_\alpha,R^m_\gamma)\\
&\quad=\sum_{(m,\gamma)}\sum_{\substack{(k,\alpha)\\ m\le k<k_0}}\mathbb{E}(\chi_{\text{good}}(R^m_\gamma))\mathbb{E}(\chi_{\text{good}}(Q^k_\alpha)\varphi(Q^k_\alpha,R^m_\gamma))\\
&\quad=\sum_{(m,\gamma)}\sum_{\substack{(k,\alpha)\\ m\le k<k_0}}\mathbb{E}(\chi_{\text{good}}(R^m_\gamma)\chi_{\text{good}}(Q^k_\alpha)\varphi(Q^k_\alpha,R^m_\gamma))\\
&\quad=\mathbb{E}\sum_{R\in\mathcal{D}'_{\text{good}}}\sum_{\substack{Q\in\mathcal{D}_{\text{good}}\\ \delta^{k_0}<\ell(Q)\le\ell(R)}}\varphi(Q,R).
\end{aligned}$$

Let us still spell out the details of the above computation. We first removed everything that is random from the summations. Then we moved the expectation inside the summation (the sum is finite by assumption), and after that we also moved the constant $\pi_{\text{good}} = 1 - C\delta^{r\gamma\eta}$ inside the summation noting then that it equals $\mathbb{E}(\chi_{\text{good}}(R^m_\gamma))$ with any (m,γ). Next we used the product rule of expectations of independent quantities. Finally, we moved the expectation out and rewrote the summation so that it again contains the random quantities. □

5. Martingale difference decomposition, Haar functions and the tangent martingale trick

We remind the reader of the reductions done in Remark 2.2. That is, we are proving Theorem 2.1 under the *additional* assumptions that $\rho = d$ is a metric, $\|T\|_{L^p(X,Y)\to L^p(X,Y)} < \infty$ (Y is a fixed UMD-space), $|\langle T(\chi_A b_1), \chi_A b_2\rangle| \lesssim \mu(A)$ for any Borel set $A \subset X$, and that $Tb_1, T^*b_2 \in \text{BMO}^q_\kappa(\mu)$ with any $1 \le q < \infty$.

Let us be given some system of cubes $\{Q^k_\alpha\}$ and some accretive function b. We set (we use the notation $\langle f\rangle_A = \mu(A)^{-1}\int_A f\,d\mu$ for the average of a function f over a set A)

$$\begin{aligned}
E^b_k f &= \sum_\alpha \langle f\rangle_{Q^k_\alpha}\langle b\rangle^{-1}_{Q^k_\alpha}\chi_{Q^k_\alpha} b,\\
E^b_{Q^k_\alpha} f &= \chi_{Q^k_\alpha} E^b_k f,\\
\Delta^b_k f &= E^b_{k+1} f - E^b_k f,\\
\Delta^b_{Q^k_\alpha} f &= \chi_{Q^k_\alpha}\Delta^b_k f.
\end{aligned}$$

Consider a cube Q. It has subcubes of the next generation Q_i, $i = 1, \dots, s(Q)$, where $s(Q) \lesssim 1$. We set $\hat{Q}_k = \bigcup_{i=k}^{s(Q)} Q_i$, and note that we can always arrange the indexation of the subcubes to be such that $|b(\hat{Q}_k)| \gtrsim \mu(Q)$ for every $k = 1, \dots, s(Q)$. Indeed, we can index so that (here a is the accretivity constant of b)

$$|b(\hat{Q}_k)| \geq \Big(1 - \frac{k-1}{s(Q)}\Big)\, a\, \mu(Q) \gtrsim \mu(Q),$$

and this can proven as Lemma 4.3 in [9]. Note also that trivially $|b(\hat{Q}_k)| \lesssim \mu(Q)$ (so $|b(\hat{Q}_k)| \sim \mu(Q)$) and $|b(Q_i)| \sim \mu(Q_i)$.

Now define

$$\Delta^b_{Q,u} f = E^b_{Q_u} f + E^b_{\hat{Q}_{u+1}} f - E^b_{\hat{Q}_u} f$$

also noting that

$$\Delta^b_Q f = \sum_{u=1}^{s(Q)-1} \Delta^b_{Q,u} f.$$

A computation shows that

$$\Delta^b_{Q,u} f = b \varphi^b_{Q,u} \langle \varphi^b_{Q,u}, f \rangle,$$

where we have the adapted Haar functions

$$\varphi^b_{Q,u} = \Big(\frac{b(Q_u) b(\hat{Q}_{u+1})}{b(\hat{Q}_u)}\Big)^{1/2} \Big(\frac{\chi_{Q_u}}{b(Q_u)} - \frac{\chi_{\hat{Q}_{u+1}}}{b(\hat{Q}_{u+1})}\Big)$$

as in [9]. Here we have to interpret $\varphi^b_{Q,u} = 0$ if $\mu(Q_u) = 0$. We also have the non-cancellative (does not, in general, have zero integral) adapted Haar function

$$\varphi^b_{Q,0} f = b(Q)^{-1/2} \chi_Q$$

using which we write $E^b_Q f = b\, \varphi^b_{Q,0}\, \langle \varphi^b_{Q,0}, f \rangle$.

We record the key properties (the last two being only important special cases):

$$\int b\, \varphi^b_{Q,u}\, d\mu = 0,$$

$$|\varphi^b_{Q,u}| \sim \mu(Q_u)^{1/2} \Big(\frac{\chi_{Q_u}}{\mu(Q_u)} + \frac{\chi_{\hat{Q}_{u+1}}}{\mu(Q)}\Big),$$

$$\|\varphi^b_{Q,u}\|_{L^p(X)} \sim \mu(Q_u)^{1/p - 1/2}$$

and

$$\|\varphi^b_{Q,u}\|_{L^1(X)}\, \|\varphi^b_{Q,u}\|_{L^\infty(X)} \sim 1.$$

Given a dyadic system $\mathcal{D} = \{Q\}$ we can write with any m that

$$f = \sum_{\substack{Q \in \mathcal{D} \\ \ell(Q) \leq \delta^m}} \Delta^{b_1}_Q f + \sum_{\substack{Q \in \mathcal{D} \\ \ell(Q) = \delta^m}} E^{b_1}_Q f = \sum_{\substack{Q \in \mathcal{D} \\ \ell(Q) \leq \delta^m}} \sum_u b_1 \varphi^{b_1}_{Q,u} \langle \varphi^{b_1}_{Q,u}, f \rangle,$$

where the u summation runs through $1, \dots, s(Q) - 1$ if $\ell(Q) < \delta^m$, and through $0, 1, \dots, s(Q) - 1$ if $\ell(Q) = \delta^m$. The unconditional convergence of this in $L^p(X, Y)$ is not at all clear, but it nevertheless follows as in Proposition 4.1 of [9] (note that in that proof certain abstract paraproducts are used, but their theory is formulated in Chapter 3 of [9] in an abstract filtered space which directly applies also in our situation).

Basically the strategy we shall use is the usual one: write the same decomposition for a function $g \in L^{p'}(X, Y^*)$ just using some other grid $\mathcal{D}' = \{R\}$ and the other test function b_2, and then decompose the pairing $\langle g, Tf \rangle$ accordingly. However, Lemma 4.8 has the restriction involving k_0 (which we have not yet fixed) and so we somehow need to get into a situation where we do not need to consider arbitrarily small cubes.

We start by choosing two boundedly supported functions $f \in L^p(X, Y)$ and $g \in L^{p'}(X, Y^*)$ so that f/b_1 and g/b_2 are Lipschitz, $\|f\|_{L^p(X,Y)} = \|g\|_{L^{p'}(X,Y^*)} = 1$ and $\|T\| \le 2|\langle g, Tf\rangle|$. Here, of course, $\|T\| = \|T\|_{L^p(X,Y)\to L^p(X,Y)}$. For the fact that Lipschitz functions are dense, see e.g. the proof of Proposition 3.4 of [8]. We now also fix m so that the supports of the functions f and g are contained in some balls $B(x_0, \delta^m)$ and $B(x_1, \delta^m)$ respectively. The k_0 can now be fixed to be so big that δ^{k_0} is small enough for the estimates that follow.

Using any two dyadic systems $\mathcal{D}$ and $\mathcal{D}'$ we decompose

$$\langle g, Tf\rangle = \langle g - E_{k_0}^{b_2} g, Tf\rangle + \langle E_{k_0}^{b_2} g, T(f - E_{k_0}^{b_1} f)\rangle + \langle E_{k_0}^{b_2} g, T(E_{k_0}^{b_1} f)\rangle,$$

and then estimate

$$\begin{aligned}|\langle g, Tf\rangle| &\le \|T\|\|g - E_{k_0}^{b_2} g\|_{L^{p'}(X,Y^*)}\|f\|_{L^p(X,Y)} \\ &\quad + \|T\|\|E_{k_0}^{b_2} g\|_{L^{p'}(X,Y^*)}\|f - E_{k_0}^{b_1} f\|_{L^p(X,Y)} + |\langle E_{k_0}^{b_2} g, T(E_{k_0}^{b_1} f)\rangle|.\end{aligned}$$

Note that $\|E_{k_0}^{b_2} g\|_{L^{p'}(X,Y^*)} \lesssim \|g\|_{L^{p'}(X,Y^*)} = 1$ so that we get

$$|\langle g, Tf\rangle| \le (C(b_2)\|f - E_{k_0}^{b_1} f\|_{L^p(X,Y)} + \|g - E_{k_0}^{b_2} g\|_{L^{p'}(X,Y^*)})\|T\| + |\langle E_{k_0}^{b_2} g, T(E_{k_0}^{b_1} f)\rangle|.$$

Next we employ the facts that f/b_1 and g/b_2 are Lipschitz (with a constant L, say). Let $h = f/b_1$. Let $x \in X$ and then let Q denote the unique $\mathcal{D}$-cube of generation k_0 containing x. We have that

$$\begin{aligned}\|E_{k_0}^{b_1} f(x) - f(x)\|_Y &\lesssim \|\langle b_1\rangle_Q h(x) - \langle b_1 h\rangle_Q\|_Y \\ &\le \frac{1}{\mu(Q)}\int_Q |b_1(z)|\,\|h(z) - h(x)\|_Y\, d\mu(z) \\ &\lesssim L d(Q) \lesssim L\delta^{k_0}.\end{aligned}$$

Noting that $\bigcup\{Q : Q \in \mathcal{D}_{k_0},\, Q \cap B(x_0, \delta^m) \neq \emptyset\} \subset B(x_0, 2\delta^m)$ we have that

$$\|f - E_{k_0}^{b_1} f\|_{L^p(X,Y)} \lesssim L\lambda(x_0, \delta^m)^{1/p}\delta^{k_0}.$$

A similar estimate holds for $\|g - E_{k_0}^{b_2} g\|_{L^{p'}(X,Y^*)}$. The k_0 is so large that we have

$$\|T\|/2 \le |\langle g, Tf\rangle| \le \|T\|/4 + |\langle E_{k_0}^{b_2} g, T(E_{k_0}^{b_1} f)\rangle|,$$

that is, $\|T\| \le 4|\langle E_{k_0}^{b_2} g, T(E_{k_0}^{b_1} f)\rangle|$ with any grids $\mathcal{D}$ and $\mathcal{D}'$ (but only with these particular fixed functions f and g, of course).

Now we write $\langle E_{k_0}^{b_2} g, T(E_{k_0}^{b_1} f)\rangle$ as the following sum:

$$\begin{aligned}
&\Big\langle \sum_{\substack{R\in\mathcal{D}'_{\rm bad}\\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g + \sum_{\substack{R\in\mathcal{D}'_{\rm bad}\\ \ell(R)=\delta^m}} E_R^{b_2} g, T(E_{k_0}^{b_1} f)\Big\rangle \\
&+\Big\langle \sum_{\substack{R\in\mathcal{D}'_{\rm good}\\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g + \sum_{\substack{R\in\mathcal{D}'_{\rm good}\\ \ell(R)=\delta^m}} E_R^{b_2} g, T\Big(\sum_{\substack{Q\in\mathcal{D}_{\rm bad}\\ \delta^{k_0}<\ell(Q)\le\delta^m}} \Delta_Q^{b_1} f + \sum_{\substack{Q\in\mathcal{D}_{\rm bad}\\ \ell(Q)=\delta^m}} E_Q^{b_1} f\Big)\Big\rangle \\
&+ \sum_{\substack{Q\in\mathcal{D}_{\rm good},\, R\in\mathcal{D}'_{\rm good}\\ \delta^{k_0}<\ell(Q),\, \ell(R)\le\delta^m}} \sum_{u,v} \langle \varphi_{R,v}^{b_2}, g\rangle \langle b_2\varphi_{R,v}^{b_2}, T(b_1\varphi_{Q,u}^{b_1})\rangle\langle \varphi_{Q,u}^{b_1}, f\rangle,
\end{aligned}$$

where the u summation runs through $1,\dots,s(Q)-1$ if $\ell(Q)<\delta^m$, and through $0,1,\dots,s(Q)-1$ if $\ell(Q)=\delta^m$, and similarly for the v summation. We thus have that $\|T\|/4$ is bounded by the sum of the following terms:

$$\|T\|\Big\| \sum_{\substack{R\in\mathcal{D}'_{\rm bad}\\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g + \sum_{\substack{R\in\mathcal{D}'_{\rm bad}\\ \ell(R)=\delta^m}} E_R^{b_2} g\Big\|_{L^{p'}(X,Y^*)} \|E_{k_0}^{b_1} f\|_{L^p(X,Y)},$$

$$\|T\|\Big\| \sum_{\substack{R\in\mathcal{D}'_{\rm good}\\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g + \sum_{\substack{R\in\mathcal{D}'_{\rm good}\\ \ell(R)=\delta^m}} E_R^{b_2} g\Big\|_{L^{p'}(X,Y^*)} \Big\| \sum_{\substack{Q\in\mathcal{D}_{\rm bad}\\ \delta^{k_0}<\ell(Q)\le\delta^m}} \Delta_Q^{b_1} f + \sum_{\substack{Q\in\mathcal{D}_{\rm bad}\\ \ell(Q)=\delta^m}} E_Q^{b_1} f\Big\|_{L^p(X,Y)}$$

and

$$\Big| \sum_{\substack{Q\in\mathcal{D}_{\rm good},\, R\in\mathcal{D}'_{\rm good}\\ \delta^{k_0}<\ell(Q),\, \ell(R)\le\delta^m}} \sum_{u,v} \langle \varphi_{R,v}^{b_2}, g\rangle \langle b_2\varphi_{R,v}^{b_2}, T(b_1\varphi_{Q,u}^{b_1})\rangle\langle \varphi_{Q,u}^{b_1}, f\rangle\Big|.$$

Note that clearly

$$\|E_{k_0}^{b_1} f\|_{L^p(X,Y)} \lesssim \|f\|_{L^p(X,Y)} = 1$$

and

$$\Big\| \sum_{\substack{R\in\mathcal{D}'_{\rm good}\\ \ell(R)=\delta^m}} E_R^{b_2} g\Big\|_{L^{p'}(X,Y^*)} \lesssim \|g\|_{L^{p'}(X,Y^*)} = 1.$$

Also, using unconditionality and the contraction principle, we have that

$$\Big\| \sum_{\substack{R\in\mathcal{D}'_{\rm good}\\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g\Big\|_{L^{p'}(X,Y^*)} \lesssim \|g\|_{L^{p'}(X,Y^*)} = 1.$$

Thus, the terms involving bad cubes are dominated by

$$\|T\| \Bigg[\Big\| \sum_{\substack{R\in\mathcal{D}'_{\text{bad}} \\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g + \sum_{\substack{R\in\mathcal{D}'_{\text{bad}} \\ \ell(R)=\delta^m}} E_R^{b_2} g \Big\|_{L^{p'}(X,Y^*)} + \Big\| \sum_{\substack{Q\in\mathcal{D}_{\text{bad}} \\ \delta^{k_0}<\ell(Q)\le\delta^m}} \Delta_Q^{b_1} f + \sum_{\substack{Q\in\mathcal{D}_{\text{bad}} \\ \ell(Q)=\delta^m}} E_Q^{b_1} f \Big\|_{L^p(X,Y)} \Bigg].$$

Taking expectations over all the random quantities in the randomization of cubes, it is easy to see that

$$\mathbb{E}\Big\| \sum_{\substack{R\in\mathcal{D}'_{\text{bad}} \\ \ell(R)=\delta^m}} E_R^{b_2} g \Big\|_{L^{p'}(X,Y^*)} + \mathbb{E}\Big\| \sum_{\substack{Q\in\mathcal{D}_{\text{bad}} \\ \ell(Q)=\delta^m}} E_Q^{b_1} f \Big\|_{L^p(X,Y)} \lesssim \eta(r),$$

where $\eta(r) \to 0$ when $r \to \infty$. Working similarly as later in Section 9 (when estimating a certain term B) we have that

$$\mathbb{E}\Big\| \sum_{\substack{R\in\mathcal{D}'_{\text{bad}} \\ \delta^{k_0}<\ell(R)\le\delta^m}} \Delta_R^{b_2} g \Big\|_{L^{p'}(X,Y^*)} + \mathbb{E}\Big\| \sum_{\substack{Q\in\mathcal{D}_{\text{bad}} \\ \delta^{k_0}<\ell(Q)\le\delta^m}} \Delta_Q^{b_1} f \Big\|_{L^p(X,Y)} \lesssim \eta(r)$$

as well. One can consult Chapter 12 of [9] too. The proof requires a certain improvement of the contraction principle recalled in Proposition 9.1 (this is Lemma 3.1 of [7]).

Choosing r large enough we thus have that

$$\|T\|/8 \le \mathbb{E}\Big| \sum_{\substack{Q\in\mathcal{D}_{\text{good}},\, R\in\mathcal{D}'_{\text{good}} \\ \delta^{k_0}<\ell(Q),\, \ell(R)\le\delta^m}} \sum_{u,v} \langle \varphi_{R,v}^{b_2}, g\rangle \langle b_2\varphi_{R,v}^{b_2}, T(b_1\varphi_{Q,u}^{b_1})\rangle \langle \varphi_{Q,u}^{b_1}, f\rangle \Big|. \tag{5.1}$$

We almost always suppress the finite summation over u, v and after that is done, simply write $\varphi_Q = \varphi_{Q,u}^{b_1}$, $\psi_R = \varphi_{R,v}^{b_2}$ and $T_{RQ} = \langle b_2\psi_R, T(b_1\varphi_Q)\rangle$. The summation condition $\delta^{k_0} < \ell(Q)$, $\ell(R) \le \delta^m$ is always in force, and thus most of the time not explicitly written. The estimation of this series involving good cubes only is now split into multiple subseries to be considered in the subsequent sections. We primarily deal with the part $\ell(Q) \le \ell(R)$ the other being symmetric. Although we have $\|f\|_{L^p(X,Y)} = \|g\|_{L^{p'}(X,Y^*)} = 1$, in some of the estimates below we explicitly write $\|f\|_{L^p(X,Y)}$ and $\|g\|_{L^{p'}(X,Y^*)}$ in place of 1 for clarity.

We still comment on some of the techniques used on the following sections. We use independent random signs ϵ_k with $\mathbb{P}(\epsilon_k = 1) = \mathbb{P}(\epsilon_k = -1) = 1/2$. The underlying probability space for these signs is denoted by Ω. They are often indexed by cubes rather than the size of cubes in situations, where there is for some reason at most one non-zero term for every generation of cubes.

Related to this vector-valued L^p-theory we combine basic randomization tricks with the more sophisticated tool called the tangent martingale trick in [9]. Let us now formulate this since it is of fundamental importance to us (this is Corollary 6.3 of [9]).

Proposition 5.1. *Let $\mathcal{A} = \bigcup_k \mathcal{A}_k$, where $\mathcal{A}_k$ is a countable partition of X into Borel sets of finite μ-measure, and the generated σ-algebras $\sigma(\mathcal{A}_k)$ satisfy $\sigma(\mathcal{A}_k) \subset \sigma(\mathcal{A}_{k+1})$. For each $A \in \mathcal{A}$ we are given a* UMD*-valued function $f_A\colon X \to Y$ supported on A, and so that f_A is $\sigma(\mathcal{A}_{k+1})$-measurable whenever $A \in \mathcal{A}_k$. For each $A \in \mathcal{A}$ we are also given a jointly measurable function $k_A\colon A \times A \to \mathbb{C}$, which is pointwise bounded by* 1*. We have*

$$\int_{\Omega\times X} \Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{A\in\mathcal{A}_k} \frac{\chi_A(x)}{\mu(A)} \int_A k_A(x,z) f_A(z)\,d\mu(z)\Big\|_Y^p \,d\mathbb{P}(\epsilon)\,d\mu(x)$$
$$\lesssim \int_{\Omega\times X} \Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{A\in\mathcal{A}_k} f_A(x)\Big\|_Y^p \,d\mathbb{P}(\epsilon)\,d\mu(x).$$

This is the only version of the trick we explicitly need in this paper. For this result and some more general theory related to this see Chapter 6 of [9]. Lastly, we record the following randomization trick which is used multiple times in the sequel. For the proof see the page 10 of [9].

Lemma 5.2. *Suppose that for each $R \in \mathcal{D}'$ we are given a subcollection $\mathcal{D}(R) \subset \mathcal{D}$. There holds*

$$\Big| \sum_{R\in\mathcal{D}'} \langle g, \psi_R\rangle \sum_{Q\in\mathcal{D}(R)} T_{RQ}\langle \varphi_Q, f\rangle \Big|$$
$$\lesssim \|g\|_{L^{p'}(X,Y^*)} \Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{R\in\mathcal{D}'_k} \psi_R \sum_{Q\in\mathcal{D}(R)} T_{RQ}\langle\varphi_Q, f\rangle \Big\|_{L^p(\Omega\times X,Y)}.$$

6. Separated cubes

We consider the part of the series where $R \in \mathcal{D}'_{\text{good}}$, $Q \in \mathcal{D}_{\text{good}}$, $\ell(Q) \le \ell(R)$ and $d(Q,R) \ge CC_0\ell(Q)$. Also the adapted Haar functions φ_Q related to the smaller cubes Q are assumed to be cancellative (by which we only mean that they have zero integral).

We begin with some estimates for the matrix elements $T_{RQ} = \langle b_2\psi_R, T(b_1\varphi_Q)\rangle$ – these follow, with some modifications, Lemma 6.1 and Lemma 6.2 of [6].

Lemma 6.1. *Let $Q \in \mathcal{D}$ and $R \in \mathcal{D}'$ be such that $\ell(Q) \le \ell(R)$ and $d(Q,R) \ge CC_0\ell(Q)$. Assume also that φ_Q is cancellative. We have the estimate*

$$|T_{RQ}| \lesssim \frac{\ell(Q)^\alpha}{d(Q,R)^\alpha \sup_{z\in Q}\lambda(z, d(Q,R))} \|\varphi_Q\|_{L^1(\mu)} \|\psi_R\|_{L^1(\mu)}.$$

Proof. Recalling that $\int b_1\varphi_Q\,d\mu = 0$, we have for an arbitrary $z \in Q$ that

$$T_{RQ} = \int_R \int_Q [K(x,y) - K(x,z)]\, b_1(y)\, \varphi_Q(y)\, b_2(x)\, \psi_R(x)\; d\mu(y)\, d\mu(x).$$

The claim follows from the kernel estimates (which we may utilize since $d(x,z) \geq d(Q,R) \geq CC_0\ell(Q) \geq Cd(y,z)$). □

We set $D(Q,R) = \ell(Q) + \ell(R) + d(Q,R)$.

Lemma 6.2. *Let $Q \in \mathcal{D}_{\text{good}}$ and $R \in \mathcal{D}'$ be such that $\ell(Q) \leq \ell(R)$ and $d(Q,R) \geq CC_0\ell(Q)$. Assume also that φ_Q is cancellative. We have the estimate*

$$|T_{RQ}| \lesssim \frac{\ell(Q)^{\alpha/2}\ell(R)^{\alpha/2}}{D(Q,R)^{\alpha}\sup_{z\in Q}\lambda(z, D(Q,R))}\|\varphi_Q\|_{L^1(\mu)}\|\psi_R\|_{L^1(\mu)}.$$

Proof. If $\ell(Q) > \delta^r\ell(R)$, then $d(Q,R) \gtrsim D(Q,R)$, and the claim follows from the previous lemma. In the case $d(Q,R) \geq \ell(R)$, we also have $d(Q,R) \gtrsim D(Q,R)$, and the claim again follows from the previous lemma.

We may thus assume that $\ell(Q) \leq \delta^r\ell(R)$ and $d(Q,R) \leq \ell(R)$. As Q is good, we have $d(Q,R) \gtrsim \ell(Q)^{\gamma}\ell(R)^{1-\gamma}$. Consider an arbitrary $z \in Q$. Using the identity

$$C_\lambda^{-\gamma\log_2\frac{\ell(R)}{\ell(Q)}} = \Big(\frac{\ell(R)}{\ell(Q)}\Big)^{-\gamma d}$$

and the doubling property of λ one gets that

$$\lambda(z, d(Q,R)) \gtrsim \Big(\frac{\ell(R)}{\ell(Q)}\Big)^{-\gamma d}\lambda(z,\ell(R)).$$

The claim then follows from the previous lemma, the identity $\gamma d + \gamma\alpha = \alpha/2$, and the fact that in our situation $\ell(R) \gtrsim D(Q,R)$. □

Let us then state and prove the main result of this section – this follows, save the technical modifications, from pages 25–26 of [9].

Proposition 6.3. *There holds*

$$\Bigg|\sum_{R\in\mathcal{D}'_{\text{good}}}\ \sum_{\substack{Q\in\mathcal{D}_{\text{good}}\\ \ell(Q)\leq\ell(R),\, d(Q,R)\geq CC_0\ell(Q)}} \langle g,\psi_R\rangle\, T_{RQ}\langle\varphi_Q, f\rangle\Bigg| \lesssim \|g\|_{L^{p'}(X,Y^*)}\|f\|_{L^p(X,Y)}$$

with the additional interpretation that the adapted Haar functions φ_Q related to the smaller cubes Q are cancellative, even on the coarsest level $\ell(Q) = \delta^m$.

Proof. We first consider the case

$$\begin{cases} \ell(R) = \delta^k, & k \in \mathbb{Z},\\ \ell(Q) = \delta^{k+m}, & m = 0,1,2,\dots,\\ \delta^{k-j} < D(Q,R) \leq \delta^{k-j-1}, & j = 0,1,2,\dots. \end{cases}$$

The last requirement says that $D(Q,R)/\ell(R) \sim \delta^{-j}$. The estimate from the previous lemma gives

$$\frac{|T_{RQ}|}{\|\varphi_Q\|_{L^1(\mu)}\|\psi_R\|_{L^1(\mu)}} \lesssim \frac{\delta^{\alpha m/2}\delta^{\alpha j}}{\sup_{z\in Q}\lambda(z,\delta^{k-j})}.$$

We suppress from our notation the requirement that $d(Q,R) \ge CC_0\ell(Q)$. Lemma 5.2 gives that

$$\Big|\sum_{k\in\mathbb{Z}}\sum_{R\in\mathcal{D}'_{\mathrm{good},k}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ D(Q,R)/\ell(R)\sim\delta^{-j}}}\langle g,\psi_R\rangle\, T_{RQ}\langle\varphi_Q,f\rangle\Big|$$
$$\lesssim \|g\|_{L^{p'}(X,Y^*)}\Big\|\sum_{k\in\mathbb{Z}}\epsilon_k\sum_{R\in\mathcal{D}'_{\mathrm{good},k}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ D(Q,R)/\ell(R)\sim\delta^{-j}}}\psi_R T_{RQ}\langle\varphi_Q,f\rangle\Big\|_{L^p(\Omega\times X,Y)}.$$

For a cube Q denote by $\tilde{Q}_\ell$ the unique cube of generation $\ell \le \mathrm{gen}(Q)$ for which $Q \subset \tilde{Q}_\ell$. Let $\theta(j)$ denote the smallest integer for which $\theta(j) \ge (j\gamma + r)(1-\gamma)^{-1}$. Recalling that R is good and r is large enough, we must have for any Q and R in the above summation that $R \subset \tilde{Q}_{k-j-\theta(j)}$. Thus, we may write

$$\sum_{R\in\mathcal{D}'_{\mathrm{good},k}} = \sum_{S\in\mathcal{D}_{k-j-\theta(j)}}\sum_{\substack{R\in\mathcal{D}'_{\mathrm{good},k}\\ R\subset S}}.$$

Also, we have

$$\mu(S) \lesssim \inf_{w\in S}\lambda(w,\delta^{k-j-\theta(j)}) \lesssim \delta^{-d\theta(j)}\inf_{w\in S}\lambda(w,\delta^{k-j}).$$

Define t_{RQ} via the identity

$$T_{RQ} = \frac{\delta^{\alpha m/2}\delta^{\alpha j-d\theta(j)}}{\mu(S)}\|\varphi_Q\|_{L^1(\mu)}\|\psi_R\|_{L^1(\mu)}t_{RQ},$$

and note that we have

$$|t_{RQ}| \lesssim \frac{\inf_{w\in S}\lambda(w,\delta^{k-j})}{\sup_{z\in Q}\lambda(z,\delta^{k-j})} \le 1.$$

Also relevant is the estimate

$$\delta^{\alpha j-d\theta(j)} \lesssim \delta^{[\alpha-d\gamma(1-\gamma)^{-1}]j} = \delta^{(\alpha^2+\alpha d)(\alpha+2d)^{-1}j}.$$

For every $S \in \mathcal{D}_{k-j-\theta(j)}$ we set

$$K_S(x,y) = \sum_{\substack{R\in\mathcal{D}'_{\mathrm{good},k}\\ R\subset S}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ D(Q,R)/\ell(R)\sim\delta^{-j}}}\psi_R(x)\|\psi_R\|_{L^1(\mu)}t_{RQ}\|\varphi_Q\|_{L^1(\mu)}\varphi_Q(y)b_1(y).$$

As $\|\varphi_Q\|_{L^1(\mu)}\|\varphi_Q\|_{L^\infty(\mu)} \lesssim 1$, $\|\psi_R\|_{L^1(\mu)}\|\psi_R\|_{L^\infty(\mu)} \lesssim 1$, $\|b_1\|_{L^\infty(\mu)} \lesssim 1$, $|t_{RQ}| \lesssim 1$ and for every fixed x and y there is at most one non-zero term in the double sum defining K_S, we have $|K_S(x,y)| \lesssim 1$. Also, K_S is supported on $S \times S$ as $\mathrm{spt}\,\psi_R \subset R \subset S$ and $\mathrm{spt}\,\varphi_Q \subset Q \subset S$.

Using the fact that $\int b_1\varphi_Q\,d\mu = 0$ one notes that $\langle \varphi_Q, f\rangle = \langle \varphi_Q, \Delta^{b_1}_{k+m} f\rangle$ for $Q \in \mathcal{D}_{k+m}$. Using this and the definitions from above, we see that

$$\Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{R\in\mathcal{D}'_{\mathrm{good},k}} \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ D(Q,R)/\ell(R)\sim\delta^{-j}}} \psi_R T_{RQ}\langle\varphi_Q, f\rangle \Big\|_{L^p(\Omega\times X,Y)}$$

can be dominated by

$$\delta^{\frac{\alpha}{2}m}\,\delta^{\frac{\alpha^2+\alpha d}{\alpha+2d}j} \Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{S\in\mathcal{D}_{k-j-\theta(j)}} \frac{\chi_S}{\mu(S)} \int_S K_S(\cdot,y) \frac{\chi_S(y)\Delta^{b_1}_{k+m} f(y)}{b_1(y)}\,d\mu(y) \Big\|_{L^p(\Omega\times X,Y)}.$$

Due to the measurability requirements of the tangent martingale trick we further split up the above sum over $k\in\mathbb{Z}$ into $m+j+\theta(j)+1 \lesssim m+j+1$ subseries:

$$\sum_{k\in\mathbb{Z}} = \sum_{k_0=0}^{m+j+\theta(j)} \sum_{\substack{k\equiv k_0\\ \mathrm{mod}\, m+j+\theta(j)+1}}.$$

The point is that $y \mapsto \frac{\chi_S(y)\Delta^{b_1}_{k+m} f(y)}{b_1(y)}$ is constant on the subcubes of generation $k+m+1 = k'-j-\theta(j)$, where $k' = k+(m+j+\theta(j)+1)$. Applying the tangent martingale trick to each of these subseries then yields that

$$\begin{aligned}
&\Big| \sum_{k\in\mathbb{Z}} \sum_{R\in\mathcal{D}'_{\mathrm{good},k}} \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ D(Q,R)/\ell(R)\sim\delta^{-j}}} \langle g,\psi_R\rangle T_{RQ}\langle\varphi_Q, f\rangle \Big| \\
&\lesssim \delta^{\frac{\alpha}{2}m}\delta^{\frac{\alpha^2+\alpha d}{\alpha+2d}j}\|g\|_{L^{p'}(X,Y^*)} \sum_{k_0=0}^{m+j+\theta(j)} \Big\| \sum_{k\equiv k_0} \epsilon_k \sum_{S\in\mathcal{D}_{k-j-\theta(j)}} \frac{\chi_S\Delta^{b_1}_{k+m} f}{b_1} \Big\|_{L^p(\Omega\times X,Y)} \\
&\lesssim \delta^{\frac{\alpha}{2}m}\delta^{\frac{\alpha^2+\alpha d}{\alpha+2d}j}(m+j+1)\|g\|_{L^{p'}(X,Y^*)}\|f\|_{L^p(X,Y)},
\end{aligned}$$

where the last inequality follows from the unconditional convergence of the adapted martingale difference decomposition (after discarding $1/b_1$). Summing over $m, j = 0, 1, 2, \ldots$ yields the claim. □

7. Cubes well inside another cube

We consider the case $R \in \mathcal{D}'_{\mathrm{good}}$, $Q \in \mathcal{D}_{\mathrm{good}}$, $Q \subset R$ and $\ell(Q) < \delta^r \ell(R)$. As usual, there is a need to introduce some cancellation. To this end, here we consider the modified matrix

$$\begin{aligned}
\tilde{T}_{RQ} &= T_{RQ} - \langle b_2, T(b_1\varphi_Q)\rangle\langle\psi_R\rangle_Q \\
&= -\langle \chi_{X\setminus S} b_2, T(b_1\varphi_Q)\rangle\langle\psi_R\rangle_S + \sum_{\substack{S'\subset R\setminus S\\ \ell(S')=\delta\ell(R)}} \langle \chi_{S'}\psi_R b_2, T(b_1\varphi_Q)\rangle,
\end{aligned}$$

where $S \subset R$ is such that $\ell(S) = \delta\ell(R)$ (that is, S is a child of R, and $\mathrm{gen}(S) = \mathrm{gen}(R) + 1$) and $Q \subset S$. The point is that Q is separated from the rest of the subcubes S' and we have introduced cancellation for this one problematic subcube S. The correction terms form a paraproduct operator, the boundedness of which will be considered in the next section.

We again begin with some estimates for the matrix $\tilde{T}_{RQ}$. Let us be brief as these estimates follow pretty much as in the pages 20–21 of [6]. Fix some $z \in Q$. Recalling that for every ball $B = B(c_B, r_B)$ and for every $\epsilon > 0$ we have the estimate (integrate over dyadic blocks $2^j r_B \le d(x, c_B) < 2^{j+1} r_B$ or see Lemma 2.4 in [6])

$$\int_{X\setminus B} \frac{d(x,c_B)^{-\epsilon}}{\lambda(c_B, d(x,c_B))}\,d\mu(x) \lesssim_\epsilon r_B^{-\epsilon},$$

we establish by changing $K(x,y)$ to $K(x,y) - K(x,z)$ (using $\int b_1\varphi_Q\,d\mu = 0$), using the kernel estimates and noting that $X \setminus S \subset X \setminus B(z, d(Q, X\setminus S))$ that

$$|\langle \chi_{X\setminus S} b_2, T(b_1\varphi_Q)\rangle| \lesssim \ell(Q)^\alpha \|\varphi_Q\|_{L^1(\mu)} d(Q, X\setminus S)^{-\alpha}.$$

To see that it was legitimate to use the kernel estimates note that in the corresponding integral $d(x,z) \ge d(X\setminus S, Q) \gtrsim \ell(Q)^\gamma \ell(S)^{1-\gamma} \ge \delta^{-r(1-\gamma)}\ell(Q)$, so that $d(x,z) \ge C d(y,z)$ choosing r large enough. Furthermore, note that $d(Q, X\setminus S) \gtrsim \ell(Q)^\gamma\ell(S)^{1-\gamma} \ge \ell(Q)^{1/2}\ell(R)^{1/2}$, and so continuing the above estimates we obtain

$$|\langle \chi_{X\setminus S} b_2, T(b_1\varphi_Q)\rangle| \lesssim \Big(\frac{\ell(Q)}{\ell(R)}\Big)^{\alpha/2} \|\varphi_Q\|_{L^1(\mu)}.$$

For the other finitely many terms involving a subcube $S' \subset R$ (where we have separation) we have using Lemma 6.2 (or actually, a trivial modification) that

$$\begin{aligned}|\langle \chi_{S'}\psi_R b_2, T(b_1\varphi_Q)\rangle| &\lesssim \Big(\frac{\ell(Q)}{\ell(S')}\Big)^{\alpha/2} \frac{\|\psi_R\|_{L^1(\mu)}}{\lambda(z,\ell(S'))}\|\varphi_Q\|_{L^1(\mu)} \\ &\lesssim \Big(\frac{\ell(Q)}{\ell(R)}\Big)^{\alpha/2} \frac{\|\psi_R\|_{L^1(\mu)}}{\mu(R)}\|\varphi_Q\|_{L^1(\mu)},\end{aligned}$$

where the last estimate follows after noting that

$$\mu(R) \le \mu(B(z, C_0\ell(R))) \le \lambda(z, C_0\ell(R)) = \lambda(z, C_0\delta^{-1}\ell(S')) \lesssim \lambda(z, \ell(S')).$$

Let us recapitulate all this as a lemma.

Lemma 7.1. *If $R \in \mathcal{D}'$, $Q \in \mathcal{D}_{\mathrm{good}}$, $Q \subset R$, $\ell(Q) < \delta^r\ell(R)$ and S is the subcube of R for which $\ell(S) = \delta\ell(R)$ and $Q \subset S$, we have*

$$|\tilde{T}_{RQ}| \lesssim \Big(\frac{\ell(Q)}{\ell(R)}\Big)^{\alpha/2}\Big[|\langle\psi_R\rangle_S| + \frac{\|\psi_R\|_{L^1(\mu)}}{\mu(R)}\Big]\|\varphi_Q\|_{L^1(\mu)}.$$

A familiar strategy involving kernels and the tangent martingale trick shall now be employed (as in the previous section and as in [9]). For this, the following lemma is both natural and useful.

Lemma 7.2. *If $R \in \mathcal{D}'$, $Q \in \mathcal{D}_{\mathrm{good}}$, $Q \subset R$, $\ell(Q) < \delta^r \ell(R)$ and S is the subcube of R for which $\ell(S) = \delta\ell(R)$ and $Q \subset S$, we have*

$$|\psi_R(x)\tilde{T}_{RQ}\varphi_Q(y)| \lesssim \Big(\frac{\ell(Q)}{\ell(R)}\Big)^{\alpha/2}\Big[\frac{\chi_{R\setminus S}(x)}{\mu(R)} + \frac{\chi_S(x)}{\mu(S)}\Big].$$

Proof. Taking the previous lemma and the estimates $\|\varphi_Q\|_{L^1(\mu)}\|\varphi_Q\|_{L^\infty(\mu)} \lesssim 1$ and $\|\psi_R\|_{L^1(\mu)}\|\psi_R\|_{L^\infty(\mu)} \lesssim 1$ into account it suffices to prove that

$$|\langle\psi_R\rangle_S||\psi_R(x)| \lesssim \frac{\chi_{R\setminus S}(x)}{\mu(R)} + \frac{\chi_S(x)}{\mu(S)}.$$

This follows by recalling that $\psi_R = \varphi^{b_2}_{R,v}$ for some v, denoting $S = R_w$, subdividing the estimation into cases ($v = w$ and $x \in S$), ($v = w$ and $x \in R \setminus S$) and $v \neq w$, and finally recalling that one has

$$|\psi_R| \sim \mu(R_v)^{1/2}\Big(\frac{\chi_{R_v}}{\mu(R_v)} + \frac{\chi_{\hat{R}_{v+1}}}{\mu(R)}\Big)$$

(or $|\psi_R| \sim \mu(R)^{-1/2}$ if $v = 0$ and no subdivision into cases is necessary). □

We are now ready to prove the main result of this section.

Proposition 7.3. *There holds that*

$$\Big|\sum_{R\in\mathcal{D}'_{\mathrm{good}}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good}},\, Q\subset R\\ \ell(Q)<\delta^r\ell(R)}}\langle g,\psi_R\rangle\tilde{T}_{RQ}\langle\varphi_Q,f\rangle\Big| \lesssim \|g\|_{L^{p'}(X,Y^*)}\|f\|_{L^p(X,Y)}.$$

Proof. Let $s(R)$ denote the number of subcubes of a cube $R \in \mathcal{D}'$ and set $s = \max_{R\in\mathcal{D}'} s(R) \lesssim 1$. Fix $w \in \{1,\ldots,s\}$ and $m \in \{r+1, r+2, \ldots\}$. The already used randomization trick gives

$$\begin{aligned}
&\Big|\sum_{k\in\mathbb{Z}}\sum_{R\in\mathcal{D}'_{\mathrm{good},k}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ Q\subset R_w}}\langle g,\psi_R\rangle\tilde{T}_{RQ}\langle\varphi_Q,f\rangle\Big|\\
&\lesssim \|g\|_{L^{p'}(X,Y^*)}\Big\|\sum_{k\in\mathbb{Z}}\epsilon_k\sum_{R\in\mathcal{D}'_{\mathrm{good},k}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ Q\subset R_w}}\psi_R\tilde{T}_{RQ}\langle\varphi_Q,f\rangle\Big\|_{L^p(\Omega\times X,Y)}.
\end{aligned}$$

We introduce the relevant kernels now. Indeed, set

$$\begin{aligned}
K^c_R &= \delta^{-\alpha m/2}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ Q\subset R_w}}\mu(R)\,\chi_{R\setminus R_w}(x)\,\psi_R(x)\,\tilde{T}_{RQ}\,\varphi_Q(y)\,b_1(y),\\
K^i_R &= \delta^{-\alpha m/2}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ Q\subset R_w}}\mu(R_w)\,\chi_{R_w}(x)\,\psi_R(x)\,\tilde{T}_{RQ}\,\varphi_Q(y)\,b_1(y).
\end{aligned}$$

The previous lemma yields at once that $|K_R^c(x,y)| \lesssim 1$ and $|K_R^i(x,y)| \lesssim 1$. Also, the supports lie in $R \times R$ and $R_w \times R_w$ respectively. There holds

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \epsilon_k \sum_{R\in\mathcal{D}'_{\mathrm{good},k}} \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ Q\subset R_w}} \psi_R(x)\, \tilde{T}_{RQ}\langle \varphi_Q, f\rangle \\
&\quad = \delta^{\alpha m/2} \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{R\in\mathcal{D}'_{\mathrm{good},k}} \frac{\chi_R(x)}{\mu(R)} \int_R K_R^c(x,y) \frac{\chi_R(y)\Delta_{k+m}^{b_1} f(y)}{b_1(y)}\, d\mu(y) \\
&\quad + \delta^{\alpha m/2} \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{R\in\mathcal{D}'_{\mathrm{good},k}} \frac{\chi_{R_w}(x)}{\mu(R_w)} \int_{R_w} K_R^i(x,y) \frac{\chi_{R_w}(y)\Delta_{k+m}^{b_1} f(y)}{b_1(y)}\, d\mu(y).
\end{aligned}$$

The tangent martingale trick cannot quite yet be used: the measurability conditions need not hold (note the important difference with the argument of the previous section – there we did not have the dyadic systems $\mathcal{D}$ and $\mathcal{D}'$ mixed in the way we have here). To fix this, one simply defines new partitions

$$\mathcal{F}_k = \{S \cap Q \neq \emptyset : S \in \mathcal{D}'_k,\ Q \in \mathcal{D}_{k-r-1}\},$$

and exploits the goodness of the cubes R via the observations

$$\mathcal{D}'_{\mathrm{good},k} \subset \mathcal{F}_k \qquad \text{and} \qquad \{R_w \in \mathcal{D}'_{k+1} : R_w \subset R \in \mathcal{D}'_{\mathrm{good},k}\} \subset \mathcal{F}_{k+1}.$$

We then extend the above sums to be over the sets $\mathcal{F}_k$ and $\mathcal{F}_{k+1}$ respectively by using zero kernels for all the new sets R. We may then apply the tangent martingale trick after passing to the obvious subseries over k yielding, just like in the previous section, the bound

$$\begin{aligned}
&\Big|\sum_{k\in\mathbb{Z}} \sum_{R\in\mathcal{D}'_{\mathrm{good},k}} \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+m}\\ Q\subset R_w}} \langle g, \psi_R\rangle\, \tilde{T}_{RQ}\langle \varphi_Q, f\rangle\Big| \\
&\qquad\qquad \lesssim \delta^{\alpha m/2}\,(m+r+1)\,\|g\|_{L^{p'}(X,Y^*)}\,\|f\|_{L^p(X,Y)},
\end{aligned}$$

from which the claim follows after summing over $m = r+1, r+2, \ldots$ and $w = 1, \ldots, s$. □

8. The correction term and the relevant paraproduct

Recall that we subtracted $\langle b_2, T(b_1\varphi_Q)\rangle\langle\psi_R\rangle_Q$ from T_{RQ} in the case $R \in \mathcal{D}'_{\mathrm{good}}$, $Q \in \mathcal{D}_{\mathrm{good}}$, $Q \subset R$ and $\ell(Q) < \delta^r \ell(R)$. Thus, we now need to consider the sum

$$\sum_{R\in\mathcal{D}'_{\mathrm{good}}} \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good}},\, Q\subset R\\ \ell(Q)<\delta^r\ell(R)}} \langle g, \psi_R\rangle\, \langle b_2, T(b_1\varphi_Q)\rangle\, \langle\psi_R\rangle_Q\, \langle \varphi_Q, f\rangle. \tag{8.1}$$

Recall also that we always have the suppressed summation over u, v and the restriction that $\delta^{k_0} < \ell(Q), \ell(R) \leq \delta^m$. Writing out the above sum unhiding these conventions and then recalling that e.g. $\Delta_Q^{b_1} f = \sum_u b_1 \varphi_{Q,u} \langle \varphi_{Q,u}, f \rangle$, we see that (writing explicitly only the relevant restrictions)

$$(8.1) = \sum_{\substack{Q \in \mathcal{D}_{\text{good}} \\ \ell(Q) > \delta^{k_0}}} \Bigg(\sum_{\substack{R \in \mathcal{D}'_{\text{good}},\, R \supset Q \\ \delta^{-r}\ell(Q) < \ell(R) \le \delta^m}} \langle \Delta_R^{b_2} g / b_2 \rangle_Q + \sum_{\substack{R \in \mathcal{D}'_{\text{good}},\, R \supset Q \\ \delta^{-r}\ell(Q) < \ell(R) = \delta^m}} \langle E_R^{b_2} g / b_2 \rangle_Q \Bigg) \langle T^* b_2, \Delta_Q^{b_1} f \rangle.$$

Now we use the trick from [9] noting that the inner summation would collapse to $\langle E_R^{b_2} g / b_2 \rangle_Q = \langle g \rangle_R / \langle b_2 \rangle_R$, where $R \in \mathcal{D}'$ is the unique cube of generation $\text{gen}(Q) - r$ for which $Q \subset R$, were it not for the restriction to good $\mathcal{D}'$-cubes in the summation. Now it is clear why Lemma 4.8 was worth proving. Indeed, we may achieve this effect just by considering the grid $\mathcal{D}'$ being fixed and averaging over all the other random quantities used in the randomization of cubes (that is: υ' is fixed and we average over υ, t_α^k, u_α^k). We use Lemma 4.8 twice. First, to remove the restriction to good R, and after collapsing the series, to put the restriction back. This yields

$$\mathbb{E}(8.1) = \mathbb{E} \sum_{Q \in \mathcal{D}_{\text{good}}} \sum_{\substack{R \in \mathcal{D}'_{\text{good}},\, R \supset Q \\ \ell(R) = \delta^{-r}\ell(Q)}} \frac{\langle g \rangle_R}{\langle b_2 \rangle_R} \langle T^* b_2, \Delta_Q^{b_1} f \rangle$$
$$= \mathbb{E} \sum_{R \in \mathcal{D}'_{\text{good}}} \sum_{\substack{Q \in \mathcal{D}_{\text{good}},\, Q \subset R \\ \ell(Q) = \delta^{r}\ell(R)}} \frac{\langle g \rangle_R}{\langle b_2 \rangle_R} \langle T^* b_2, b_1 \varphi_Q \rangle \langle \varphi_Q, f \rangle,$$

where the standard summation conditions were yet again suppressed.

Notice now that the right hand side of this is the expectation of a pairing $\langle \Pi g, f \rangle$, where we have (for every fixed choice of the random quantities) the paraproduct

$$\Pi g = \sum_{R \in \mathcal{D}'_{\text{good}}} \sum_{\substack{Q \in \mathcal{D}_{\text{good}},\, Q \subset R \\ \ell(Q) = \delta^{r}\ell(R)}} \frac{\langle g \rangle_R}{\langle b_2 \rangle_R} \langle T^* b_2, b_1 \varphi_Q \rangle \, \varphi_Q.$$

We shall next study this with any fixed choice of the random quantities. Note that in [6] the paraproduct had the inessential difference that instead of the requirement of Q being good we had the requirement $d(Q, X \setminus R) \ge C C_0 \ell(Q)$ (which follows from the goodness), and the essential difference that the bigger cubes were not restricted to good cubes. As was noted in [9], this restriction is useful in this vector valued context.

Lemma 8.1. *If $\varphi \in \text{BMO}_\kappa^p(\mu)$, then*

$$\Bigg\| \sum_{\substack{Q \in \mathcal{D}_{\text{good}},\, Q \subset R \\ \ell(Q) \le \delta^r \ell(R)}} \epsilon_Q \, \langle \varphi, b_1 \varphi_Q \rangle \, \varphi_Q \Bigg\|_{L^p(\Omega \times X)} \lesssim \mu(R)^{1/p} \|\varphi\|_{\text{BMO}_\kappa^p(\mu)}.$$

Proof. This can be proven similarly as Lemma 7.1 in [6], borrowing some minor additional ingredients related to this vector valued context from the proof of Lemma 9.3 of [9]. □

Since $T^*b_2 \in \mathrm{BMO}^q_\kappa(\mu)$ for any $1 \le q < \infty$ (see Remark 2.2 and Theorem 3.3), the previous lemma is important in proving that the paraproduct Π is bounded. We will not provide the exact details instead citing [9] as this part of the argument no longer has anything special to do with the metric space structure or with our use of more general measures. Indeed, having been able to do all these reductions in the metric space setting, one can now follow the argument found in pages 32–33 of [9] pretty much word to word (when reading that, notice that §3 of [9] is already in an abstract form suitable for us), and this yields:

Proposition 8.2. *We have*

$$\|\Pi g\|_{L^{p'}(X,Y^*)} \lesssim \|T^*b_2\|_{\mathrm{BMO}^{p'}_\kappa(\mu)} \|g\|_{L^{p'}(X,Y^*)} \lesssim \|g\|_{L^{p'}(X,Y^*)}.$$

The main result of this section now readily follows.

Proposition 8.3. *We have that*

$$\Big|\mathbb{E} \sum_{R \in \mathcal{D}'_{\mathrm{good}}} \sum_{\substack{Q \in \mathcal{D}_{\mathrm{good}},\, Q \subset R \\ \ell(Q) < \delta^r \ell(R)}} \langle g, \psi_R\rangle \langle b_2, T(b_1\varphi_Q)\rangle \langle \psi_R\rangle_Q \langle \varphi_Q, f\rangle\Big| \lesssim \|g\|_{L^{p'}(X,Y^*)} \|f\|_{L^p(X,Y)},$$

where we average over all the random quantities used in the randomization of the cubes (υ, υ', t^k_α, u^k_α).

9. Estimates for adjacent cubes of comparable size

We shall now deal with the part of the series where good cubes $Q \in \mathcal{D}_{\mathrm{good}}$ and $R \in \mathcal{D}'_{\mathrm{good}}$ are adjacent ($d(Q,R) < CC_0 \min(\ell(Q), \ell(R))$) and of comparable size ($|\mathrm{gen}(Q) - \mathrm{gen}(R)| \le r$). We denote the last condition by $\ell(Q) \sim \ell(R)$. Also, only the size, and not the cancellation, properties of the adapted Haar functions are used.

We are given some fixed small $\epsilon > 0$. Given cubes Q and R define $\Delta = Q \cap R$,

$$\delta_Q = \{x : d(x,Q) \le \epsilon\ell(Q) \text{ and } d(x, X \setminus Q) \le \epsilon\ell(Q)\},$$

and

$$\delta_R = \{x : d(x,R) \le \epsilon\ell(R) \text{ and } d(x, X \setminus R) \le \epsilon\ell(R)\}.$$

Set also

$$Q_s = Q \setminus \Delta \setminus \delta_R, \quad Q_\partial = Q \setminus \Delta \setminus Q_s, \quad R_s = R \setminus \Delta \setminus \delta_Q \quad \text{and} \quad R_\partial = R \setminus \Delta \setminus R_s.$$

Given $R \in \mathcal{D}'_{\text{good}}$, there are only finitely many $Q \in \mathcal{D}_{\text{good}}$ which are adjacent to R and of comparable size. Thus, one needs only to study finitely many subseries

$$\sum_{R \in \mathcal{D}'_{\text{good}}} \langle g, \psi_R \rangle T_{RQ} \langle \varphi_Q, f \rangle,$$

where $Q = Q(R)$ is implicitly a function of R – a convention that is used throughout this section. We shall also act like the mapping $R \mapsto Q(R)$ is invertible – this only amounts to identifying some terms with zero (if there are no preimages) or splitting into finitely many new subseries using the triangle inequality (if there are multiple preimages).

Recall that $T_{RQ} = \langle \psi_R b_2, T(b_1 \varphi_Q) \rangle$. We note that

$$b_1 \varphi_Q \langle \varphi_Q, f \rangle = \sum_{\substack{Q' \in \mathcal{D}: Q' \subset Q \\ \ell(Q') = \delta \ell(Q)}} b_1 \chi_{Q'} \langle \varphi_Q \rangle_{Q'} \langle \varphi_Q, f \rangle = \sum_{\substack{Q' \in \mathcal{D}: Q' \subset Q \\ \ell(Q') = \delta \ell(Q)}} b_1 \chi_{Q'} A_{Q'},$$

where $A_{Q'} = \langle \varphi_Q \rangle_{Q'} \langle \varphi_Q, f \rangle$. Similarly there holds

$$b_2 \psi_R \langle g, \psi_R \rangle = \sum_{\substack{R' \in \mathcal{D}': R' \subset R \\ \ell(R') = \delta \ell(R)}} b_2 \chi_{R'} B_{R'},$$

where $B_{R'} = \langle \psi_R \rangle_{R'} \langle g, \psi_R \rangle$. Thus, we are left with finitely many new subseries of the form

$$\sum_{R \in \mathcal{D}'} B_R \langle \chi_R b_2, T(b_1 \chi_Q) \rangle A_Q,$$

where $Q = Q(R)$ is a new function of R but one still has $\ell(Q) \sim \ell(R)$. Note also that the parents of these cubes are always good.

Given R and then $Q = Q(R)$ as in the above sum, we shall now split the pairing $\langle \chi_R b_2, T(b_1 \chi_Q) \rangle$ into five terms. Such a simple decomposition is only possible because we have exploited the work that has already been done in the scalar case [6] to update our WBP into a stronger one (see Remark 2.2). While this part of the argument could be made self-contained, this saves us from a lot of problems which are even worse in this vector-valued setting than in the scalar setting.

We now decompose

$$\begin{aligned} \langle \chi_R b_2, T(b_1 \chi_Q) \rangle &= \langle \chi_{R_s} b_2, T(b_1 \chi_Q) \rangle + \langle \chi_{R_\partial} b_2, T(b_1 \chi_Q) \rangle \\ &\quad + \langle \chi_\Delta b_2, T(b_1 \chi_\Delta) \rangle + \langle \chi_\Delta b_2, T(b_1 \chi_{Q_\partial}) \rangle + \langle \chi_\Delta b_2, T(b_1 \chi_{Q_s}) \rangle \\ &= A + B + C + D + E. \end{aligned}$$

It is time to deal with these terms now. These belong to various different groups: we have the terms A and E with separation, the terms B and D involving bad boundary regions, and the diagonal term C, which needs the stronger WBP if one wants to avoid complicated additional surgery. (For the details of the more complicated surgery in the metric situation see §8 and §9 in [6].)

Also, when we sum over R we have to use different kinds of strategies involving simple randomization (for the diagonal term), the tangent martingale trick (for the separated terms) and a certain improvement of the contraction principle (for the bad boundary region terms). In the bad boundary region terms control is gained only after using the a priori boundedness of T, and in these cases it is essential to get a small constant in front so that these may later be absorbed. This requires that we average over all the dyadic grids.

We have $C = \alpha_\Delta \mu(\Delta)$, where $|\alpha_\Delta| \lesssim 1$, since actually $|\langle T(\chi_A b_1), \chi_A b_2\rangle| \lesssim \mu(A)$ with any Borel set $A \subset X$. Using randomization, Hölder's inequality and the contraction principle, we obtain (denoting the dyadic parent of Q by $\tilde{Q}$ and similarly for R) that

$$\begin{aligned}
\Big|\sum_R B_R C(R) A_Q\Big| &= \Big|\int_\Omega \int_X \sum_R \epsilon_R \chi_R B_R \sum_{R'} \epsilon_{R'} \alpha_\Delta A_{Q(R')} \chi_{Q(R')}\,d\mu\,d\mathbb{P}\Big| \\
&\le \Big\|\sum_R \epsilon_R \chi_R B_R\Big\|_{L^{p'}(\Omega\times X, Y^*)} \Big\|\sum_Q \epsilon_Q \alpha_\Delta A_Q \chi_Q\Big\|_{L^p(\Omega\times X, Y)} \\
&\lesssim \Big\|\sum_R \epsilon_R \psi_{\tilde{R}} \langle g, \psi_{\tilde{R}}\rangle\Big\|_{L^{p'}(\Omega\times X, Y^*)} \Big\|\sum_Q \epsilon_Q \varphi_{\tilde{Q}} \langle \varphi_{\tilde{Q}}, f\rangle\Big\|_{L^p(\Omega\times X, Y)} \\
&\lesssim \|g\|_{L^{p'}(X,Y^*)} \|f\|_{L^p(X,Y)}.
\end{aligned}$$

Let us then estimate the separated terms A and E. However, these are so similar that we only explicitly handle A here. The first kernel estimate yields

$$|A| = |\langle \chi_{R_s} b_2, T(b_1 \chi_Q)\rangle| \lesssim \int_{R_s} \int_Q \frac{1}{\lambda(y, d(x,y))}\,d\mu(y)\,d\mu(x).$$

Then we note that

$$\lambda(y, d(x,y)) \ge \lambda(y, d(x,Q)) \ge \lambda(y, \epsilon\ell(Q)) \gtrsim \epsilon^d \lambda(y, \ell(Q)).$$

Thus, we may write

$$A = \beta_Q \frac{\mu(Q)\mu(R)}{\inf_{y\in Q} \lambda(y, \ell(Q))},$$

where $|\beta_Q| \lesssim \epsilon^{-d}$ (note that the infimum may be zero only if $\mu(Q) = 0$). Now we may write

$$\begin{aligned}
\sum_R B_R A(R) A_Q &= \sum_R \langle g, \psi_{\tilde{R}}\rangle\, \langle \psi_{\tilde{R}}\rangle_R\, \beta_Q\, \frac{\mu(Q)\mu(R)}{\inf_{y\in Q} \lambda(y, \ell(Q))}\, \langle \varphi_{\tilde{Q}}\rangle_Q\, \langle \varphi_{\tilde{Q}}, f\rangle \\
&= \sum_R \langle g, \psi_{\tilde{R}}\rangle\, \|\psi_{\tilde{R}}\|_{L^1(\mu)}\, \frac{\tilde{\beta}_Q}{\inf_{y\in Q} \lambda(y, \ell(Q))}\, \|\varphi_{\tilde{Q}}\|_{L^1(\mu)}\, \langle \varphi_{\tilde{Q}}, f\rangle,
\end{aligned}$$

where $|\tilde{\beta}_Q| \le |\beta_Q| \lesssim \epsilon^{-d}$. Recall that these parents $\tilde{R}$ and $\tilde{Q}$ are again good cubes. Also recall that every cube has at most $\lesssim 1$ children. So it remains to study the series

$$\sum_R \langle g, \psi_R\rangle\, \|\psi_R\|_{L^1(\mu)}\, \frac{\sigma_Q}{\inf_{y\in Q} \lambda(y, \ell(Q))}\, \|\varphi_Q\|_{L^1(\mu)}\, \langle \varphi_Q, f\rangle,$$

where again $|\sigma_Q| \lesssim \epsilon^{-d}$ (note that $\lambda(y,\ell(\tilde{Q})) \lesssim \lambda(y,\ell(Q))$). Using a randomization trick and then reindexing the summation we see that this may be dominated by

$$\|g\|_{L^{p'}(X,Y^*)}\Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{S\in\mathcal{D}_k} \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+2r}\\ Q\subset S}} \|\psi_R\|_{L^1(\mu)}\psi_R(x) \cdot \frac{\sigma_Q}{\inf_{y\in Q}\lambda(y,\ell(Q))}\|\varphi_Q\|_{L^1(\mu)}\,\langle \varphi_Q, f\rangle \Big\|_{L^p(\Omega\times X,Y)}.$$

Since R is good, $\ell(R) \le \delta^{-r}\ell(Q) = \delta^r\ell(S)$ and $CC_0\ell(R) > d(Q,R)$, one easily checks that $R \subset S$ (if r is large enough). We then set for $S \in \mathcal{D}_k$ that

$$K_S(x,y) = \epsilon^d \sum_{\substack{Q\in\mathcal{D}_{\mathrm{good},k+2r}\\ Q\subset S}} \|\psi_R\|_{L^1(\mu)}\psi_R(x)\frac{\mu(S)}{\inf_{w\in Q}\lambda(w,\ell(Q))}\sigma_Q\|\varphi_Q\|_{L^1(\mu)}\varphi_Q(y)b_1(y),$$

and note that the previous majorant can now be written in the form

$$\epsilon^{-d}\Big\|\sum_{k\in\mathbb{Z}}\epsilon_k\sum_{S\in\mathcal{D}_k}\frac{\chi_S(x)}{\mu(S)}\int_S K_S(x,y)\,\frac{\chi_S(y)\Delta^{b_1}_{k+2r}f(y)}{b_1(y)}\,d\mu(y)\Big\|_{L^p(\Omega\times X,Y)},$$

which is amenable to the tangent martingale trick as is next demonstrated. Indeed, just note that K_S is supported on $S\times S$ and that $|K_S(x,y)| \lesssim 1$ holds, and then divide the summation over k into $\lesssim 1$ appropriate pieces to get that

$$\Big|\sum_R B_R A(R) A_Q\Big| \lesssim \epsilon^{-d}\|g\|_{L^{p'}(X,Y^*)}\|f\|_{L^p(X,Y)}.$$

The same, as already stated earlier, works with A replaced by E.

It still remains to deal with the terms B and D involving bad boundary regions. The small term in front of $\|T\|$ is gained only after averaging over the dyadic grids $\mathcal{D}$ and $\mathcal{D}'$. We only deal with the term B – the term D is handled completely analogously.

We turn to the details. Using randomization, Hölder's inequality and the a priori boundedness of T one gets that

$$\Big|\sum_R B_R B(R) A_Q\Big| \le \|T\|\Big\|\sum_R \epsilon_R B_R \chi_{R_\partial} b_2\Big\|_{L^{p'}(\Omega\times X,Y^*)}\Big\|\sum_Q \epsilon_Q A_Q b_1\chi_Q\Big\|_{L^p(\Omega\times X,Y)}.$$

Now, the second term is easily seen to be dominated by $\|f\|_{L^p(X,Y)}$ using the contraction principle and unconditionality.

The first term is more involved since it is here that the small factor needs to be extracted. Let us define

$$\delta(k) = \bigcup_{j=k-r}^{k+r}\bigcup_{Q\in\mathcal{D}_j}\delta_Q.$$

Note that if $\mathrm{gen}(R) = k$, then $\mathrm{gen}(Q(R)) \in [k-r, k+r]$, and so we must have $\chi_{R_\partial} = \chi_{R_\partial}\chi_{\delta(k)}\chi_R$. Throwing χ_{R_∂} and b_2 away using the contraction principle, we get

$$\Big\| \sum_R \epsilon_R B_R \chi_{R_\partial} b_2 \Big\|_{L^{p'}(\Omega\times X, Y^*)} \lesssim \Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \chi_{\delta(k)} \sum_{R\in\mathcal{D}'_k} B_R\chi_R \Big\|_{L^{p'}(\Omega\times X, Y^*)}.$$

Now, keeping everything else fixed, we take the conditional expectation of this over the grids $\mathcal{D}$. Using Jensen's inequality and Fubini's theorem, we get

$$\begin{aligned} &\mathbb{E}\Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \chi_{\delta(k)} \sum_{R\in\mathcal{D}'_k} B_R\chi_R \Big\|_{L^{p'}(\Omega\times X, Y^*)} \\ &\quad\lesssim \Big(\int_X \mathbb{E}\Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \chi_{\delta(k)}(x) \sum_{R\in\mathcal{D}'_k} B_R\chi_R(x) \Big\|_{L^{p'}(\Omega, Y^*)}^{p'} d\mu(x)\Big)^{1/p'}. \end{aligned}$$

In order to gain access to a certain improvement of the contraction principle (to be formulated shortly), it is still beneficial to further dominate this by

$$\Big(\int_X \Big[\mathbb{E}\Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \chi_{\delta(k)}(x) \sum_{R\in\mathcal{D}'_k} B_R\chi_R(x) \Big\|_{L^{p'}(\Omega, Y^*)}^{t}\Big]^{p'/t} d\mu(x)\Big)^{1/p'},$$

where $t \ge p'$. We now fix t once and for all demanding only that it is larger than p, p', the cotype of Y and the cotype of Y^* (recall that the dual of a UMD space is UMD and that a UMD space has nontrivial cotype). The requirements involving p and the cotype of Y are only needed when handling the similar term D.

We now formulate the contraction principle we need (this is Lemma 3.1 of [7]).

Proposition 9.1. *Suppose Z is a Banach space of cotype $s \in [2,\infty)$, $\xi_j \in Z$, $s < u < \infty$ and $\theta_j \in L^u(\tilde\Omega)$ (here $\tilde\Omega$ is a probability space). Then*

$$\Big\| \sum_{j=1}^\infty \epsilon_j \theta_j \xi_j \Big\|_{L^u(\tilde\Omega, L^2(\Omega, Z))} \lesssim \sup_j \|\theta_j\|_{L^u(\tilde\Omega)} \Big\| \sum_{j=1}^\infty \epsilon_j \xi_j \Big\|_{L^2(\Omega, Z)}.$$

Utilizing the above contraction principle together with Lemma 4.3 and Kahane's inequality gives (here the L^t norm is taken over the probability space used in the randomization of $\mathcal{D}$)

$$\begin{aligned} &\mathbb{E}\Big\| \sum_R \epsilon_R B_R \chi_{R_\partial} b_2 \Big\|_{L^{p'}(\Omega\times X, Y^*)} \\ &\quad\lesssim \Big(\int_X \sup_{k\in\mathbb{Z}} \|\chi_{\delta(k)}(x)\|_{L^t}^{p'} \Big\| \sum_{k\in\mathbb{Z}} \epsilon_k \sum_{R\in\mathcal{D}'_k} B_R\chi_R(x) \Big\|_{L^{p'}(\Omega, Y^*)}^{p'} d\mu(x)\Big)^{1/p'} \\ &\quad\lesssim \epsilon^{\eta/t} \Big\| \sum_R \epsilon_R B_R \chi_R \Big\|_{L^{p'}(\Omega\times X, Y^*)} \lesssim \epsilon^{\eta/t} \|g\|_{L^{p'}(X, Y^*)}. \end{aligned}$$

We now formulate the above considerations as a proposition.

Proposition 9.2. *Let $\epsilon > 0$. We have the estimate*

$$\mathbb{E}\Big|\sum_{R\in\mathcal{D}'_{\mathrm{good}}}\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good}}:\,\ell(Q)\sim\ell(R)\\ d(Q,R)<CC_0\min(\ell(Q),\ell(R))}}\langle g,\psi_R\rangle T_{RQ}\langle\varphi_Q,f\rangle\Big|$$
$$\lesssim C(\epsilon)\,\|g\|_{L^{p'}(X,Y^*)}\,\|f\|_{L^p(X,Y)}+\|T\|\,c(\epsilon)\,\|g\|_{L^{p'}(X,Y^*)}\,\|f\|_{L^p(X,Y)},$$

where we average over all the random quantities used in the randomization of the cubes, and $c(\epsilon)$ can be made arbitrarily small by choosing ϵ small enough.

Remark 9.3. Recall that when we dealt with the separated cubes in Proposition 6.3 we had the assumption that the adapted Haar functions related to the smaller cubes are cancellative (have zero integral). Note that there are only boundedly many terms with $\ell(Q)=\ell(R)=\delta^m$ where the contrary can happen (due to the assumptions about the supports of the functions f and g). Thus, the relevant arguments involving separated sets used in the present section let us also remove this assumption.

10. Completion of the proof

Combining all that we have done in the previous sections shows that

$$\mathbb{E}\Big|\sum_{\substack{Q\in\mathcal{D}_{\mathrm{good}},\,R\in\mathcal{D}'_{\mathrm{good}}\\ \delta^{k_0}<\ell(Q),\,\ell(R)\le\delta^m}}\sum_{u,v}\langle\varphi^{b_2}_{R,v},g\rangle\langle b_2\varphi^{b_2}_{R,v},T(b_1\varphi^{b_1}_{Q,u})\rangle\langle\varphi^{b_1}_{Q,u},f\rangle\Big|\lesssim C(\epsilon)+c(\epsilon)\|T\|,$$

where $c(\epsilon)\to 0$ when $\epsilon\to 0$. Recalling (5.1) the estimate $\|T\|\lesssim 1$ follows by taking ϵ small enough. We have proved what we set out to prove, namely Theorem 2.1.

References

[1] AUSCHER, P. AND HYTÖNEN, T.: Orthonormal bases of regular wavelets in spaces of homogeneous type. *Appl. Comput. Harmon. Anal.*, in press.

[2] CHRIST, M.: A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral. *Colloq. Math.* **60/61** (1990), no. 2, 601–628.

[3] DAVID, G., JOURNÉ, J.-L. AND SEMMES, S.: Opérateurs de Calderón–Zygmund, fonctions para-accrétives et interpolation. *Rev. Mat. Iberoamericana* **1** (1985), no. 4, 1–56.

[4] HYTÖNEN, T. AND KAIREMA, A.: Systems of dyadic cubes in a doubling metric space. *Colloq. Math.* **126** (2012), no. 1, 1–33.

[5] HYTÖNEN, T., LIU, S., YANG, D. AND YANG, D.: Boundedness of Calderón–Zygmund operators on non-homogeneous metric measure spaces. *Canad. J. Math.* **64** (2012), no. 4, 892–923.

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

[7] Hytönen, T. and Veraar, M.: R-boundedness of smooth operator-valued functions. *Integral Equations and Operator Theory* **63** (2009), no. 3, 373–402.

[8] Hytönen, T.: A framework for non-homogeneous analysis on metric spaces, and the RBMO space of Tolsa. *Publ. Mat.* **54** (2010), no. 2, 485–504.

[9] Hytönen, T.: The vector-valued non-homogeneous Tb theorem. To appear in *Int. Math. Res. Not.*

[10] Müller, P. F. X. and Passenbrunner, M.: A decomposition theorem for singular integral operators on spaces of homogeneous type. *J. Funct. Anal.* **262** (2012), no. 4, 1427–1465.

[11] Macías, R. A. and Segovia, C.: Lipschitz functions on spaces of homogeneous type. *Adv. in Math.* **33** (1979), no. 3, 257–270.

[12] Nazarov, F., Treil, S. and Volberg, A.: Weak type estimates and Cotlar inequalities for Calderón-Zygmund operators on nonhomogeneous spaces. *Internat. Math. Res. Notices* **1998**, no. 9, 463–487.

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

[14] Stein, E.: *Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals.* Princeton Mathematical Series 43, Monographs in Harmonic Analysis III, Princeton University Press, Princeton, New Jersey, 1993.

[15] Volberg, A. and Wick, B. D.: Bergman-type singular integral operators on metric spaces. Preprint, `arXiv:1001.0038`, 2009.

[16] Volberg, A. and Wick, B. D.: Bergman-type singular integral operators and the characterization of Carleson measures for Besov–Sobolev spaces on the complex ball. *Amer. J. Math.* **134** (2012), no. 4, 949–992.

Received May 17, 2010; revised February 20, 2012.

Henri Martikainen: Department of Mathematics and Statistics, University of Helsinki, P.O.B. 68, 00014 Helsinki, Finland.
E-mail: henri.martikainen@helsinki.fi

The author is supported by the Academy of Finland through the project "L^p methods in harmonic analysis". The paper is part of the author's doctoral thesis project written under the supervision of Tuomas Hytönen – the guidance of whom is gratefully acknowledged.