Acta Math., 190 (2003), 151–239

The Tb-theorem on non-homogeneous spaces

by

F. NAZAROV
Michigan State University
East Lansing, MI, U.S.A.

S. TREIL
Brown University
Providence, RI, U.S.A.

and

A. VOLBERG
Michigan State University
East Lansing, MI, U.S.A.

Contents

All authors are partially supported by the NSF Grant DMS 9970395.

Notation

$:=$	equal by definition;
μ	measure on $\mathbf{R}^N$;
$B(x_0,r)$	the open ball, $B(x_0,r):=\{x\in\mathbf{R}^N: \lvert x-x_0\rvert<r\}$;
$\langle f,g\rangle$	standard linear duality, $\langle f,g\rangle:=\int fg\,d\mu$;
$\mathrm{BMO}^p_\lambda(\mu)$	BMO-space, see §1.1;
$\mathcal{D}$	a collection of dyadic cubes, see below;
χ_Q	characteristic function (indicator) of the set Q;
$l(Q)$	"size" of the cube $Q\subset\mathbf{R}^N$, i.e., the length of its side;
E_Q, E_k	averaging operators, see §4. For a cube Q, $E_Q f:=(\mu(Q)^{-1}\int_Q f\,d\mu)\cdot\chi_Q$; the operator E_k is defined by $E_k f:=\sum_{Q\in\mathcal{D},\, l(Q)=2^k} E_Q f$;
Δ_Q, Δ_k	martingale difference operators, see §4: $\Delta_k:=E_{k-1}-E_k$; for a cube Q of size 2^k ($l(Q)=2^k$) define $\Delta_Q f:=\chi_Q\cdot\Delta_k f$;
E^b_Q, E^b_k	weighted averaging operators, see §4: $E^b_Q f:=(\int_Q b\,d\mu)^{-1}\cdot(\int_Q f\,d\mu)\cdot b\chi_Q$, $E^b_k f:=\sum_{Q\in\mathcal{D},\, l(Q)=2^k} E^b_Q f$;
Δ^b_Q, Δ^b_k	weighted martingale difference operators, see §4: $\Delta^b_k:=E^b_{k-1}-E^b_k$; for a cube Q of size 2^k ($l(Q)=2^k$) define $\Delta^b_Q f:=\chi_Q\cdot\Delta^b_k f$;
$\mathbf{f}_Q, \langle f\rangle_Q$	average of the function f, $\mathbf{f}_Q=\langle f\rangle_Q:=\mu(Q)^{-1}\int_Q f\,d\mu$;
Π	paraproduct, see §7.1.

Cubes and dyadic lattices. Throughout the paper we will speak about dyadic cubes and dyadic lattices, so let us first fix some terminology. A *cube* in $\mathbf{R}^N$ is an object obtained from the *standard cube* $[0,1)^N$ by dilations and shifts.

For a cube Q we denote by $l(Q)$ its *size*, i.e. the length of its side. Given a cube Q one can split it into 2^N cubes Q_k of size $\frac12 l(Q)$: we will call such cubes Q_k the *subcubes* (*of* Q) *of the first generation*, or just simply *subcubes*.

For a cube Q and $\lambda>0$ we denote by λQ the cube Q dilated λ times with respect to its center.

Now, let us define the *standard dyadic lattice*: for each $k\in\mathbf{Z}$ let us consider the cube $[0,2^k)^N$ and all its shifts by elements of $\mathbf{R}^N$ with coordinates of form $j\cdot 2^k$, $j\in\mathbf{Z}$. The collection of all such cubes (union over all k) is called the *standard dyadic lattice.*

A *dyadic lattice* is just a shift of the standard dyadic lattice. The collection of all cubes from a dyadic lattice $\mathcal{D}$ of a fixed size 2^k is called a *dyadic grid.*

0. Introduction: main objects and results

The goal of this paper is to present a (more or less) complete theory of Calderón–Zygmund operators on non-homogeneous spaces. The theory can be developed in an abstract metric space with measure, but we will consider the interesting case for applications when our space is just a subset of $\mathbf{R}^N$.

Let μ be a Borel measure on $\mathbf{R}^N$. Let d be a positive number (not necessarily integer) and let the measure μ behave like a d-dimensional measure:

$$\mu(B(x,r))\leqslant r^d$$

for any ball $B(x,r)$ of radius r with center at x. A Calderón–Zygmund kernel (of dimension d) is a function $K(s,t)$ of two variables satisfying:

(i) $|K(s,t)|\leqslant C|s-t|^{-d}$;

(ii) there exists $\alpha>0$ such that

$$|K(s,t)-K(s_0,t)|,|K(t,s)-K(t,s_0)|\leqslant C\frac{|s-s_0|^\alpha}{|t-s_0|^{d+\alpha}},$$

whenever $|t-s_0|\geqslant 2|s-s_0|$.

If $d=N$ (N is the dimension of the underlying space $\mathbf{R}^N$), we have just a classical Calderón–Zygmund kernel.

We are interested in the question of when a Calderón–Zygmund operator (integral operator with kernel K, $Tf(x)=\int K(x,y)f(y)\,d\mu(y)$) is bounded on $L^p(\mu)$.

0.1. Main results

The main results that we state below look like they are just copied from some classical book. But let the reader not be misled, the results are new. We intentionally defined BMO in such a way that our theorems could be stated exactly as the corresponding classical results. However, the BMO we use is not exactly the space the reader is probably familiar with. Actually, there is a whole plethora of BMO-spaces generalizing the classical BMO-space to the non-homogeneous situation from the point of view of singular integral operators. There is one "more equal than others"—the RBMO of Xavier Tolsa, which is discussed and used in §1.2. But we feel that—at least at this stage of our understanding—it is a good idea to work with all definitions of BMO at once.

Our first two theorems deal with Calderón–Zygmund operators with antisymmetric kernels.

Let us mention that there is no canonical way to assign an operator to a *general* Calderón–Zygmund kernel. We cannot just say that $Tf(x)=\int K(x,y)f(y)\,d\mu(y)$, because for almost all x the functions $K(x,\cdot\,)$ and $K(\,\cdot\,,x)$ are not integrable, not even locally in the neighborhood of the singularity x.

However, if the kernel is antisymmetric ($K(x,y)=-K(y,x)$) there exists a canonical way to define an operator.

Namely, since the kernel K is antisymmetric, we have (formally)

$$\langle Tf,g\rangle=\iint K(x,y)f(y)g(x)\,d\mu(x)\,d\mu(y)=-\iint K(x,y)f(x)g(y)\,d\mu(x)\,d\mu(y),$$

and so

$$\langle Tf,g\rangle=\frac{1}{2}\iint K(x,y)[f(y)g(x)-f(x)g(y)]\,d\mu(x)\,d\mu(y).$$

But for smooth (even Lipschitz) compactly supported functions the last expression is well defined.

Namely, the integrand has the singularity bounded by $C/|x-y|^{d-1}$ for $x-y$ close to 0. By the Comparison Lemma (see Lemma 2.1 below) such a singularity is integrable (say, with respect to x), so the integral is well defined.

So, for an antisymmetric kernel one can canonically define a bilinear form $\langle Tf,g\rangle$ for compactly supported Lipschitz functions. The corresponding operator is called the *principal value*.

We think that unfortunately the terminology is confusing here, because *principal value* also means $\lim_{\varepsilon\to 0}\int_{|x-y|>\varepsilon}K(x,y)f(y)\,d\mu(y)$. We would prefer to use, and we will use in this paper, the term *canonical value*, or *canonical operator*. Unfortunately, *principal value* is now a widely accepted term.

Similarly, one can also define for antisymmetric kernels the bilinear form $\langle Tbf,bg\rangle$, $b\in L^\infty$, as

$$\langle Tbf,bg\rangle=\frac{1}{2}\iint K(x,y)[f(y)g(x)-f(x)g(y)]b(x)b(y)\,d\mu(x)\,d\mu(y),$$

where $b\in L^\infty$.

THEOREM 0.1 ($T1$-theorem). *Let $1<p<\infty$. The canonical value of a Calderón–Zygmund operator T with antisymmetric kernel extends to a bounded operator on $L^p(\mu)$ if and only if $T1$ belongs to* $\mathrm{BMO}=\mathrm{BMO}(\mu)$.

Moreover, the upper bound of the norm of T depends only on the dimensions N, d, the exponent p, the Calderón–Zygmund constants of the kernel K, and the BMO*-norm of $T1$.*

The definition of the space BMO is rather involved and requires separate discussion. We will discuss this space in detail later in §1.1.

Although the $T1$-theorem above gives a necessary and sufficient condition for a Calderón–Zygmund operator T to be bounded, it is not always easy to verify the condition $T1 \in \mathrm{BMO}$. But sometimes it is trivial to see that $Tb \in \mathrm{BMO}$ for some $b \in L^\infty$.

Let us call a bounded (complex-valued) function b weakly accretive (with respect to the measure μ) if there exists $\delta > 0$ such that for any cube Q,

$$\mu(Q)^{-1}\left|\int_Q b(s)\,d\mu(s)\right| \geqslant \delta.$$

Note that if b is weakly accretive then $|b| \geqslant \delta$ μ-a.e.

THEOREM 0.2 (Tb-theorem). *Let $1<p<\infty$, and let b be a weakly accretive function. The canonical value of a Calderón–Zygmund operator T with antisymmetric kernel extends to a bounded operator on $L^p(\mu)$ if and only if Tb belongs to* $\mathrm{BMO}=\mathrm{BMO}(\mu)$.

Moreover, the upper bound of the norm of T depends only on the dimensions N, d, the exponent p, the Calderón–Zygmund constants of the kernel K, the constant δ from the definition of weak accretivity, $\|b\|_\infty$, and the BMO*-norm of Tb.*

A similar Tb-theorem in the homogeneous case (the measure μ is doubling) was used to prove the boundedness of the Cauchy transform on Lipschitz curves.

The following two theorems should be treated as some kind of *meta-theorems*. As we already mentioned above, there is no canonical way to define a Calderón–Zygmund operator in the general case. There are several possible interpretations, which we will discuss in §0.3. So for each interpretation of Calderón–Zygmund operators, Theorems 0.3 and 0.4 below should be interpreted accordingly.

THEOREM 0.3 ($T1$-theorem). *Let $1<p<\infty$. A Calderón–Zygmund operator T extends to a bounded operator on $L^p(\mu)$ if and only if it is weakly bounded and $T1, T^*1$ belong to* $\mathrm{BMO}=\mathrm{BMO}(\mu)$.

Being weakly bounded in the simplest case means that there exist $\Lambda \geqslant 1$, $C<\infty$ such that $|\langle T\chi_Q, \chi_Q\rangle| \leqslant C\mu(\Lambda Q)$ for any cube Q. There are alternative definitions (not equivalent) that would also work. We will discuss them later in §0.3.

Again, the estimate of the norm of T depends only on the constants involved, namely the dimensions N and d, the exponent p, the Calderón–Zygmund constants of the kernel, the BMO-norms of $T1$, T^*1, and the constant C from the definition of weak boundedness.

Suppose that we are given two weakly accretive functions b_1 and b_2.

THEOREM 0.4 (Tb-theorem). *Let $1<p<\infty$, and let b_1, b_2 be two weakly accretive functions. A Calderón–Zygmund operator T extends to a bounded operator on $L^p(\mu)$ if and only if the operator b_2Tb_1 is weakly bounded and Tb_1, T^*b_2 belong to* $\mathrm{BMO}=\mathrm{BMO}(\mu)$.

Again, the upper bound on the norm of T depends only on constants involved.

We postpone the discussion of weak boundedness to §0.3, and one can find a more specific discussion in §11. The subtle point here is that as one makes a weaker assumption of "weak boundedness", the assumptions of accretivity one should require become stronger.

Our Tb-theorems are the extensions to the case of non-doubling measures of the Tb-theorems obtained by G. David, J.-L. Journé and S. Semmes [6], [9], [10] for Calderón–Zygmund operators on $\mathbf{R}^N$ with respect to Lebesgue measure. It was clear that such Tb-theorems apply to arbitrary spaces of homogeneous type, a general setting for singular integral theory introduced by Coifman and Weiss [4]. In particular, the boundedness of the Cauchy operator on chord-arc curves could have been obtained directly from homogeneous Tb-theorems. (Notice that the more general case of Ahlfors–David curves required extra important ideas [5].) The Calderón–Zygmund theory on homogeneous spaces acquired a new approach from the work of M. Christ [2], where an *accretive system* Tb-theorem for homogeneous spaces has been proved (the difference with the Tb-theorems of David, Journé and Semmes is in using a collection of b's instead of one such function). This allowed one, for example, to obtain the boundedness of the Cauchy operator on Ahlfors–David curves from the homogeneous Tb-theorems of Christ's type. More generally this allowed one to obtain a Tb-proof of T. Murai's [19] theorem which characterized compact *homogeneous* sets of finite length on the plane for which the Cauchy operator is bounded. So *almost everything* homogeneous became clear.

However, quite unexpectedly, the homogeneity is something one can dispense with. The first results in this direction dealt with the Cauchy integral operator. A version of the $T1$-theorem for the Cauchy integral operator in the non-homogeneous setting was proved independently and with different methods by X. Tolsa [26] and by the authors [21]. Note that in [21] the case of more general Calderón–Zygmund operators was also treated.

An alternative and very interesting approach to the $T1$-theorem for the Cauchy operator was introduced by J. Verdera in [32].

Then in [22] Cotlar inequalities and weak type $(1,1)$ estimates were proved for Calderón–Zygmund operators bounded on $L^2(\mu)$. In particular, this implied that, as in the classical case, if a Calderón–Zygmund operator is bounded on $L^2(\mu)$, then it also extends to a bounded operator on all $L^p(\mu)$, $1<p<\infty$. Thus, the theory of Calderón–Zygmund operators on non-homogeneous spaces was almost complete.

Our $T1$- and Tb-theorems complete the theory.(1) Also, in [23] we prove a non-homogeneous analogue of Christ's Tb-theorem, which allows us, for example, to extend Murai's theorem [19], and to fully describe compact sets of finite length on the plane for which the Cauchy operator is bounded. The technique in [23] is an extension of the technique we use in the present article.

So, the main goal of this article (as well as articles [22], [21] and some subsequent ones) is to build a non-homogeneous theory for Calderón–Zygmund operators. There are several possible applications of such a theory. One is presented below in §0.2. Also, for motivation, see the introduction to [22].

The proofs in the paper are rather technical and can be very complicated, but essentially everything is based on two main ideas: estimating the matrix in the Haar basis (weighted Haar basis for the Tb-theorem) and eliminating bad cases by averaging over random dyadic lattices.

Neither idea is new. The Haar system was used by Coifman–Jones–Semmes [3] in their elementary proof of the L^2-boundedness of the Cauchy integral operator on Lipschitz curves (it is the earliest use of the Haar system for estimates of singular integral operators we are aware of). Later it became commonplace and was used by many authors, see [6], [7], [11], [29], [20], [21].

The idea of averaging over dyadic grids is not new either. In [13], for example, it was used by Garnett and Jones to pass from results about dyadic BMO to classical ones. The idea of averaging was also used by E. Sawyer [24] in his proof of two weight estimates for the maximal operator.

However, we introduced a new twist to this idea: we use averaging to show that one can ignore bad situations if they have small probability (pulling yourself up by the hair). This trick was first used in [20] to simplify the presentation: an "honest" estimate, not resorting to the averaging trick, is also possible there. Later in [21] we noticed that the same trick can be used to deal with Calderón–Zygmund operators on non-homogeneous spaces (measure without doubling).

It is very well known to everybody who was working with singular integral operators that if one tries to estimate the matrix of a Calderón–Zygmund operator in the Haar basis, it is impossible to get good estimates when the support of one Haar function is close to the jumps of another. But, one has to be especially unlucky to really have the worst case estimate for any given pair of Haar functions: shifting a bit the boundaries of the cubes improves the estimates with high probability.

---

(1) Of course, there are still some open questions, for example, about existence of *principal values*. For the Cauchy operator on a non-homogeneous space it was proved by X. Tolsa [25], but for general Calderón–Zygmund operators the question is still open.

Thus, to improve the estimates, G. David and P. Mattila (see [7], [11]) used curved "squares" avoiding areas where the measure is concentrated.

In our approach, we consider random dyadic grids to show that with non-zero probability the "bad" part of the Haar expansion has small norm. This allows us to get an estimate of the norm as soon as we have some a priori estimate (usually very weak) of the operator. For example, it works if we know that its bilinear form is bounded for smooth compactly supported functions, see §0.3 below for details.

This idea for Calderón–Zygmund operators on non-homogeneous spaces was introduced by us in [21] where we proved the $T1$-theorem for Cauchy-type operators. Here we further refine it: we are relaxing all the assumptions, proving the results in most generality. The surprising thing for us was that this trick allows us to relax significantly the weak boundedness assumptions (in comparison with [21]), see §10.

0.2. An application of the $T1$-theorem: electric intensity capacity

As a possible application of our non-homogeneous $T1$-theorem we will cite the following result about the so-called *electric intensity* capacity (also known as *harmonic Lipschitz* capacity, see [11], [14], [16], [31] for some interesting related results). Let us consider the following problem.

Suppose that we have a compact set K in $\mathbf{R}^3$. We want to find what maximal possible charge one can put on K such that the intensity of the resulting electric field is bounded by 1. Note that if we require the *potential* to be bounded by 1, we get the usual capacity from physics. But in engineering it is often very important to have the intensity of the electric field bounded, so our capacity has very good physical meaning.

In this problem we forbid negative densities.

Let us now formally state the problem. Given a compact set K in $\mathbf{R}^N$, $N\geqslant 3$, consider the class S of all subharmonic functions φ ($-\varphi$ is the potential) in $\mathbf{R}^N$ such that

(i) φ is harmonic in $\mathbf{R}^N\setminus K$, $\varphi(\infty)=0$;

(ii) $|\nabla\varphi(x)|\leqslant 1$ for almost all (with respect to N-dimensional Lebesgue measure) $x\in\mathbf{R}^N$ (intensity is bounded by 1).

The *electric intensity capacity* (also known as *positive harmonic Lipschitz* capacity) $\operatorname{cap}_{\mathrm{ei}}(K)$ of the compact set K is defined by

$$\operatorname{cap}_{\mathrm{ei}}(K) := \sup_{\varphi\in S} |C_\varphi|,$$

where C_φ is the leading coefficient in the asymptotic expansion

$$\varphi(x) = C_\varphi/|x|^{N-2} + o(1/|x|^{N-2})$$

of the function $\varphi \in S$ at ∞ (any function φ harmonic in a neighborhood of ∞ and satisfying $\varphi(\infty)=0$ admits such an expansion). Note that in $\mathbf{R}^3$ the constant $-C_\varphi$ is exactly the charge on K.

To state our result we need to introduce one more capacity, the so-called *operator* capacity. Given a Borel measure μ, consider the "Cauchy" transforms T_j^μ, $1 \leqslant j \leqslant N$, $T_j^\mu f(x) := \int K_j(x,y) f(y)\, d\mu(y)$, where $K_j(x,y) = (x_j - y_j)/|x-y|^N$. Note that the kernels K_j are antisymmetric Calderón–Zygmund kernels of dimension $d=N-1$. Since the kernels are antisymmetric, we have no problems defining the operator (just use the canonical value).

The *operator* capacity $\operatorname{cap}_{\mathrm{op}}(K)$ is defined by

$$\operatorname{cap}_{\mathrm{op}}(K) = \sup\{\mu(K) : \mu \geqslant 0,\ \operatorname{supp} \mu \subset K,\ \|T_j^\mu\|_{L^2(\mu)} \leqslant 1 \text{ for } 1 \leqslant j \leqslant N\};$$

here μ stands for a non-negative Borel measure.

THEOREM 0.5. *The capacities* $\operatorname{cap}_{\mathrm{op}}$ *and* $\operatorname{cap}_{\mathrm{ei}}$ *are equivalent, i.e., there exist constants* c, C, $0 < c \leqslant C < \infty$, *depending only on the dimension* N, *such that*

$$c \cdot \operatorname{cap}_{\mathrm{op}}(K) \leqslant \operatorname{cap}_{\mathrm{ei}}(K) \leqslant C \cdot \operatorname{cap}_{\mathrm{op}}(K).$$

As an immediate corollary of this result we obtain that the capacity $\operatorname{cap}_{\mathrm{ei}}$ is semiadditive, i.e.,

$$\operatorname{cap}_{\mathrm{ei}}(K_1 \cup K_2) \leqslant C \cdot (\operatorname{cap}_{\mathrm{ei}}(K_1) + \operatorname{cap}_{\mathrm{ei}}(K_2)).$$

This follows immediately from Theorem 0.5, because for the capacity $\operatorname{cap}_{\mathrm{op}}$ we trivially have

$$\operatorname{cap}_{\mathrm{op}}(K_1 \cup K_2) \leqslant \operatorname{cap}_{\mathrm{op}}(K_1) + \operatorname{cap}_{\mathrm{op}}(K_2).$$

Sketch of the proof of Theorem 0.5. Let $\mu := \Delta\varphi = \sum_{j=1}^N \partial^2\varphi/\partial x_j^2$ be the Riesz measure of the function φ.

Since $|\nabla\varphi(x)| \leqslant 1$, it is an easy exercise using Green's formula to check that $\mu(B) \leqslant Cr^{N-1}$ for any ball of radius r. Indeed, let us apply Green's formula

$$\int_\Omega (u\Delta v - v\Delta u)\, dV = \int_{\partial\Omega} \left(u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n} \right) dS$$

to $u \equiv 1$, $v = \varphi$ and $\Omega = B = B(x_0, r)$. We get

$$\int_B d\mu = \int_B \Delta\varphi\, dV = \int_{\partial B} \frac{\partial\varphi}{\partial n}\, dS.$$

Since $|\partial\varphi/\partial n| \leqslant |\nabla\varphi| \leqslant 1$, the measure $\mu(B)$ is estimated by the $(N-1)$-dimensional measure of the sphere ∂B, which is $C_N r^{N-1}$.

We know that the jth coordinate of the gradient $\nabla\varphi$ is given (up to a multiplicative constant) by $K_j^\mu 1=\int K_j(x,y)\,1\,d\mu(y)$. From here we conclude that $T_j^\mu 1\in L^\infty$, $\|T_j^\mu 1\|_\infty\leqslant 1$, $1\leqslant j\leqslant N$. Since $L^\infty\subset\mathrm{BMO}$, the $T1$-theorem (Theorem 0.1) implies that the operators T_j are bounded, and therefore

$$c\cdot\mathrm{cap_{op}}(K)\leqslant\mathrm{cap_{ei}}(K).$$

The reverse estimate is rather standard and well known (at least in the homogeneous case). First of all, it was proved in [22] (for the non-homogeneous case) that if a Calderón–Zygmund operator T extends to a bounded operator on $L^2(\mu)$, then it extends to a bounded operator on all $L^p(\mu)$, $1<p<\infty$, and, moreover, it is of weak type $(1,1)$. It was also proved there that in this case the truncated operators T_r (integrals are taken over the set $|x-y|>r$) are also bounded on $L^p(\mu)$, $1<p<\infty$, and are of weak type $(1,1)$ uniformly in r.

Therefore, the "Cauchy" transform $T^\mu=(T_1^\mu,T_2^\mu,\dots,T_N^\mu)^T$ (which maps scalar-valued functions to $\mathbf{R}^N$-valued), as well as its adjoint and the corresponding truncated operators are (uniformly) of weak type $(1,1)$.

So, applying to the truncated "Cauchy" transform Theorem 0.6 below (see [1, Theorem VII.23] for the scalar version), we get the desired estimate.

Let $\mathcal{M}$ denote the space of all finite measures (signed, or complex) on a locally compact Hausdorff space $\mathcal{X}$, and let $C(\mathcal{X},\mathbf{R}^N)$ be the space of all functions continuous on $\mathcal{X}$ with values in $\mathbf{R}^N$. The dual of this space, $\mathcal{M}(\mathcal{X},\mathbf{R}^N)$, is the space of all $\mathbf{R}^N$-valued finite Borel measures on $\mathcal{X}$.

THEOREM 0.6. *Let $\mathcal{X}$ be a locally compact Hausdorff space, let μ be a Radon measure on $\mathcal{X}$, and $T:\mathcal{M}\to C(\mathcal{X},\mathbf{R}^N)$ a bounded linear operator. Suppose that the adjoint operator T^* is of weak type $(1,1)$, that is, there exists $A<\infty$ such that*

$$\mu\{x:|T^*\nu(x)|>\alpha\}\leqslant A\alpha^{-1}\|\nu\|$$

for all $\alpha>0$ and $\nu\in\mathcal{M}(\mathcal{X},\mathbf{R}^N)$. Then for any Borel set $E\subset\mathcal{X}$ with $0<\mu(E)<\infty$, there exists $h:\mathcal{X}\to[0,1]$ satisfying

$$h(x)=0 \quad \textit{for all } x\notin E,\qquad \int_E h\,d\mu\geqslant\tfrac12\mu(E)\quad \textit{and}\quad \|T(h\,d\mu)\|_\infty\leqslant 4A.$$

This theorem is well known in the scalar case, although we cannot be sure where it first appeared. It can be found in [30], [12]. The proof below is presented only for the sake of completeness, since we follow the scalar case proof presented in [1] (see the proof of Theorem VII.23 there) almost up to the letter.

It is interesting that the argument below involves dualizing a weak type $(1,1)$ inequality, although the corresponding weak space $L^{1,\infty}$ has no reasonable dual.

Proof. Suppose that the conclusion of the theorem fails for some E. Define

$$\begin{aligned} B_0 &:= \left\{ f : \mathcal{X} \to [0,1] : f = 0 \ \mu\text{-a.e. on } \mathcal{X} \setminus E \text{ and } \int_E f\,d\mu \geqslant \tfrac{1}{2}\mu(E) \right\}, \\ B_1 &:= \{ T(f\,d\mu) : f \in B_0 \}, \\ B_2 &:= \{ g \in C(\mathcal{X}, \mathbf{R}^N) : \|g\|_\infty < 4A \}. \end{aligned}$$

The Hahn–Banach Theorem implies that there exists a bounded linear functional l on $C(\mathcal{X}, \mathbf{R}^N)$ separating B_1 and B_2, i.e., $l(g) \leqslant l(h)$ for all $h \in B_1$, $g \in B_2$. Let λ be the $\mathbf{R}^N$-valued measure representing l, so

$$\int [T(f\,d\mu)]^T\,d\lambda \geqslant \int g^T\,d\lambda \quad \text{for all } f \in B_0,\ g \in B_2.$$

Taking the supremum on the right-hand side and using the identity $\int [T(f\,d\mu)]^T\,d\lambda = \int f \cdot T^*\lambda\,d\mu$, we get

$$\int f \cdot T^*\lambda\,d\mu \geqslant 4A\|\lambda\|. \tag{0.1}$$

To use the weak $(1,1)$ estimate set $\alpha = 3A\|\lambda\|/\mu(E)$ and note that

$$\mu\{x \in E : |T^*\lambda(x)| > \alpha\} \leqslant A\|\lambda\|/\alpha = \tfrac{1}{3}\mu(E),$$

so that

$$\mu\{x \in E : |T^*\lambda(x)| \leqslant \alpha\} \geqslant \tfrac{2}{3}\mu(E).$$

Therefore we can find a closed set $F \subset E$ such that $\mu(F) \geqslant \frac{1}{2}\mu(E)$ and $|T^*\lambda| \leqslant \alpha$ a.e. on F. Take $f := \chi_F$. Then $f \in B_0$ and

$$\left| \int f \cdot T^*\lambda\,d\mu \right| \leqslant \alpha\mu(F) \leqslant 3A\|\lambda\|,$$

which contradicts (0.1). □

That completes the proof of Theorem 0.5. □

0.3. How to interpret a Calderón–Zygmund operator T

Let us discuss here how one can interpret the above results, first how one can define the operator T for general kernels. Recall that for antisymmetric kernels we can define the operator as the *canonical value*, see §0.1.

A typical Calderón–Zygmund kernel (think, for example, of the kernel $1/(x-y)$ on the real line $\mathbf{R}$ with Lebesgue measure) is such that for almost all x the functions $K(x,\cdot)$, $K(\cdot,x)$ are not in L^1, not even locally, in a neighborhood of the singularity x.

In the case of the kernel $1/(x-y)$ on the real line one can still define the operator on smooth functions with compact support if one interprets the integral $\int_{-\infty}^{\infty} K(x,y)f(y)\,dy$ as *principal value*, i.e., as the limit

$$\text{p.v.}\int_{-\infty}^{\infty} K(x,y)f(y) := \lim_{\varepsilon\to 0}\int_{|y-x|>\varepsilon} K(x,y)f(y)\,dy.$$

However, if one considers a general Calderón–Zygmund kernel, it is not clear why the principal value exists.(2)

The classical way to interpret T was to assume that the bilinear form $\langle Tf,g\rangle$ of the operator T (or of the operator b_2Tb_1 in the case of the Tb-theorem) is initially well defined on a dense set of nice functions f,g, for example for $f,g\in C_0^\infty$ (C^∞-functions with compact support). In other words, the bilinear form $\langle Tf,g\rangle$ is well defined and continuous (with respect to the topology of C_0^∞) for $f,g\in C_0^\infty$.

One can replace here C_0^∞ by the Schwartz class $\mathcal{S}$ of rapidly decaying C^∞-functions: it really does not matter.

The words that T is an integral operator with kernel K mean only that

$$\langle Tf,g\rangle = \iint K(x,y)g(x)f(y)\,d\mu(x)\,d\mu(y) \tag{0.2}$$

for compactly supported f,g with *separated* compact supports, when the integral is well defined. Notice that the kernel K does not determine the operator uniquely: for example, any multiplication operator $f\mapsto\varphi f$ is a Calderón–Zygmund operator with kernel 0.

This observation is a commonplace for specialists, but it can be really surprising for a beginner.

Now we are going to give three ways to interpret a Calderón–Zygmund operator T with kernel K. In all cases we assume that a bilinear form of the operator T is defined for some class of functions, and that for functions with separated compact supports equality (0.2) holds.

0.3.1. *The bilinear form is defined on Lipschitz functions.* Since for antisymmetric kernels the bilinear form $\langle Tf,g\rangle$ (or $\langle b_2Tb_1f,g\rangle$ for the Tb-theorem) is well defined for Lipschitz functions f,g (see §0.1 above), it seems reasonable to assume that this is the case for general kernels as well.

(2) We should mention here a remarkable result of X. Tolsa [25] that if a Cauchy integral $Tf(z):=\int_{\mathbf{C}}(f(\xi)/(\xi-z))\,d\mu(\xi)$ is a bounded operator on $L^2(\mu)$, then for any $g\in L^2(\mu)$ the principal value of Tg exists μ-a.e.

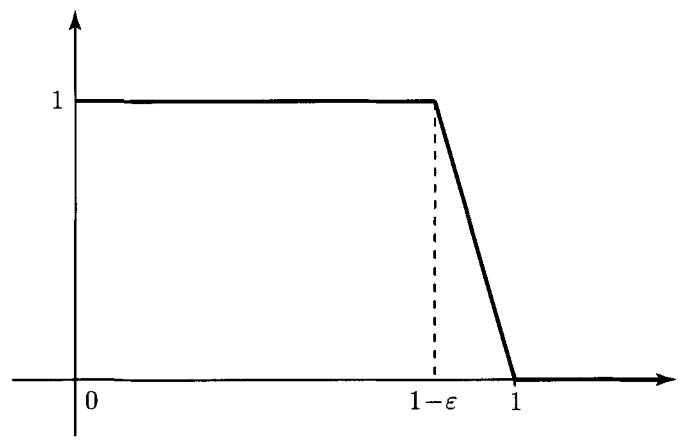

Fig. 1. The function σ^ε.

Weak boundedness in this case means that the following two conditions are satisfied:

(i) For all pairs of Lipschitz functions φ_1, φ_2 satisfying $|\varphi_{1,2}(x)-\varphi_{1,2}(y)| \leqslant L\cdot|x-y|$, supported by bounded sets D_1, D_2, respectively, and such that $\|\varphi_{1,2}\|_\infty \leqslant 1$, the inequalities

$$|\langle Tb_1\varphi_1, b_2\varphi_2\rangle|, |\langle Tb_1\varphi_2, b_2\varphi_1\rangle| \leqslant CL\cdot\|b_1\|_\infty\cdot\|b_2\|_\infty\cdot\operatorname{diam}(D_1)\cdot\mu(D_2)$$

should hold for weakly accretive functions b_1, b_2 (this is for the Tb-theorem, for the $T1$-theorem $b_1=b_2=1$).

As Lemma 11.3 below shows, this condition (with $b_1=b_2$ as in the corresponding Tb-theorem) holds for antisymmetric kernels.

(ii) Let σ^ε be a function as in Figure 1. For a cube Q let ϱ_Q be its Minkowski functional

$$\varrho_Q(x) := \inf\{\lambda>0 : \lambda Q \ni x\}$$

and let

$$\sigma_Q^\varepsilon(x) := \sigma^\varepsilon(\varrho_Q(x)).$$

(Clearly σ_Q^ε is a Lipschitz function with Lipschitz norm at most $C/r\varepsilon$.)

We will require that for all cubes Q,

$$|\langle Tb_1\sigma_Q^\varepsilon, b_2\sigma_Q^\varepsilon\rangle| \leqslant C\mu(\lambda' Q)$$

for some $\lambda' \geqslant 1$, uniformly in ε and Q.

The last condition (with $b_1=b_2=b$) definitely holds for antisymmetric kernels, since for such kernels $\langle Tb\sigma_Q^\varepsilon, b\sigma_Q^\varepsilon\rangle = 0$.

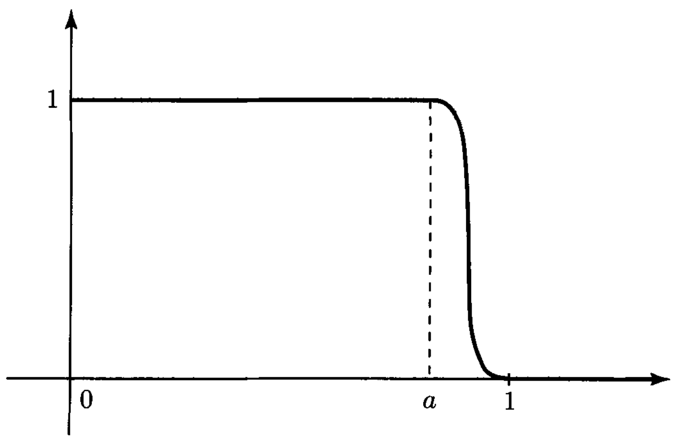

Fig. 2. The function σ.

0.3.2. *The bilinear form is defined on smooth functions.* We do not think that it makes much sense in our situation to assume that the bilinear form $\langle b_2 T b_1 f, g\rangle$ (or $\langle Tf, g\rangle$) is defined for smooth (say C_0^∞) functions f and g. We really do not see how additional smoothness (in comparison with Lipschitz functions) can help.

However, it is still possible to assume that the bilinear form is defined for C_0^∞-functions. In this case we have to assume more about the functions b_1, b_2 in the Tb-theorem: we want them to be *sectorial*. Let us recall that a function b is called *sectorial* if $b\in L^\infty$, and there exists a constant $\xi\in\mathbf{C}$, $|\xi|=1$, such that $\operatorname{Re}\xi b\geqslant\delta>0$.

The advantage is that we can relax the assumptions of the week boundedness in this case. Namely, fix a C^∞-function σ on $[0,\infty)$ such that $0\leqslant\sigma\leqslant 1$, $\sigma\equiv 1$ on $[0,a]$ $(0<a<1)$ and $\sigma\equiv 0$ on $[1,\infty)$, see Figure 2. The parameter a is not essential here, but we already have too many parameters in what follows, so let us fix some a, say $a=0.9$. For a ball $B=B(x_0,r)$ let $\sigma_B(x):=\sigma(|x-x_0|/r)$. Clearly, σ_B is supported by the ball B and is identically 1 on the ball $0.9B$. We will require that for any concentric balls B_1, B_2 of comparable sizes, say $\frac{1}{2}\operatorname{diam}(B_1)\leqslant\operatorname{diam}(B_2)\leqslant 2\operatorname{diam}(B_1)$, the following inequality holds:

$$|\langle T\sigma_{B_1}b_1, \sigma_{B_2}b_2\rangle|\leqslant C\mu(B), \tag{0.3}$$

where B is the largest of the two balls B_1, B_2. We can even replace $\mu(B)$ by $\mu(\lambda B)$, $\lambda>1$ here.

0.3.3. *A priori boundedness.* We feel that the most natural way to interpret a Calderón–Zygmund operator T is to think that we are not given the operator T per se, but that its kernel K is "approximated" in some sense by "nice" kernels K_ε, and we are interested in the question of when the operators T_ε with kernels K_ε are uniformly bounded.

A typical example one should think of is to consider *truncated* operators T_ε,

$$T_\varepsilon f(x) := \int_{|x-y|>\varepsilon} K(x-y) f(y)\, d\mu(y).$$

Such truncated operators are clearly well defined on compactly supported functions. Moreover, for compactly supported f and g, $\operatorname{diam}(\operatorname{supp}(f)) \leqslant A$, $\operatorname{diam}(\operatorname{supp}(g)) \leqslant A$, one has

$$|\langle T_\varepsilon f, g\rangle| \leqslant C(\varepsilon, A) \|f\|_2 \|g\|_2. \tag{0.4}$$

That will be our main way of interpretation. It was shown in [22] that under our assumptions about the measure and the kernel, if a Calderón–Zygmund operator T is bounded on $L^2(\mu)$ (or in some L^{p_0}, $1<p_0<\infty$), then it is bounded on all $L^p(\mu)$, $1<p<\infty$, and the maximal operator $T^\#$,

$$T^\# f(x) = \sup_{\varepsilon>0} \left| \int_{|x-y|>\varepsilon} K(x-y) f(y)\, d\mu(y) \right|,$$

is bounded on all $L^p(\mu)$ as well.

This implies that all truncated operators T_ε are uniformly bounded, so it is reasonable to think of boundedness of T as the uniform boundedness of T_ε.

So Theorem 0.3 can be interpreted in the following way: *a sequence of truncated operators T_ε is uniformly bounded if and only if the sequence is weakly bounded* (*with uniform estimates*) *and $T1, T^*1 \in \mathrm{BMO}$ with uniform estimates on the norms.*

There is a small technical problem with such an interpretation: the truncated operators T_ε are not Calderón–Zygmund operators (their kernels do not satisfy the property (ii) above).

Fortunately, this is not a real problem, and we know at least two ways of coping with it. First, two lemmas below where we use property (ii), namely Lemma 6.1 and Lemma 7.3, are true for truncated Calderón–Zygmund kernels as well: one just has to integrate a positive function not over a cube, but over a "truncated" cube, and that can only yield a better estimate.

Another possibility is to replace truncated operators by nicer regularizations of the operator T which have Calderón–Zygmund kernels. Namely, let

$$\Phi_\varepsilon(t) = \begin{cases} t/\varepsilon, & t \in [0, \varepsilon], \\ 1, & t \geqslant \varepsilon, \end{cases}$$

see Figure 3. Then the kernels $K_\varepsilon(x,y) := K(x,y)\Phi_\varepsilon(|x-y|)$ are clearly Calderón–Zygmund kernels with uniform estimates on all Calderón–Zygmund constants.

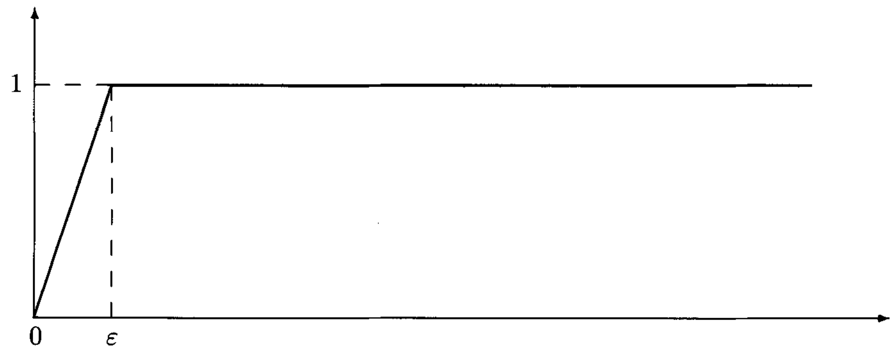

Fig. 3. The function Φ_ε.

It is also easy to see that for $|x-y| \leqslant \varepsilon$ we have $|K_\varepsilon(x,y)| \leqslant C/|x-y|^{d-1}$. So, applying the Comparison Lemma below (see Lemma 2.1), we get that for measures with compact support, $\int |K_\varepsilon(x,y)|\, d\mu(x) \leqslant C$, $\int |K_\varepsilon(x,y)|\, d\mu(y) \leqslant C$, and by the Schur Lemma the operators with kernels K_ε are bounded (but not necessarily uniformly in ε). Moreover, the same Comparison Lemma together with the Schur Test imply that the Calderón–Zygmund operator with kernel K_ε and the corresponding truncated operator T_ε differ by a bounded operator (uniformly in ε).

One can also consider two-sided truncations $T_{r,\varepsilon}$ of the operator T,

$$T_{r,\varepsilon} f(x) := \int_{\varepsilon < |x-y| < r} K(x-y) f(y)\, d\mu(y).$$

Such operators are clearly bounded. Moreover, such operators $T_{r,\varepsilon}$ are uniformly bounded (or $T_{r,\varepsilon}1$, $T^*_{r,\varepsilon}1$ are uniformly in BMO) if and only if the corresponding property holds for all one-sided truncations T_ε.

However, it is possible that we only have information about the truncations T_ε for small ε. Therefore, it makes sense to consider the case of one-sided truncations T_ε separately. So we will prove the main results under the assumption of boundedness only on compactly supported functions.

For two-sided truncations one can also replace (without losing anything) the truncated operator $T_{r,\varepsilon}$ by a nicer regularization, for example by the integral operator with kernel $K(x,y)\Phi_{\varepsilon,r}(|x-y|)$, where $\Phi_{\varepsilon,r}$ is the function in Figure 4.

There are several possible definitions of weak boundedness for the regularized operators T_ε (or $T_{r,\varepsilon}$). The simplest is to call the operator T weakly bounded if there exist $\Lambda \geqslant 1$, $C < \infty$ such that

$$|\langle T\chi_Q, \chi_Q \rangle| \leqslant C\mu(\Lambda Q)$$

for any cube Q. Another possibility is to consider the cube $Q' := aQ$ (for some fixed $a > 1$)

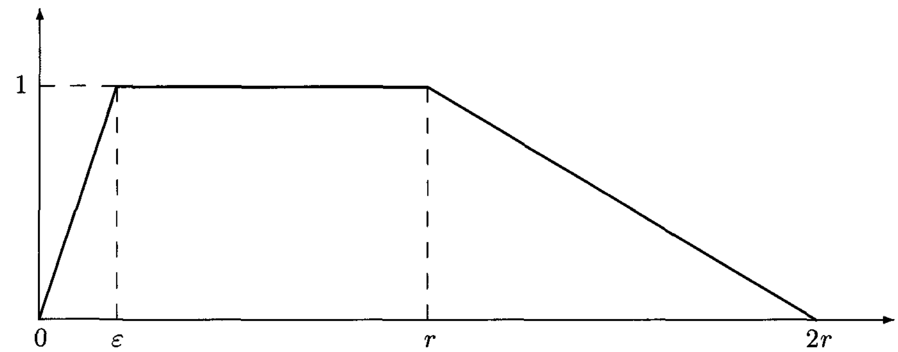

Fig. 4. The function $\Phi_{\varepsilon,r}$.

and require that

$$|\langle T\chi_{Q'},\chi_Q\rangle|\leqslant C\mu(\Lambda Q),\quad |\langle T\chi_Q,\chi_{Q'}\rangle|\leqslant C\mu(\Lambda Q').$$

One can also replace cubes by balls, to obtain two more definitions.

None of the four definitions above follows from any other (at least formally, we have not constructed any counterexamples), but any one of the definitions works if we assume a priori boundedness on compactly supported functions.

0.4. Plan of the paper

§1 is devoted to a discussion of the different BMO-spaces and the relations between them.

In §2 we deal with necessity. We will prove that if a Calderón–Zygmund operator T is bounded on $L^p(\mu)$, then for $b\in L^\infty$ we have $Tb\in\mathrm{BMO}^p_\lambda(\mu)$. In the same §2 we will also prove that if $Tb\in\mathrm{BMO}^p_\lambda(\mu)$ for some p, $1\leqslant p<\infty$, then $Tb\in\mathrm{RBMO}(\mu)$, and, therefore, $Tb\in\mathrm{BMO}^p_\lambda(\mu)$ for *all* p, $1\leqslant p<\infty$. We would like to emphasize that for an arbitrary function f the condition $f\in\mathrm{BMO}^p_\lambda(\mu)$, $p<2$, doesn't imply $f\in\mathrm{BMO}^2_\lambda(\mu)$, see §1.1.1. But Tb is not an arbitrary function, it possesses some additional regularity.

The rest of the paper is devoted to the sufficiency. We will only need to prove that the operator T is bounded on $L^2(\mu)$, because it was already proved in [22] that the boundedness on $L^2(\mu)$ implies the boundedness on all $L^p(\mu)$, $1<p<\infty$.

The idea of the proof is quite simple: consider a basis of "Haar functions" with respect to the measure μ (or weighted "Haar functions" for the Tb-theorem), and estimate the matrix of the operator T in this basis. To simplify the notation, it is more convenient to use the "coordinate-free" form of the decomposition with respect to the "Haar system", the so-called *martingale difference decomposition.*

In §§3–8 we introduce the main technical tools and gather all necessary estimates. Let us mention that in §3 we prove a generalization of the famous Carleson Embedding

Theorem to weighted Triebel–Lizorkin spaces. Although we only need the classical case $p=2$, we think that the theorem and its proof are of independent interest. However, the reader can skip this section if he or she wants, and use his or her favorite proof of the Carleson Embedding Theorem.

Then we do all necessary (and rather standard) constructions and estimates, such as decomposing functions into a martingale difference decomposition, estimating the matrix of the operator, constructing paraproducts, and getting the Carleson measure condition from $Tb\in\mathrm{BMO}$. All the ingredients should be very well known to a specialist, although non-homogeneity (non-doubling) of the measure adds quite a bit of specifics.

Finally, in §§9–10 we gather everything together to prove the theorems.

One of the main difficulties that appear when one works with non-doubling measures is an absence of good estimates of $\langle T\varphi_Q, \psi_R\rangle$ for functions φ_Q, ψ_R supported by the cubes Q and R respectively, when the cubes are close to each other. To overcome this difficulty we use averaging over random dyadic lattices and the "pulling yourself up by the hair" trick. One needs to use the trick several times to get the most general version of the theorem.

To give the reader a better understanding of the trick without getting lost in technical details, we first prove in §9 a weaker version of the Tb-theorem, where we use a stronger condition of weak boundedness. §10 deals with the full version of the theorem.

In §§9–10 we assume that the operator T is bounded on compactly supported functions (one should think of the truncated operators T_ε), i.e., $|\langle Tf, g\rangle|\leqslant C(A)\|f\|\cdot\|g\|$, where $A=\max\{\mathrm{diam}(\mathrm{supp}\, f), \mathrm{diam}(\mathrm{supp}\, g)\}$.

For many readers that will be enough because, as we already discussed above, the most natural way to interpret a Calderón–Zygmund operator T is to think of the sequence of truncated operators T_ε.

And finally, in the last section (§11) we reduce everything to the case of truncated operators. We consider the most general case, when the bilinear form of the operator is defined for smooth functions, or for Lipschitz functions, as in the case of the *canonical value* of an antisymmetric operator. We show that if such an operator satisfies the assumptions of our Tb-theorem, then the sequence of the truncated operators T_ε also satisfies these assumptions (uniformly in ε).

§11 has some ideas in common with §2.3, and uses some lemmas from this section, so it would be logical to place §11 right after §2.3.

However, the section is rather long and technical. Since we think that for many it is enough to just consider truncated operators T_ε, we decided to put §11 at the end of the paper.

1. Definitions of BMO-spaces

There are infinitely many different BMO-spaces that can be used in our theorems.

In the classical case, when μ is N-dimensional Lebesgue measure in $\mathbf{R}^N$, all of the definitions below give the well-known classical BMO.

First of all, there is a two-parameter family of spaces $\mathrm{BMO}^p_\lambda(\mu)$, $1\leqslant p<\infty$, $\lambda>1$, defined below in §1.1. The spaces $\mathrm{BMO}^p_\lambda(\mu)$ are quite different from classical BMO: in particular, the John–Nirenberg inequality fails for such spaces.

Then, there is a *regular* BMO-space $\mathrm{RBMO}(\mu)$ that was introduced by X. Tolsa [27]. This space is contained in $\bigcap_{1\leqslant p<\infty,\lambda>1}\mathrm{BMO}^p_\lambda(\mu)$, and it seems to be the most natural generalization of the classical BMO.

So, what space should we use in our theorems? And the answer is: it doesn't matter, one can use any one of the above spaces!

The space RBMO seems to be the most natural analogue of the classical BMO. However, the condition $T1\in\mathrm{RBMO}(\mu)$ (or $Tb\in\mathrm{RBMO}(\mu)$) is rather hard to verify. Therefore, let us think that BMO in the statements of our results means one of the spaces $\mathrm{BMO}^p_\lambda(\mu)$.

1.1. BMO^p_λ

Let $1\leqslant p<\infty$ and $\lambda>1$. We say that an $L^1_{\mathrm{loc}}(\mu)$-function f belongs to $\mathrm{BMO}^p_\lambda(\mu)$ if for any cube Q there exists a constant a_Q such that

$$\left(\int_Q |f-a_Q|^p\,d\mu\right)^{1/p}\leqslant C\mu(\lambda Q)^{1/p},$$

where the constant C does not depend on Q. The best constant C is called the $\mathrm{BMO}^p_\lambda(\mu)$-norm of f.

Using the standard reasoning from the classical BMO-theory one can replace the constant a_Q in the definition by the average $f_Q=\mu(Q)^{-1}\int_Q f\,d\mu$. Indeed,

$$\begin{aligned}|f_Q-a_Q| &= \left|\mu(Q)^{-1}\int_Q (f-a_Q)\,d\mu\right| \\ &\leqslant \left(\mu(Q)^{-1}\int_Q |f-a_Q|^p\,d\mu\right)^{1/p}\leqslant \|f\|_{\mathrm{BMO}^p_\lambda(\mu)}\left(\frac{\mu(\lambda Q)}{\mu(Q)}\right)^{1/p},\end{aligned}$$

and so, if we replace a_Q by f_Q in the definition, we just get an equivalent norm in $\mathrm{BMO}^p_\lambda(\mu)$.

Now we make several observations about the properties of BMO-spaces.

First we have the trivial inclusions: $\mathrm{BMO}^{p_2}_\lambda(\mu)\subset\mathrm{BMO}^{p_1}_\lambda(\mu)$ if $p_1<p_2$ (Hölder inequality) and $\mathrm{BMO}^p_\lambda(\mu)\subset\mathrm{BMO}^p_\Lambda(\mu)$ if $\lambda<\Lambda$.

It is not so trivial, but we will show this just below, that both inclusions are proper. Namely, the space $\mathrm{BMO}^p_\lambda(\mu)$ depends on p for $\lambda>1$.

Also, the space $\mathrm{BMO}^p_\lambda(\mu)$ does depend on λ. However, in the statement of the theorem any $\lambda>1$ would work.

And finally, $\mathrm{BMO}^p_1(\mu)$ ($\lambda=1$) is a wrong object for our theory: boundedness of T does not imply $T1\in\mathrm{BMO}^p_1(\mu)$.

Notice that one can introduce BMO-spaces where averages are taken over balls, not over cubes. But it is easy to see that if a function belongs to such a "ball" $\mathrm{BMO}^p_\lambda(\mu)$ then it belongs to the "cube" $\mathrm{BMO}^p_\Lambda(\mu)$ with $\Lambda=\sqrt{N}\,\lambda$. So, in the statements of the main results one can use the "ball" BMO as well.

Also, it does not matter whether we consider closed or open cubes (balls) in the definition of $\mathrm{BMO}^p_1(\mu)$. Formally, definitions with open cubes and with closed ones give us different spaces (because the boundary can have non-zero measure), but if the $\mathrm{BMO}^p_\lambda(\mu)$-condition is satisfied for open cubes, then for all closed ones the condition $\mathrm{BMO}^p_\Lambda(\mu)$ holds true for any $\Lambda>\lambda$.

Strangely enough, we will be using the assumptions $Tb\in\mathrm{BMO}^p_\lambda$ without requiring that T maps b to locally integrable functions. The interpretation follows the classical one—see §2 below.

This makes it slightly difficult to interpret $Tb\in\mathrm{RBMO}$, where RBMO is the "right" BMO-space found by Xavier Tolsa for non-homogeneous measures. The space RBMO is used in §1.2, and it is extremely useful because the space RBMO has the John–Nirenberg property (unlike BMO^p_λ—see the subsection below).

1.1.1. *Example*: $\mathrm{BMO}^p_\lambda(\mu)$ *does depend on* p. Let us explain why $\mathrm{BMO}^p_\lambda(\mu)$ does depend on p. Notice that the Hölder inequality implies that there is a trivial inclusion $\mathrm{BMO}^{p_2}_\lambda(\mu)\subset\mathrm{BMO}^{p_1}_\lambda(\mu)$ if $p_1<p_2$.

Let us have a careful look at the proof of the inclusion $\mathrm{BMO}^2_\lambda(\mu)\subset\mathrm{BMO}^1_\lambda(\mu)$:

$$\int_Q |f-f_Q|\,d\mu\leqslant\left(\int_Q |f-f_Q|^2\,d\mu\right)^{1/2}\mu(Q)^{1/2}\leqslant C\mu(\lambda Q)^{1/2}\mu(Q)^{1/2}.$$

Clearly $\mu(\lambda Q)^{1/2}\mu(Q)^{1/2}\leqslant\mu(\lambda Q)$, but since the measure μ is not doubling, the inverse inequality (with a constant) does not hold, and moreover, the gap can be huge. This can lead to the following example.

Let μ be a measure on $\mathbf{R}$ defined by $d\mu=w\,dt$ where $w=\varepsilon\chi_{[0,1]}+\chi_{\mathbf{R}\setminus[0,1]}$. Take $f=\chi_{[0,1]}$. It is an easy exercise to show that

$$\|f\|_{\mathrm{BMO}^p_\lambda(\mu)}\sim\varepsilon^{1/p}$$

(the interval $I=[0,1+\varepsilon]$ almost gives a supremum). Therefore the norms for different p are not equivalent as $\varepsilon\to 0$.

Now take a sequence $\varepsilon_k\searrow 0$ and a sequence of intervals I_k such that the intervals $2I_k$ are disjoint. Put $w=\sum_k \varepsilon_k\chi_{I_k}+\chi_E$ where $E=\mathbf{R}\setminus\bigcup_k I_k$. We leave to the reader to check that for $d\mu=w\,dt$,

$$\mathrm{BMO}_\lambda^{p_2}(\mu)\subsetneq \mathrm{BMO}_\lambda^{p_1}(\mu),\quad p_1<p_2.$$

1.1.2. *Example*: $\mathrm{BMO}_\lambda^p(\mu)$ *does depend on* λ. Let us consider the following measure μ on $\mathbf{R}$: on the intervals $[-2,-1]$ and $[1,2]$ it is just Lebesgue measure dx; on the intervals $\left[-\frac12,-\frac{5}{12}\right]$ and $\left[\frac{5}{12},\frac12\right]$ it is $\varepsilon\,dx$ where $\varepsilon>0$ is small; everywhere else μ is zero.

Define the function $f:=\varepsilon^{-1/p}(\chi_{[5/12,1/2]}-\chi_{[-1/2,-5/12]})$. Then for $\lambda\leqslant 2$ we have

$$\|f\|_{\mathrm{BMO}_\lambda^p(\mu)}\sim\varepsilon^{-1/p}$$

(consider the interval $I=\left[-\frac12,\frac12\right]$). However, $\|f\|_{\mathrm{BMO}_3^p(\mu)}\sim 1$. Indeed, if for an interval I we have $f\not\equiv f_I$, then I has to contain one of the following three intervals: $\left[-\frac{5}{12},\frac{5}{12}\right]$, $\left[\frac12,1\right]$ or $\left[-1,-\frac12\right]$. Then $\mu(3I)\geqslant\frac14$, so if we put a_I from the definition of $\mathrm{BMO}_\lambda^p(\mu)$ to be 0, we get

$$\int_I |f-a_I|^p\,d\mu=\int_I|f-0|^p\,d\mu\leqslant\frac16\leqslant\mu(3I).$$

Take a sequence $\varepsilon_k\to 0$, and let μ_k, f_k be the pair constructed above for $\varepsilon=\varepsilon_k$. Put $d\mu(x)=\sum_k d\mu_k(x-10k)$, $f(x)=\sum_k f_k(x-10k)$. Then clearly $f\in\mathrm{BMO}_3^p(\mu)$, but $f\notin\mathrm{BMO}_\lambda^p(\mu)$ with $\lambda\leqslant 2$.

Of course, the constants 3 and 2 are not essential here: for any pair λ_1,λ_2 satisfying $1<\lambda_1<\lambda_2$ one can easily modify the example to get a function $f\in\mathrm{BMO}_{\lambda_2}^p(\mu)$ such that $f\notin\mathrm{BMO}_{\lambda_1}^p(\mu)$.

1.1.3. *Example*: T *is bounded on* $L^p(\mu)\not\Leftrightarrow T1\in\mathrm{BMO}_1^p(\mu)$. Let us notice that this was proved independently, using another method, by J. Verdera [32].

Define a measure μ on $\mathbf{R}$ as follows: on the intervals $[1,2]$ and $[-2,-1]$ it is just Lebesgue measure dx; on the intervals $[-1,-1+\varepsilon]$ and $[1-\varepsilon,1]$ it is $0.1\,dx$; everywhere else μ is zero. Let T be the operator with kernel $K(s,t)=1/(t-s)$ (defined as principal value, i.e., as $\lim_\varepsilon\int_{|t-s|>\varepsilon}\dots$).

The operator T is bounded on $L^p(\mu)$, $1<p<\infty$, because the operator with kernel $1/(t-s)$ is bounded on $L^p(\mathbf{R},dx)$ (the operator is just the Hilbert transform up to a constant).

On the middle third $\left[1-\frac23\varepsilon,1-\frac13\varepsilon\right]$ of the interval $[1-\varepsilon,1]$ we can estimate (for small ε) $T1\geqslant c\log(1/\varepsilon)$, where c is some absolute constant.

Similarly, on the middle third of the interval $[-1,-1+\varepsilon]$ we have $T1\leqslant -c\log(1/\varepsilon)$. This implies that the norm of $T1$ in $\mathrm{BMO}_1^p(\mu)$ is at least $c_1^{1/p}\log(1/\varepsilon)$.

Again let $\varepsilon_k\to 0$, and let μ_k be the measures constructed above with $\varepsilon=\varepsilon_k$. We leave to the reader as an easy exercise to check that for $d\mu(x)=\sum_k d\mu(x-10^k)$ we have $T1\notin\mathrm{BMO}_1^p(\mu)$. (It is trivial that T is bounded on $L^p(\mu)$.)

1.2. RBMO and related spaces with the John–Nirenberg property

Recall that the measure μ under consideration satisfies

$$\mu(B(x,r))\leqslant r^d. \tag{1.1}$$

We will consider $f\in L^1_{\mathrm{loc}}(\mu)$ having the following property: for each cube Q there exists a number f_Q such that

$$\int_Q |f-f_Q|\leqslant B_1\mu(\varrho Q) \tag{1.2}$$

and such that for all cubes $Q\subset R$,

$$|f_R-f_Q|\leqslant B_2\cdot\left(1+\int_{2R\setminus Q}\frac{d\mu(x)}{|x-c_Q|^d}\right). \tag{1.3}$$

Such functions f will be called RBMO-functions, and the infimum of B_1+B_2 can be called the RBMO-norm. Let us make four remarks.

First, if we change $2R$ to λR, $\lambda>1$, the space does not change. This follows immediately from (1.1).

Second, we can change the parameter ϱ in (1.2) without changing the space. This follows from the following important lemma (we repeat the proof of [27] for the convenience of the reader).

LEMMA 1.1. *Let $1<\lambda<\varrho$ and let $f\in\mathrm{RBMO}$ in the sense that* (1.2) *and* (1.3) *are satisfied. Then*

$$\int_Q |f-f_Q|\leqslant B(B_1,B_2,\varrho,\lambda)\,\mu(\lambda Q).$$

Before proving the lemma, let us make a remark. For two cubes Q, R, we denote by $Q(R)$ the smallest cube concentric with Q and containing R. We call Q, R *neighbors* if the size of $Q(R)$ is at most 10 times the size of Q, and the size of $R(Q)$ is at most 10 times the size of Q. Given a function from RBMO (with its collection of f_Q's), it is easy to see from (1.3) and (1.1) that, if Q and R are neighbors, then

$$|f_Q-f_R|\leqslant B_3. \tag{1.4}$$

Let us also notice that (1.3) can be replaced by

$$R|f_Q - f_R| \leqslant B_4\left(1 + \int_{2R(Q)\setminus Q} \frac{d\mu(x)}{|x-c_Q|^d} + \int_{2Q(R)\setminus R} \frac{d\mu(x)}{|x-c_R|^d}\right) \quad \text{for all } Q. \tag{1.5}$$

Proof. It is convenient to think that ϱ is a large number, and that λ is only slightly bigger than 1. Fix a cube R in our Euclidean space $\mathbf{R}^N$, fix a first integer M greater than $2/\varrho(\lambda-1)$, and divide R into M^N equal cubes Q_i. Each Q_i can be connected with R by a chain of neighbors, and the length of the chain L_i is bounded by a constant depending only on ϱ, λ: $L_i \leqslant L = L(\varrho, \lambda)$. In particular,

$$|f_{Q_i} - f_R| \leqslant B(B_3, L). \tag{1.6}$$

We know by (1.2) that

$$\int_{Q_i} |f - f_{Q_i}| \leqslant B_1 \mu(\varrho Q_i).$$

By (1.6) we can replace f_{Q_i} here by f_R:

$$\int_{Q_i} |f - f_R| \leqslant B(B_1, B_3, L) \mu(\varrho Q_i).$$

Now let us sum up all these inequalities. The cubes Q_i constitute a disjoint covering of R, the cubes ϱQ_i all lie in λR, and their multiplicity is bounded by $C(d)\varrho^{-d}$. This follows trivially from volume considerations. Thus we have

$$\int_R |f - f_R| \leqslant B(B_1, B_3, \varrho, \lambda) \mu(\lambda R),$$

and the lemma is proved. □

Our third remark: we could have changed the definition by considering only cubes centered at the support of μ. Again this does not change the space. In fact, if we are given a cube Q not centered at $K := \operatorname{supp} \mu$ and such that $2Q$ intersects K, we assign f_Q by the following rule: fix a point $x \in K \cap 2Q$, and let R be the smallest cube centered at x and $4Q \subset R$. Then put $f_Q := f_R$. The amount of ambiguity is very small, because any other $\bar{R}$ (with a different center) will have almost the same $f_{\bar{R}}$ by (1.4). If $K \cap 2Q = \varnothing$, then we put $f_Q = 0$. It is easy to see that if a function f and its collection of f_Q's satisfy (1.2), (1.3) with a certain $\varrho > 1$ and only for cubes centered at K, then, by extending the collection of f_Q's to *all* cubes as it has been done above, we obtain (1.2), (1.3) with a certain $\varrho' > 1$ which is a constant times bigger than ϱ. But Lemma 1.1 claims the independence from $\varrho > 1$. So our remark follows.

And the fourth remark is that we could have replaced cubes by balls without changing the space RBMO. This is an easy consequence of Lemma 1.1. By the way, the similar lemma is true when cubes are changed to balls (just instead of disjoint coverings we will have coverings of finite multiplicity), which means that the corresponding "ball" space also allows the change of $\varrho\in(1,\infty)$ without changing the space.

Now we are ready to formulate the main result of this section: the John–Nirenberg property of functions from RBMO. It has been proved by Tolsa in [27]. For the sake of completeness we will prove this result.

THEOREM 1.2. *Let $f\in\mathrm{RBMO}$, let $\lambda>1$, $1\leqslant p<\infty$. Then for any cube Q,*

$$\int_Q |f-f_Q|^p \leqslant B(\lambda,p,\|f\|_{\mathrm{RBMO}})(\mu(\lambda Q))^p.$$

To prove this result we will use the notion of *doubling cube* (exactly as in [27]). Fix any $\alpha>1$ and $\beta>\alpha^d$ (d is from (1.1)). A cube Q is called (α,β)-doubling (just doubling if the parameters are clear, or when they do not matter) if

$$\mu(\alpha Q)\leqslant \beta\mu(Q). \tag{1.7}$$

For a given Q we consider $Q_j:=\alpha^j Q$, $j\geqslant 0$, and the first Q_j which is (α,β)-doubling is called Q' (we omit parameters for the sake of brevity). We will use the notation Q'' for $(\alpha Q')'$. Every cube has a supercube which is (α,β)-doubling. This follows immediately from (1.1).

On the other hand, if $\beta>\alpha^N$, (N is the dimension of the ambient space), then almost every point of $K=\operatorname{supp}\mu$ has a nest of cubes centered at it and shrinking to it such that they are (α,β)-doubling.

Indeed, consider a cube Q, $l(Q)=l$, and let $M:=\mu(3Q)$. Take a point $x\in Q$, and let Q_x^r be the cube of size $\alpha^{-r}l$ centered at x.

Let us call the point x *bad*, if none of the cubes $\alpha^k Q_x^r$, $0\leqslant k\leqslant r$, is doubling. Then (since $2^r Q_x^r\subset 3Q$) $\mu(Q_x^r)\leqslant M\cdot\beta^{-r}=M\cdot(\alpha^N/\beta)^r\alpha^{-Nr}=M\cdot(\alpha^N/\beta)^r\operatorname{Vol}Q_x^r$.

Applying the Besicovitch Covering Lemma, we get that the set of all bad points x is covered by a family of cubes $Q_{x_j}^r$, $\sum_j \mu(Q_{x_j}^r)\leqslant C(\alpha^N/\beta)^r\to 0$ as $r\to\infty$. This implies that μ-almost all points x have a doubling cube of size at most l centered at x. Since this is true for arbitrary l, almost all points have a sequence of doubling cubes centered at this point and shrinking to it.

LEMMA 1.3. *Let $f\in\mathrm{RBMO}$, and let $\alpha>1$ and $\beta>\alpha^d$ be fixed arbitrarily. Then*

$$|f_Q-f_{Q'}|\leqslant C(\|f\|_{\mathrm{RBMO}},\alpha,\beta).$$

Proof. Let $Q'=Q_j:=\alpha^j Q$. Then

$$\int_{2Q'\setminus Q}\frac{d\mu(x)}{|x-c_Q|^d}\leqslant\int_{2Q'\setminus Q'}\ldots+\sum_{i=1}^{j}\int_{Q_i\setminus Q_{i-1}}\ldots\leqslant C+2^d\sum_{i=1}^{j}\frac{\mu(Q_i)}{l(Q_{i-1})^d}.$$

But $\mu(Q_i)\leqslant\beta^{i-j}\mu(Q_j)$ and $l(Q_{i-1})^{-d}\leqslant\alpha^d\alpha^{d(j-i)}l(Q_j)^{-d}$. We substitute these two inequalities into the previous one to obtain a convergent geometric progression (recall that $\alpha^d/\beta<1$). The lemma is proved. $\square$

LEMMA 1.4. *If $f\in$RBMO with fixed B_1, B_2, ϱ, then there exist numbers f_Q and positive numbers C', C'', C''' depending only on $B_1, B_2, \varrho, \alpha, \beta, d, N$ such that*

$$\frac{1}{\mu(\varrho Q)}\int_Q|f-f_Q|\leqslant C', \tag{1.8}$$

$$|f_Q-f_{Q'}|\leqslant C'', \tag{1.9}$$

$$|f_{Q_1}-f_{Q_2}|\leqslant C''' \quad \textit{for all neighbors } Q_1, Q_2. \tag{1.10}$$

There is nothing to prove—f_Q's are the numbers from the definition of RBMO, and Lemma 1.3 completes the explanation.

Notice that one could have considered (1.8), (1.9) and (1.10) as the definition of the "right" BMO-space (we will see that the John–Nirenberg property is satisfied under this definition). However, the disadvantage is that it would probably depend on two parameters: α, β. Such a space should have been called BMO(α,β) (dependence on ϱ does not exist—the analogue of Lemma 1.1 applies). Being a scale of spaces (unlike RBMO which is *one* canonical space) BMO(α,β) has the advantage that it can be described in terms of averages of our function over cubes (while RBMO involves *some* f_Q, which, as the reader will see, are often not averages at all). Here is this description:

For a function f let $\langle f\rangle_Q$ denote its average over Q, $\langle f\rangle_Q:=\mu(Q)^{-1}\int_Q f\,d\mu$.

LEMMA 1.5. *If $f\in$BMO(α,β), there exist positive numbers A', A'', A''' such that*

$$\int_Q|f-\langle f\rangle_Q|\leqslant A'\mu(\alpha Q), \tag{1.11}$$

$$|\langle f\rangle_Q-\langle f\rangle_{Q'}|\leqslant A''\frac{\mu(\alpha Q)}{\mu(Q)}, \tag{1.12}$$

$$|\langle f\rangle_{(Q_1)'}-\langle f\rangle_{(Q_2)'}|\leqslant A''' \quad \textit{for all neighbors } Q_1, Q_2. \tag{1.13}$$

Conversely, any function f satisfying (1.11), (1.12) *and* (1.13) *belongs to* BMO(α,β).

Proof. We remarked already that Lemma 1.1 holds in the setting of BMO(α,β) (the proof does not change at all). So, if f belongs to BMO(α,β), then in (1.8) we can replace ϱ by α with a cost of maybe changing a constant. Now

$$|\langle f\rangle_Q - f_Q| \leqslant C\frac{\mu(\alpha Q)}{\mu(Q)}.$$

This follows immediately from (1.8). The same for Q':

$$|\langle f\rangle_{Q'} - f_{Q'}| \leqslant C\frac{\mu(\alpha Q')}{\mu(Q')} \leqslant C(\beta).$$

Now (1.11) and (1.12) follow from these inequalities and from (1.8). To prove (1.13) we write $|\langle f\rangle_{Q'} - f_{Q'}| \leqslant C$ for $Q=Q_1, Q_2$ and compare $f_{(Q_1)'}, f_{(Q_2)'}$: $|f_{(Q_1)'} - f_{(Q_2)'}| \leqslant |f_{Q_1} - f_{(Q_1)'}| + |f_{Q_2} - f_{(Q_2)'}| + |f_{Q_1} - f_{Q_2}|$. The first two terms are bounded by (1.9). The third term is bounded by (1.10).

Conversely, let f satisfy (1.11), (1.12) and (1.13). Put $f_Q := \langle f\rangle_{Q'}$. Then

$$\int_Q |f-f_Q| \leqslant \int_Q |f-\langle f\rangle_Q| + |\langle f\rangle_Q - \langle f\rangle_{Q'}|\,\mu(Q) \leqslant C\mu(\alpha Q). \tag{1.14}$$

Also (1.10) follows immediately from (1.13) by definition. So f belongs to $\mathrm{BMO}(\alpha,\beta)$ because, as we already pointed out, the constant α on the right-hand side of (1.14) can be replaced by ϱ (by changing C). □

Remark. Let us emphasize that Lemmas 1.4 and 1.5 describe the same space. The change of $\varrho>1$ in (1.8) does not change the space, and because of that the change of α to $\alpha'\in(1,\alpha)$ on the right-hand sides of the inequalities of Lemma 1.5 does not diminish the space. It is the same $\mathrm{BMO}(\alpha,\beta)$. But the dependence on α,β probably persists, because the definition of the doubling cube Q' depends on these parameters. What we proved is that $\mathrm{RBMO}\subset\mathrm{BMO}(\alpha,\beta)$ for all parameters.

Now we are ready to prove Theorem 1.2. It follows immediately from the following lemma.

LEMMA 1.6. *Let f satisfy all assumptions of Lemma 1.4 with certain ϱ, C', C'', C'''. Then*

$$\mu\{x\in Q: |f(x)-f_Q|>t\} \leqslant D_1\mu(\alpha Q)\exp(-t/D_2), \tag{1.15}$$

where D_1, D_2 depend only on $\varrho,\alpha,\beta,N,d,C',C'',C'''$ but not on t.

Proof. Recall that $Q''=(\alpha Q')'$. Let L be a very large constant depending only on $\varrho,\alpha,\beta,m,C',C'',C'''$, which will be chosen during the proof. Find n such that $nL\leqslant t<(n+1)L$. Consider all *maximal* q' having the following properties: they are centered at $x\in Q$, $q'\subset\sqrt{\alpha}Q$ and $|f_{q'}-f_Q|>t$. We can freely change ϱ in (1.8), which implies $|\langle f\rangle_{q'}-f_{q'}|\leqslant C\mu(\alpha q')/\mu(q')\leqslant C(\beta)$. This inequality and $|f_{q'}-f_Q|>L$ imply

$$|\langle f\rangle_{q'}-f_Q|>\tfrac{1}{2}L$$

if L is large enough. In particular,

$$\int_{q'}|f-f_Q|\geqslant\tfrac{1}{2}L\mu(q'). \tag{1.16}$$

The maximality of q' implies that either $|f_{q''}-f_Q|\leqslant L$, or, if it happened that q'' is not in $\sqrt{\alpha}Q$, we can consider first $q_i:=\alpha^i q'$, which is not inside $\sqrt{\alpha}Q$. The cube q'' equals q_j for a certain j, and $j\geqslant i$. If $j=i$, then q'' has a size comparable to Q, and thus, $|f_{q''}-f_Q|\leqslant C$. If $j>i$, then still $|f_{q_i}-f_Q|\leqslant C$ because the sizes are comparable. But also $q''=(q_i)'$, and so $|f_{q''}-f_{q_i}|\leqslant C$ because of (1.9). So in all cases,

$$|f_{q''}-f_Q|\leqslant L \tag{1.17}$$

if L is large enough.

The choice of q', the inequality $|f_{q''}-f_{q'}|\leqslant C$ and (1.17) imply

$$L<|f_{q'}-f_Q|\leqslant 2L. \tag{1.18}$$

So,

$$\{x\in Q:|f(x)-f_Q|>t\}\subset\bigcup_{q'}\{x\in q':|f(x)-f_{q'}|>t-2L\geqslant(n-2)L\}, \tag{1.19}$$

where the union is taken over our maximal q' chosen above.

From our cover by q' of $\{x\in Q:|f(x)-f_Q|>t\}$ let us choose the subcover Q^i of finite multiplicity (by the theorem of Besicovitch).

Then, using (1.16), we conclude that (the C are different, but depend only on $\alpha,\beta,m,d,C',C'',C'''$)

$$\begin{aligned}\sum_i\mu(\alpha Q^i)&\leqslant C(\beta)\sum_i\mu(Q^i)\leqslant\frac{C}{L}\sum_i\int_{Q^i}|f-f_Q|\leqslant\frac{C}{L}\int_{\sqrt{\alpha}Q}|f-f_Q|\\&\leqslant\frac{C}{L}\int_{\sqrt{\alpha}Q}|f-f_{\sqrt{\alpha}Q}|+|f_Q-f_{\sqrt{\alpha}Q}|\mu(\sqrt{\alpha}Q)\\&\leqslant\frac{C}{L}\mu(\alpha Q)\leqslant\frac{1}{2}\mu(\alpha Q),\end{aligned}$$

if L is sufficiently large. The estimate before the last follows again by the fact that we can freely change $\varrho>1$ in (1.8) if $f\in\mathrm{BMO}(\alpha,\beta)$. Here we used $\varrho=\sqrt{\alpha}$.

Now we repeat our consideration for each Q^i instead of Q. By (1.9) and the last inequality we will get

$$\mu\{x\in Q:|f(x)-f_Q|>t\}\leqslant\left(\tfrac{1}{2}\right)^{n/2-1}\mu(\alpha Q),$$

which proves the lemma. □

2. Necessary conditions

2.1. How to interpret the condition $Tb\in\mathrm{BMO}^p_\lambda(\mu)$

Even if we assume that the operator T is bounded on $L^2(\mu)$, it takes some time to define what it means for $T1$ (or for Tb, $b\in L^\infty$) to belong to $\mathrm{BMO}^2_\lambda(\mu)$, since for infinite measures μ, $1\notin L^2(\mu)$, and the expression $T1$ formally is not defined for such measures. However, one can make perfect sense of the above condition, even without assuming that T is bounded.

We will need the following simple lemma, see also [22]. It means simply that if $\mu(B(x,r))\leqslant r^d$, then radially symmetric singularities (like $|x-x_0|^\alpha$) admit the same estimates as in the case of Lebesgue measure in $\mathbf{R}^d$. In particular, the singularity $|x|^{-r}$ is integrable at ∞ if $r>d$, and is integrable at 0 if $r<d$.

LEMMA 2.1 (Comparison Lemma). *Let $F\geqslant 0$ be a decreasing function on $(0,\infty)$, and let the measure μ satisfy $\mu(B(x_0,r))\leqslant r^d$ (here $d>0$) for a fixed x_0 and for all $r\geqslant 0$. Then for $\delta>0$,*

$$\int_{x:|x-x_0|\geqslant\delta} F(|x-x_0|)\,d\mu(x)\leqslant F(\delta)\delta^d+d\int_\delta^\infty F(t)t^{d-1}\,dt.$$

In particular, for $F(t)=t^{-d-\alpha}$ we have

$$\int_{x:|x-x_0|\geqslant\delta}|x-x_0|^{-d-\alpha}\,d\mu(x)\leqslant(d/\alpha+1)\delta^{-\alpha}.$$

Proof. We can assume $\lim_{t\to\infty}F(t)=0$, since otherwise we have ∞ on the right-hand side, and the lemma is trivial. Clearly

$$\begin{aligned}\int_{x:|x-x_0|\geqslant\delta} F(|x-x_0|)\,d\mu(x)&\leqslant\int_0^{F(\delta)}\mu(\{x:F(|x-x_0|\geqslant t\})\,dt\\ &\leqslant\int_0^{F(\delta)}[F^{-1}(t)]^d\,dt=-\int_\delta^\infty\tau^d\,dF(\tau)\\ &=\tau^dF(\tau)|_\delta^\infty+d\int_\delta^\infty F(\tau)\tau^{d-1}\,d\tau\\ &\leqslant F(\delta)\delta^d+d\int_\delta^\infty F(\tau)\tau^{d-1}\,d\tau.\end{aligned}$$ □

Let us suppose (for the case of the Tb-theorem) that the bilinear form $\langle Tb_1f,b_2g\rangle$ of the operator $M_{b_2}TM_{b_1}$ (M_b stands for the operator of multiplication on b) is well defined for smooth (say C^∞) compactly supported f and g. Note that the bilinear form is well defined for *arbitrary* $L^2(\mu)$-functions with separated compact supports.

Let φ be an arbitrary smooth function supported by a cube Q, satisfying $\int\varphi b_2\,d\mu=0$. Then we claim that the expression $\langle Tb_1,b_2\varphi\rangle$ is well defined.

LEMMA 2.2. *Let $\varphi=\varphi_Q$ be a function supported by the cube Q and orthogonal to constants, i.e., such that $\int_Q \varphi\,d\mu=0$. Then for x outside the cube Q,*

$$|(T\varphi_Q)(x)| \leqslant C\frac{l(Q)^\alpha}{\operatorname{dist}(x,Q)^{d+\alpha}}\cdot\|\varphi_Q\|_{L^1(\mu)}.$$

As one can see from the proof below, the lemma holds for truncated Calderón–Zygmund operators T_r as well.

Proof. Let y_0 be the center of the cube Q. If $\operatorname{dist}(x,Q)\geqslant l(Q)$, then by property (ii) of Calderón–Zygmund kernels,

$$\begin{aligned}
|T\varphi(x)| &= \left|\int K(x,y)\varphi(y)\,d\mu(y)\right| \\
&= \left|\int [K(x,y)-K(x,y_0)]\,\varphi(y)\,d\mu(y)\right| \\
&= \left|\int \frac{|y-y_0|^\alpha}{|x-y_0|^{d+\alpha}}\,\varphi(y)\,d\mu(y)\right| \\
&\leqslant C\frac{l(Q)^\alpha}{\operatorname{dist}(x,Q)^{d+\alpha}}\cdot\|\varphi\|_{L^1(\mu)}.
\end{aligned}$$

If $\operatorname{dist}(x,Q)\leqslant l(Q)$, then we have a trivial estimate using property (i) of Calderón–Zygmund kernels:

$$|T\varphi(x)| \leqslant \frac{C}{\operatorname{dist}(x,Q)^d}\cdot\|\varphi\|_{L^1(\mu)} \leqslant C\frac{l(Q)^\alpha}{\operatorname{dist}(x,Q)^{d+\alpha}}\cdot\|\varphi\|_{L^1(\mu)}. \qquad \square$$

Let ψ_1 be a smooth compactly supported function, identically equal to 1 on $2Q$, satisfying $0\leqslant\psi_1\leqslant 1$. Let $\psi_2=1-\psi_1$.

The above Lemma 2.2, applied to the function φb_2 and the operator T^*, implies

$$(T^*\varphi b_2)(x) \leqslant C\mu(Q)\,\|b_2\|_\infty \frac{l(Q)^\alpha}{\operatorname{dist}(x,Q)^{1+\alpha}}.$$

Then, the Comparison Lemma (Lemma 2.1) implies

$$\int |(T^*\varphi b_2)\psi_2 b_1|\,d\mu \leqslant CC\mu(Q)\,\|b_2\|_\infty\|b_1\|_\infty \int_{\mathbf{R}^N\setminus 2Q} \frac{l(Q)^\alpha}{\operatorname{dist}(x,Q)^{d+\alpha}}\,d\mu(x) < \infty,$$

so $\langle T\psi_2 b_1, \varphi b_2\rangle$ is well defined.

Since by the assumption $\langle T\psi_1 b_1, \varphi b_2\rangle$ is well defined (ψ_1 is a smooth compactly supported function), the expression $\langle Tb_1, \varphi b_2\rangle$ is well defined as well.

It is not difficult to show that the above expression does not depend on a choice of the function ψ_1. One can also replace the requirement $\psi_1 \equiv 1$ on $2Q$ by $\psi_1 \equiv 1$ on kQ for some $k>1$.

Now we can say that the condition $Tb_1 \in \mathrm{BMO}^2_\lambda(\mu)$ means that for any cube Q,

$$|\langle Tb_1, \varphi b_2\rangle| \leqslant C\|\varphi b_2\|_{L^2(\mu)}\mu(\lambda Q)^{1/2}$$

for any smooth function φ supported by the cube Q and satisfying $\int \varphi b_2\,d\mu=0$.

Notice that if Tb_1 is well defined, then the last condition means exactly that $Tb_1 \in \mathrm{BMO}^2_\lambda(\mu)$.

Similarly, condition $Tb_1 \in \mathrm{BMO}^p_\lambda(\mu)$ can be interpreted as

$$|\langle Tb_1, \varphi b_2\rangle| \leqslant C\|\varphi b_2\|_{L^q(\mu)}\mu(\lambda Q)^{1/p}$$

$(1/p+1/q=1)$ for all cubes Q and for all smooth functions φ supported by the cube Q and satisfying $\int \varphi b_2\,d\mu=0$.

Notice that if the bilinear form $\langle b_2 T b_1 f, g\rangle$ is defined for Lipschitz compactly supported functions, or simply for bounded compactly supported functions (as for truncated operators T_ε), we can assume that the above function φ belongs to the same class.

2.2. Necessary conditions

THEOREM 2.3. *Let a Calderón–Zygmund operator T be bounded on $L^p(\mu)$, $1<p<\infty$, and let $b\in L^\infty(\mu)$, $\|b\|_\infty \leqslant 1$. Then $Tb\in \mathrm{BMO}^p_\lambda(\mu)$, and moreover, $\|Tb\|_{\mathrm{BMO}^p_\lambda(\mu)}$ is bounded by a constant depending on the norm of T and the constants in the definition of the Calderón–Zygmund kernel.*

Proof. Take $g\in L^q(\mu)$, $1/p+1/q=1$, supported by a cube Q, and such that $\int g\,d\mu=0$. Here $g=\varphi b_2$ in terms of Lemma 2.2. Since we already know that T is bounded on $L^p(\mu)$, we do not have to worry about smoothness.

Decompose b as

$$b = b\chi_{\lambda Q} + b\cdot(1-\chi_{\lambda Q}) = b^1 + b^2.$$

It is easy to estimate

$$|\langle b^1, T^* g\rangle| \leqslant \|b\|_p \cdot \|T\|\cdot\|g\|_q \leqslant \|b\|_\infty \cdot \mu(\lambda Q)^{1/p}\cdot \|T\|\cdot\|g\|_q.$$

Let us now estimate $|\langle b^2, T^* g\rangle|$. By Lemma 2.2,

$$|T^* g(y)| \leqslant C\frac{l(Q)^\alpha}{\operatorname{dist}(y,Q)^{d+\alpha}}\cdot\|g\|_{L^1(\mu)} \leqslant C\frac{l(Q)^\alpha}{\operatorname{dist}(y,Q)^{d+\alpha}}\cdot\mu(Q)^{1/p}\|g\|_{L^q(\mu)}$$

(the last inequality is just the Hölder inequality).

Using the Comparison Lemma (Lemma 2.1) we get

$$|\langle b^2, T^*g\rangle| \leqslant C\mu(Q)^{1/p}\|g\|_{L^q(\mu)}\left|\int_{\mathbf{R}^N\setminus\lambda Q}\frac{l(Q)^\alpha}{\operatorname{dist}(y,Q)^{d+\alpha}}\,d\mu(y)\right| \leqslant C\mu(Q)^{1/p}\|g\|_q. \qquad \square$$

2.3. $Tb\in\mathrm{BMO}^1_\lambda(\mu) \Rightarrow Tb\in\mathrm{RBMO}(\mu) \Rightarrow Tb\in\mathrm{BMO}^2_\lambda(\mu)$

In this section we show that it does not matter what BMO-space to pick. We will show here that if Tb belongs to the largest possible BMO-space $\mathrm{BMO}^1_\lambda(\mu)$, then it belongs to $\mathrm{RBMO}(\mu)$ and, since the space RBMO satisfies the John–Nirenberg property, see Theorem 1.2, it belongs to the space $\mathrm{BMO}^2_\lambda(\mu)$.

Let us discuss how to interpret the condition $Tb\in\mathrm{RBMO}$. The problem is, that even if we know that the operator T is bounded on $L^p(\mu)$, Tb is not defined generally. In §2.1 we avoided this difficulty interpreting the condition $Tb\in\mathrm{BMO}^p_\lambda(\mu)$ by duality. Unfortunately, we do not know any such simple interpretation for the case of RBMO. So our interpretation will be a bit more complicated.

Namely, given a cube G, we say that a function f belongs to $\mathrm{RBMO}(G,\mu)$ if the inequalities (1.2) and (1.3) defining RBMO hold for all cubes $Q\subset R\subset G$.

It is easy to say what it means that $Tb_1\in\mathrm{RBMO}(G,\mu)$: consider a smooth compactly supported function φ, $0\leqslant\varphi\leqslant 1$, such that $\varphi(x)\equiv 1$ on the cube $10G$. Since $\langle b_2Tb_1\varphi, f\rangle$ is defined for all smooth compactly supported f, the function $Tb_1\varphi$ is well defined.

We say that Tb_1 belongs to $\mathrm{RBMO}(G,\mu)$ if $Tb_1\varphi\in\mathrm{RBMO}(G,\mu)$. It is not difficult to see that this condition does not depend on the choice of cut-off function φ.

And finally, we say that $Tb_1\in\mathrm{RBMO}(\mu)$ if Tb_1 belongs to $\mathrm{RBMO}(G,\mu)$ (with uniform estimates on the norms) for all cubes G.

Clearly, if $Tb_1\in\mathrm{RBMO}(\mu)$ then $Tb_1\in\mathrm{BMO}^p_\lambda(\mu)$ for all $p\in[1,\infty)$, $\lambda>1$; in particular, $Tb_1\in\mathrm{BMO}^2_\lambda(\mu)$.

In this section we treat the a priori bounded case, i.e., the case when the operator T is well defined on bounded compactly supported functions. One can think about truncated operators T_ε here.

THEOREM 2.4. *Let the bilinear form $\langle Tb_1f, b_2g\rangle$ be defined for bounded compactly supported f and g. Let also $b_1\in L^\infty$ and let $b_2\in L^\infty$ be a weakly accretive function.*

Suppose that $Tb_1\in\mathrm{BMO}^p_\lambda(\mu)$ for some p, $1\leqslant p<\infty$, and suppose that $M_{b_2}TM_{b_1}$ is weakly bounded, in the sense that there exist $\lambda'\geqslant 1$, $a<1$ such that

$$|\langle Tb_1\chi_Q, b_2\chi_{aQ}\rangle| \leqslant C\mu(\lambda' Q) \tag{2.1}$$

for all cubes Q.

Then $Tb_1\in\mathrm{RBMO}(\mu)$ (and therefore $Tb_1\in\mathrm{BMO}^2_\lambda(\mu)$).

LEMMA 2.5. *Under the assumptions of the previous theorem,*

$$\int_Q |Tb_1\chi_{2Q}|^p\,d\mu \leqslant C\mu(\Lambda Q),$$

where $\Lambda=\max(\lambda,\lambda')$.

Proof. First notice that if the weak boundedness condition (2.1) holds for some $a<1$, then it holds for any other value of a, probably with different C.

Fix a cube Q. Pick $g\in L^\infty$ supported by the cube Q, such that $\|g\|_q=1$, where $1/p+1/q=1$. We want to show that $\langle Tb_1\chi_{2Q}, b_2g\rangle$ is bounded. So, let us assume that $a=\frac{1}{2}$.

Pick a constant c such that

$$c\int_Q b_2\,d\mu=\int_Q b_2g\,d\mu,$$

i.e., such that $\int(b_2g-cb_2\chi_Q)\,d\mu=0$.

Since $\left|\int_Q b_2\,d\mu\right|\geqslant\delta\mu(Q)$ (b_2 is weakly accretive),

$$|c|\leqslant\delta^{-1}\mu(Q)^{-1}\int_Q|b_2g|\,d\mu\leqslant\delta^{-1}\mu(Q)^{-1}\|b_2\|_\infty\|g\|_{L^q(\mu)}\mu(Q)^{1/p}=\delta^{-1}\|b_2\|_\infty\cdot\mu(Q)^{-1/q},$$

and so $\|c\chi_Q\|_{L^q(\mu)}\leqslant C$.

Therefore $\|b_2\cdot(g-c\chi_Q)\|_{L^p(\mu)}\leqslant C+1$ and the condition $Tb_1\in\mathrm{BMO}^p_\lambda(\mu)$ implies

$$|\langle Tb_1\chi_{2Q}, b_2\cdot(g-c\chi_Q)\rangle|\leqslant C\mu(\lambda Q)\leqslant C\mu(\Lambda Q).$$

We know (weak boundedness) that

$$|\langle Tb_1\chi_{2Q}, b_2\chi_Q\rangle|\leqslant C\mu(\lambda' 2Q)\leqslant C\mu(\Lambda Q).$$

It follows that

$$|\langle Tb_1\chi_{2Q}, b_2g\rangle|\leqslant C,$$

and that is exactly what we need. □

Proof of Theorem 2.4. Let $Q\subset R\subset G$. Property (ii) of Calderón–Zygmund kernels and the Comparison Lemma 2.1 imply that for any cube $Q\subset G$ the function $\varphi:=Tb_1\chi_{10G\setminus 2Q}$ is almost constant on Q, namely

$$|\varphi(x)-\varphi(x')|\leqslant C,\quad x,x'\in Q.$$

The above Lemma 2.5 implies that for $a_Q=\varphi(c_Q)$, where c_Q is the center of the cube Q,

$$\begin{aligned}\int_Q |Tb_1\chi_{10G}-a_Q|\,d\mu &\leqslant \mu(Q)^{1/q}\biggl(\int_Q |Tb_1\chi_{10G}-a_Q|^p\,d\mu\biggr)^{1/p}\\ &\leqslant C\mu(Q)^{1/q}\mu(\Lambda Q)^{1/p}\leqslant C\mu(\Lambda Q).\end{aligned}$$

Let us compare

$$\begin{aligned}|a_Q-a_R| &= |(Tb_1\chi_{10G\setminus Q})(c_Q)-(Tb_1\chi_{10G\setminus R})(c_R)|\\ &\leqslant |(Tb_1\chi_{10G\setminus 2Q})(c_Q)-(Tb_1\chi_{10G\setminus 2R})(c_Q)|+C.\end{aligned}$$

Hence

$$|a_Q-a_R|\leqslant C+\int_{2R\setminus 2Q}|K(c_Q,y)|\,d\mu(y)\leqslant C\biggl(1+\int_{2R\setminus Q}\operatorname{dist}(y,c_Q)^{-d}\,d\mu(y)\biggr). \qquad \square$$

Now we are going to prove an analogue of Theorem 2.4 under the classical assumption of weak boundedness

$$|\langle Tb_1\chi_Q, b_2\chi_Q\rangle|\leqslant C\mu(\lambda' Q), \quad \lambda'>1, \tag{2.2}$$

for all cubes Q.

THEOREM 2.6. *Let the bilinear form $\langle Tb_1 f, b_2 g\rangle$ be defined for bounded compactly supported f and g. Let also $b_1\in L^\infty$ and let $b_2\in L^\infty$ be a weakly accretive function.*

Suppose that $Tb_1\in \mathrm{BMO}^1_\lambda(\mu)$ for some p, $1\leqslant p<\infty$, and that $M_{b_2}TM_{b_1}$ is weakly bounded, in the sense that (2.2) *holds for all cubes Q.*

Then $Tb_1\in \mathrm{RBMO}(\mu)$ (and therefore $Tb_1\in \mathrm{BMO}^2_\lambda(\mu)$).

To prove the theorem we will need the following analogue of Lemma 2.5.

LEMMA 2.7. *Under the assumptions of Theorem* 2.6,

$$\int_Q |Tb_1\chi_{2Q}|^p\,d\mu\leqslant C\mu(\Lambda Q),$$

where $\Lambda=\max(2\lambda, 2\lambda', 3)$.

If this lemma is proved, Theorem 2.6 follows immediately; one has simply to repeat the proof of Theorem 2.4.

If one tries to repeat the proof of Lemma 2.5 to prove Lemma 2.7, one would encounter a problem: at some point we need to estimate $\langle Tb_1\chi_{2Q}, b_2\chi_Q\rangle$, and we only know that $\langle Tb_1\chi_Q, b_2\chi_Q\rangle$ is bounded.

The following two lemmas below help us to cope with this problem. In these two lemmas $|\cdot|$ denotes a fixed norm in $\mathbf{R}^N$, and "ball" means the ball in this norm, $B(x_0,r):=\{x\in\mathbf{R}^N:|x-x_0|<r\}$. We will need the lemmas for the case when the norm $|\cdot|$ is the l^∞-norm, $|x|=\max\{|x_k|:1\leqslant k\leqslant N\}$, so the "balls" are cubes.

LEMMA 2.8. *Let $B(x_0,R)$ be a ball. There exists R_0, $R\leqslant R_0\leqslant 1.2R$, such that for all $s\in[0,1.5]$,*

$$\mu(\{x: R_0-Rs<|x-x_0|<R_0+Rs\})\leqslant Cs\mu(B(x_0,3R)).$$

Proof. Define the measure ν on $[0,3R)$ as the radial projection of the measure $\mu|B(x_0,3R)$,

$$\nu([0,t)) := \mu(B(x_0,t)),\quad 0\leqslant t\leqslant 3R.$$

Consider the centered maximal operator M, $M\nu(x):=\sup_{s>0}\nu((x-s,x+s))/2s$. It is well known that M is of weak type $(1,1)$, i.e., that

$$\operatorname{meas}_1\{x: M\nu(x)>\lambda\}\leqslant \frac{A}{\lambda}\nu[0,3R),\quad \lambda>0,$$

where meas_1 is one-dimensional Lebesgue measure on $\mathbf{R}$, and A is some absolute constant. Therefore

$$M\nu(x)>\frac{10A\mu(B(x_0,3R))}{R}$$

on a set of length at most $0.1R$. Therefore for some $R_0\in[R,1.2R]$ the inequality $M\nu(R_0)\leqslant 10A\mu(B(x_0,3R))/R$ holds. That implies the conclusion of the lemma. □

LEMMA 2.9. *Let R_0 be as above in Lemma 2.8, and let K be a Calderón–Zygmund kernel. Then*

$$\begin{aligned}\iint_{B(x_0,R_0)\times[B(x_0,3R)\setminus B(x_0,R_0)]}|K(x,y)|\,d\mu(x)\,d\mu(y)&\leqslant C\sqrt{\mu(B(x_0,R_0))}\,\sqrt{\mu(B(x_0,3R))}\\&\leqslant C\mu(B(x_0,3R)).\end{aligned}$$

Note that the lemma is not true for arbitrary R_0. We use the fact that the measure behaves regularly, as it is described in Lemma 2.8, in a neighborhood of the sphere $S_{R_0}:=\{x:|x-x_0|=R_0\}$.

Proof of Lemma 2.9. Consider

$$f(x):=\int_{B(x_0,3R)\setminus B(x_0,R_0)}|K(x,y)|\,d\mu(y).$$

Let $x\in B(x_0,R_0)$ and let $\delta:=\operatorname{dist}(x,S_{R_0})$, where $S_{R_0}:=\{x:|x-x_0|=R_0\}$. Clearly

$$f(x)\leqslant\int_{\delta\leqslant|y-x|\leqslant 5R}\frac{1}{|y-x|^d}\,d\mu(y),$$

and the Comparison Lemma (Lemma 2.1) implies

$$f(x) \leqslant 1+\int_{\delta}^{5R} \frac{dt}{t} \leqslant C \log \frac{R}{\delta} = C \log \frac{R}{\operatorname{dist}(x, S_{R_0})}.$$

The Cauchy–Schwarz inequality implies

$$\int_{B(x_0,R_0)} f(x)\, d\mu(x) \leqslant C\mu(B(x_0,R_0))^{1/2} \left(\int_{B(x_0,R_0)} \log^2 \frac{R}{\operatorname{dist}(x, S_{R_0})}\, d\mu(x) \right)^{1/2}.$$

Since the measure of the strip $\{x\in B(x_0,R_0) : \operatorname{dist}(x,S_{R_0})<\tau\}$ is at most

$$C\frac{\tau}{R}\cdot \mu(B(x_0,3R)),$$

see Lemma 2.8, we get

$$\int_{B(x_0,R_0)} \log^2 \frac{R}{\operatorname{dist}(x,S_{R_0})}\, d\mu(x) \leqslant C\mu(B(x_0,3R))\frac{1}{R}\int_0^{R_0} \log^2 \frac{R}{\tau}\, d\tau \leqslant C'\mu(B(x_0,3R)).$$

We are done. □

Proof of Lemma 2.7. Let $|\cdot|$ denote the l^∞-norm on $\mathbf{R}^N$, $|x|=\max\{|x_k| : 1\leqslant k\leqslant N\}$, so a cube Q is just a ball in this norm, $Q=B(x_0,R)=\{x\in\mathbf{R}^N : |x-x_0|<R\}$. Let $Q'=B(x_0,R_0)$ be the cube (ball) from Lemma 2.8 above.

By Lemma 2.9,

$$|\langle Tb_1\chi_{2Q\setminus Q'}, b_2\chi_{Q'}\rangle| \leqslant C\mu(3Q) \leqslant C\mu(\Lambda Q),$$

so, since $|\langle Tb_1\chi_{Q'}, b_2\chi_{Q'}\rangle| \leqslant C\mu(\lambda' Q') \leqslant C\mu(\Lambda Q)$, we have

$$|\langle Tb_1\chi_{2Q}, b_2\chi_{Q'}\rangle| \leqslant C\mu(3Q) \leqslant C\mu(\Lambda Q). \tag{2.3}$$

The rest of the proof goes exactly the same way as the proof of Lemma 2.5: take a bounded function g supported by the cube Q, and pick a number c such that

$$c\int_{Q'} b_2\, d\mu = \int_Q g b_2\, d\mu.$$

As in Lemma 2.5, $\|cb_2\chi_{Q'}\|_{L^p(\mu)} \leqslant C$. The condition $Tb_1 \in \mathrm{BMO}^p_\lambda(\mu)$ implies that

$$|\langle Tb_1\chi_{2Q}, b_2\cdot(g-c\chi_{Q'})\rangle| \leqslant C\mu(\lambda Q') \leqslant C\mu(\Lambda Q),$$

and together with (2.3) this implies $|\langle Tb_1\chi_{2Q}, b_2 g\rangle| \leqslant C\mu(\Lambda Q)$. □

3. An embedding theorem

As people familiar with proofs of classical $T1$- or Tb-theorems can remember, the Carleson Embedding Theorem plays an important role there.

Here we present and prove a version of the theorem we need. We will use Theorem 3.1 below only with $p=2$. In this case it is just the classical Carleson Embedding Theorem, and any known proof (with obvious modifications) would work.

We think that this theorem is of independent interest, so we will present the proof of the general case.

Let $\mathcal{D}$ be a collection of dyadic cubes in $\mathbf{R}^N$. Let $\{a_Q\}_{Q\in\mathcal{D}}$ be a collection of non-negative numbers, and let $\mathbf{f}_Q$ be the average, $\mathbf{f}_Q:=\mu(Q)^{-1}\int_Q f\,d\mu$. Consider a (non-linear) operator S defined on, say, locally μ-integrable functions by

$$Sf(x):=\biggl(\sum_{Q\in\mathcal{D}} a_Q\mathbf{f}_Q^2\chi_Q(x)\biggr)^{1/2}.$$

We are interested in the question of when this operator is bounded on $L^p(\mu)$, i.e., when $\|Sf\|_{L^p(\mu)}\leqslant C\|f\|_{L^p(\mu)}$ for all $f\in L^p(\mu)$.

THEOREM 3.1. *The following statements are equivalent*:

(i) *the operator S is bounded on $L^p(\mu)$*;

(ii) $\displaystyle\sup_{Q\in\mathcal{D}}\frac{1}{\mu(Q)}\int_Q\biggl(\sum_{R\subset Q}a_R\chi_R(x)\biggr)^{p/2}d\mu(x)=C<\infty$;

(iii) *the family $\{a_Q\}_{Q\in\mathcal{D}}$ satisfies the following "Carleson measure condition"*:

$$\sup_{Q\in\mathcal{D}}\frac{1}{\mu(Q)}\sum_{R\subset Q}a_R\mu(R)=C_1<\infty.$$

Moreover, the constants $C^{2/p}$, C_1 and $\|S\|^2$ are equivalent in the sense of two-sided estimates with absolute constants.

When $p=2$, condition (i) means that $\sum_{Q\in\mathcal{D}}a_Q\mathbf{f}_Q^2\mu(Q)\leqslant C\|f\|^2_{L^2(\mu)}$, and the theorem is simply a dyadic version of the famous Carleson Embedding Theorem. For $p\neq2$ the theorem can be interpreted as a result about embedding an L^p-space into a weighted Triebel–Lizorkin space.

Proof of Theorem 3.1. (i) $\Rightarrow$ (ii). Take $f=\chi_Q$. Then

$$\int_Q\biggl(\sum_{R\subset Q}a_R\chi_R(x)\biggr)^{p/2}d\mu(x)\leqslant\|S\chi_Q\|^p_{L^p(\mu)}\leqslant\|S\|^p\cdot\|\chi_Q\|^p_{L^p(\mu)}=\|S\|^p\cdot\mu(Q),$$

i.e., condition (ii) holds with $C=\|S\|^p$.

(ii) $\Rightarrow$ (iii). If $p\geqslant 2$, the Hölder inequality implies

$$\frac{1}{\mu(Q)}\sum_{R\subset Q} a_R\cdot\mu(R)=\frac{1}{\mu(Q)}\int_Q\sum_{R\subset Q} a_R\chi_R(x)\,d\mu(x)$$
$$\leqslant\left(\frac{1}{\mu(Q)}\int_Q\left(\sum_{R\subset Q} a_R\chi_R(x)\right)^{p/2}d\mu(x)\right)^{2/p}\leqslant C^{2/p}.$$

Let us now consider the case $p<2$. First notice that in this case the inequality

$$X^{p/2}-(X-\Delta X)^{p/2}\geqslant \tfrac12 pX^{p/2-1}\Delta X \tag{3.1}$$

holds for $X, X-\Delta X>0$.

For a cube Q let us define the function $\varphi_Q(x):=\sum_{R\subset Q} a_R\chi_R(x)$.

Let Q_k, $1\leqslant k\leqslant 2^N$, be the cubes of size $\frac12 l(Q)$ contained in Q. Notice that for $x\in Q_k$ we have $\varphi_{Q_k}(x)=\varphi_Q(x)-a_Q$, so the inequality (3.1) (with $X=\varphi_Q$, $\Delta X=a_Q$) implies

$$\varphi_Q^{p/2}(x)-\varphi_{Q_k}^{p/2}(x)\geqslant\tfrac12 p\varphi_Q^{p/2-1}(x)\cdot a_Q\quad\text{for } x\in Q_k.$$

Integrating over Q_k and summing up over k we get

$$\frac{1}{\mu(Q)}\int_Q\varphi_Q^{p/2}\,d\mu\geqslant\frac{p}{2}a_Q\frac{1}{\mu(Q)}\int_Q\varphi_Q^{p/2-1}\,d\mu+\sum_{k=1}^{2^N}\frac{1}{\mu(Q)}\int_{Q_k}\varphi_{Q_k}^{p/2}\,d\mu. \tag{3.2}$$

Let us notice that $(1/\mu(Q))\int_Q\varphi_Q^{p/2-1}\,d\mu$ is bounded below. Indeed

$$1\leqslant\frac{1}{\mu(Q)}\int_Q\varphi_Q^{1-p/2}\,d\mu\cdot\frac{1}{\mu(Q)}\int_Q\varphi_Q^{p/2-1}\,d\mu.$$

On the other hand, Hölder's inequality implies

$$\frac{1}{\mu(Q)}\int_Q\varphi_Q^{1-p/2}\,d\mu\leqslant\left(\frac{1}{\mu(Q)}\int_Q\varphi_Q^{p/2}\,d\mu\right)^{2/p-1}\leqslant C^{2/p-1},$$

and so

$$\frac{1}{\mu(Q)}\int_Q\varphi_Q^{p/2-1}\,d\mu\geqslant C^{1-2/p}.$$

Therefore (3.2) implies

$$a_Q\leqslant\frac{2}{p}C^{p/2-1}\left(\frac{1}{\mu(Q)}\int_Q\varphi_Q^{p/2}\,d\mu-\sum_{k=1}^{2^N}\frac{1}{\mu(Q)}\int_{Q_k}\varphi_{Q_k}^{p/2}\,d\mu\right).$$

Writing such inequalities for all dyadic cubes $R\subset Q$, multiplying them by $\mu(R)$ and summing them up, we get

$$\sum_{R\subset Q} a_R\mu(R) \leqslant \frac{2}{p} C^{p/2-1}\int_Q \varphi_Q^{p/2}\,d\mu \leqslant \frac{2}{p} C^{p/2-1}\mu(Q)C = \frac{2}{p} C^{p/2}\mu(Q),$$

which is exactly condition (iii).

(iii) $\Rightarrow$ (i). To prove the implication, we use the Bellman function method. What a Bellman function is and how to find it is discussed in great detail in [20], so here our presentation will be very sketchy.

Clearly, it is enough to consider only $f\geqslant 0$.

For a dyadic cube Q consider the averages

$$\mathbf{F}_Q := \mu(Q)^{-1}\int_Q f^p\,d\mu, \qquad \mathbf{f}_Q := \mu(Q)^{-1}\int_Q f\,d\mu, \tag{3.3}$$

$$\mathbf{A}_Q := \mu(Q)^{-1}\sum_{R\subset Q} a_R|R|, \quad \mathbf{c}_Q := \Bigl(\sum_{R:R\supsetneq Q} a_R\mathbf{f}_R^2\Bigr)^{1/2}. \tag{3.4}$$

Our goal is to construct a function $\mathcal{B}=\mathcal{B}(\mathbf{f},\mathbf{F},\mathbf{c},\mathbf{A})$ of four real variables. We want the function to be defined on the set

$$0\leqslant \mathbf{f}\leqslant \mathbf{F}^{1/p}, \quad 0\leqslant \mathbf{A}\leqslant 1, \quad \mathbf{c}\geqslant 0.$$

We want it to satisfy

$$\gamma\mathbf{c}^p \leqslant \mathcal{B}(\mathbf{f},\mathbf{F},\mathbf{c},\mathbf{A}) \leqslant \Gamma\cdot(\mathbf{F}+\mathbf{c}^p) \tag{3.5}$$

where $\Gamma\geqslant\gamma\geqslant 0$ are some constants. We also want it to satisfy

$$\mathcal{B}(\mathbf{f},\mathbf{F},\mathbf{c},\mathbf{A}) \geqslant \tfrac{1}{2}(\mathcal{B}(\mathbf{f}_1,\mathbf{F}_1,\mathbf{c}_1,\mathbf{A}_1)+\mathcal{B}(\mathbf{f}_2,\mathbf{F}_2,\mathbf{c}_2,\mathbf{A}_2)) \tag{3.6}$$

for any three sets of arguments satisfying

$$\begin{aligned} \mathbf{F} &= \tfrac{1}{2}(\mathbf{F}_1+\mathbf{F}_2), & \mathbf{f} &= \tfrac{1}{2}(\mathbf{f}_1+\mathbf{f}_2),\\ \mathbf{A} &= \tfrac{1}{2}(\mathbf{A}_1+\mathbf{A}_2)+a, & \mathbf{c}_1 &= \mathbf{c}_2 = \sqrt{\mathbf{c}^2+a\mathbf{f}^2}\,. \end{aligned}$$

If we construct such a function $\mathcal{B}$, we are done!

To show this, let us first notice that if a function $\mathcal{B}$ satisfies (3.6), then

$$\mathcal{B}(\mathbf{f},\mathbf{F},\mathbf{c},\mathbf{A}) \geqslant \sum_{k=1}^{M}\mu_k\mathcal{B}(\mathbf{f}_k,\mathbf{F}_k,\mathbf{c}_k,\mathbf{A}_k) \tag{3.7}$$

for any $\mu_k \geqslant 0$ such that $\sum_k \mu_k = 1$, and any $M+1$ sets of variables satisfying

$$\mathbf{F} = \sum_{k=1}^{M} \mu_k \mathbf{F}_k, \qquad \mathbf{f} = \sum_{k=1}^{M} \mu_k \mathbf{f}_k, \tag{3.8}$$

$$\mathbf{A} = \sum_{k=1}^{M} \mu_k \mathbf{A}_k + a, \quad \mathbf{c}_1 = \mathbf{c}_2 = \ldots = \mathbf{c}_M = \sqrt{\mathbf{c}^2 + a\mathbf{f}^2}. \tag{3.9}$$

Suppose that we are given a family $\{a_R\}_{R\in\mathcal{D}}$. Without loss of generality we can always assume that its Carleson constant is 1, i.e., that

$$\mu(Q)^{-1} \sum_{R\subset Q} a_R \mu(R) \leqslant 1 \quad \text{for all } Q \in \mathcal{D}.$$

Clearly, it is enough to prove the implication (iii) $\Rightarrow$ (i) for finite families, so we assume that only finitely many a_R are non-zero and that $a_R = 0$ for $R \not\subset Q$.

Fix this cube Q, and let Q_k^n, $k=1,2,\ldots,2^{Nn}$, be the cubes of size $2^{-n}l(Q)$ contained in Q. Pick a non-negative function f in $L^p(\mu)$. Condition (3.7) implies that

$$\mathcal{B}(\mathbf{f}_Q, \mathbf{F}_Q, \mathbf{c}_Q, \mathbf{A}_Q) \geqslant \sum_{k=1}^{2^N} \mu_k \mathcal{B}(\mathbf{f}_{Q_k^1}, \mathbf{F}_{Q_k^1}, \mathbf{c}_{Q_k^1}, \mathbf{A}_{Q_k^1}),$$

where $\mathbf{f}_Q, \mathbf{F}_Q, \mathbf{c}_Q, \mathbf{A}_Q$ are the averages defined above in (3.3), (3.4), and $\mu_k := \mu(Q_k)/\mu(Q)$. Notice that the averages satisfy (3.8), (3.9) with $\mathbf{f}_k = \mathbf{f}_{Q_k^1}, \ldots$, and $a = a_Q$.

Let us apply this inequality for each cube Q_k^1, then for each cube Q_k^2, etc. Going n generations down we get

$$\mathcal{B}(\mathbf{f}_Q, \mathbf{F}_Q, \mathbf{c}_Q, \mathbf{A}_Q) \geqslant \sum_{k=1}^{2^{Nn}} \frac{\mu(Q_k^n)}{\mu(Q)} \mathcal{B}(\mathbf{f}_{Q_k^n}, \mathbf{F}_{Q_k^n}, \mathbf{c}_{Q_k^n}, \mathbf{A}_{Q_k^n}).$$

The inequality (3.5) implies

$$\gamma \frac{1}{\mu(Q)} \int \sum_{k=1}^{2^{Nn}} (\mathbf{c}_{Q_k^n})^p \chi_{Q_k^n}(x)\, d\mu(x) = \gamma \sum_{k=1}^{2^{Nn}} \frac{\mu(Q_k^n)}{\mu(Q)} \cdot (\mathbf{c}_{Q_k^n})^p \leqslant \Gamma \cdot \mathbf{F}_Q$$

($\mathbf{c}_Q = 0$ since $a_R = 0$ for $R \not\subset Q$). Since the family $\{a_R\}_{R\in\mathcal{D}}$ is finite, for sufficiently large n the function $\sum_{k=1}^{2^{Nn}} (\mathbf{c}_{Q_k^n})^p \chi_{Q_k^n}$ coincides with $|Sf|^p$. So we get

$$\frac{\gamma}{\mu(Q)} \int |Sf|^p \, d\mu \leqslant \frac{\Gamma}{\mu(Q)} \int_Q |f|^p \, d\mu,$$

which is exactly what we need.

So, to complete the proof we need to present a Bellman function $\mathcal{B}$. Here is one of the possible choices:

$$\mathcal{B}(\mathbf{f},\mathbf{F},\mathbf{c},\mathbf{A}) := K\mathbf{F} - \frac{\mathbf{f}^{1+\varepsilon}\mathbf{c}^{p-\varepsilon-1}}{(1+\mathbf{A})^{\varepsilon}} + 2\gamma\mathbf{c}^{p},$$

where $K>0$ is large and $\varepsilon>0$ is small, such that $p-\varepsilon>1$. The function $\mathcal{B}$ satisfies estimates (3.5): the upper estimate is trivial, and the lower one hold for sufficiently large K (it follows from Young's inequality $ab\leqslant a^p/p+b^{p'}/p'$ with appropriate p).

Let us show that (3.6) holds. Since the function $\mathbf{f}^{1+\varepsilon}/(1+\mathbf{A})^{\varepsilon}$ is convex, it is enough to check that the term

$$\frac{\mathbf{f}^{1+\varepsilon}\mathbf{c}^{p-\varepsilon-1}}{(1+\mathbf{A})^{\varepsilon}}$$

increases more than $\gamma\mathbf{c}^p$ when one replaces $\mathbf{c}\mapsto\mathbf{c}'=\sqrt{\mathbf{c}^2+a\mathbf{f}^2}$, $\mathbf{A}\mapsto\mathbf{A}-a$.

Notice that for any $\alpha>0$,

$$C_1(\mathbf{c}')^{\alpha-2}a\mathbf{f}^2 \leqslant (\mathbf{c}')^{\alpha}-\mathbf{c}^{\alpha} \leqslant C_2(\mathbf{c}')^{\alpha-2}a\mathbf{f}^2.$$

Therefore, all we need to show is the inequality

$$\underbrace{\mathbf{f}^{1+\varepsilon}(\mathbf{c}')^{p-\varepsilon-3}a\mathbf{f}^2}_{\text{increase } \mathbf{c} \text{ first}} + \underbrace{\mathbf{f}^{1+\varepsilon}(\mathbf{c}')^{p-1-\varepsilon}a}_{\text{decrease } \mathbf{A} \text{ then}} \geqslant \underbrace{\gamma'(\mathbf{c}')^{p-2}a\mathbf{f}^2}_{\text{increment of } \gamma\mathbf{c}^p}.$$

This inequality follows immediately from Young's inequality

$$xy \leqslant \frac{x^r}{r}+\frac{y^{r'}}{r'}$$

with $r=2/(1-\varepsilon)$ and $x=\mathbf{f}^{(3+\varepsilon)(1-\varepsilon)/2}\mathbf{c}'^{(p-3-\varepsilon)(1-\varepsilon)/2}$.

There is also a simple way to see, without computations, that Young's inequality with some r would work. First, notice that the sum of exponents of $\mathbf{f}$ and $\mathbf{c}'$ is p for each term. Then, compare exponents, say of f, of each term:

$$1+\varepsilon < 2 < 3+\varepsilon. \qquad \square$$

4. Martingale difference decomposition

Fix a dyadic lattice $\mathcal{D}$ in $\mathbf{R}^N$. Just for our convenience we will consider only lattices constructed of cubes with sides 2^k, $k\in\mathbf{Z}$ (we consider cubes of all sizes, not only with a fixed k).

Denote by E_k the averaging operator over dyadic cubes of size (length of the side) 2^k, namely $E_kf(x)=\mu(Q)^{-1}\int_Q f\,d\mu$, where Q is a dyadic cube of size 2^k containing x (for the sake of definiteness, we consider cubes of the form $x_0+[a,b)^N$). If Q is a cube of size 2^k, we denote by E_Qf the restriction of E_kf to Q: $E_Qf=(\mu(Q)^{-1}\int_Q f\,d\mu)\chi_Q=\chi_Q E_kf$.

Let $\Delta_k:=E_{k-1}-E_k$. Again for a dyadic cube Q of size 2^k, denote by Δ_Qf the restriction of Δ_kf to Q. Clearly, for any $f\in L^2(\mu)$, the functions Δ_Qf, $Q\in\mathcal{D}$, are orthogonal to each other, and for any fixed n,

$$f=\sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}}\Delta_Qf+\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}}E_Qf,$$

$$\|f\|^2_{L^2(\mu)}=\sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}}\|\Delta_Qf\|^2+\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}}\|E_Qf\|^2.$$

For the Tb-theorem we need a weighted version of the above decomposition. Namely, let b be a weakly accretive function. Define

$$E_k^bf(x):=\left(\int_Q b\,d\mu\right)^{-1}\cdot\left(\int_Q f\,d\mu\right)\cdot b(x),$$

where Q is the dyadic cube of size 2^k containing x. Again for a cube Q of size 2^k let E_Q^b denote the restriction of E_k^b to Q. Similarly to the non-weighted case define operators $\Delta_k^b:=E_{k-1}^b-E_k^b$, and for a dyadic cube Q of size 2^k, let the symbol Δ_Q^bf denote the restriction of Δ_k^bf to Q.

Notice that all operators E_k^b, E_Q^b, Δ_Q^b, Δ_k^b are (generally non-orthogonal) projections. Notice also that for any $f\in L^2(\mu)$ the function Δ_Q^bf is always orthogonal to constants, i.e., $\int\Delta_Q^bf\,d\mu=0$.

Similarly to the non-weighted case, for any $f\in L^2(\mu)$ one can write down a decomposition

$$f=\sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}}\Delta_Q^bf+\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}}E_Q^bf$$

(we discuss the convergence a bit later). Unfortunately, the terms in this decomposition are not orthogonal, so we cannot get such a nice formula for the norm $\|f\|_{L^2(\mu)}$ as in the non-weighted case. Fortunately, the system of subspaces $\{\operatorname{Range}\Delta_Q^b: l(Q)\leqslant 2^n\}$, $\{\operatorname{Range}E_Q^b: l(Q)=2^n\}$ forms an *unconditional basis* in $L^2(\mu)$, i.e., the following lemma holds.

LEMMA 4.1. *Let b be a weakly accretive function, and let $n \in \mathbf{Z}$. Then, any $f \in L^2(\mu)$ can be decomposed as*

$$f = \sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}} \Delta_Q^b f + \sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}} E_Q^b f,$$

where the series converges in $L^2(\mu)$. Moreover,

$$A^{-1}\|f\|^2_{L^2(\mu)} \leqslant \sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}} \|\Delta_Q^b f\|^2_{L^2(\mu)} + \sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}} \|E_Q^b f\|^2_{L^2(\mu)} \leqslant A\|f\|^2_{L^2(\mu)},$$

where the constant $A=A(b)$ depends only on b (more precisely on $\|b\|_\infty$ and the constant δ in the definition of weak accretivity).

Proof. If $f=\sum_{Q\in\mathcal{D},\, l(Q)=2^{-k}} c_Q \chi_Q \cdot b$ (the sum is finite), then the decomposition converges, because the sum contains only finitely many terms. So, the decomposition converges on a dense subset of $L^2(\mu)$, and to prove the lemma we only need to prove the estimates.

Let us first prove the estimate from above. Notice that the estimate for the second sum is trivial, so to prove the estimate it is enough to show that

$$\sum_{Q\in\mathcal{D}} \|\Delta_Q^b f\|^2_{L^2(\mu)} \leqslant C\|f\|^2_{L^2(\mu)}, \tag{4.1}$$

or equivalently,

$$\sum_k \|\Delta_k^b f\|^2_{L^2(\mu)} \leqslant C\|f\|^2_{L^2(\mu)}.$$

Notice that

$$\begin{aligned}\Delta_k^b f &= E_{k-1}^b f - E_k^b f = [(E_{k-1}b)^{-1}\cdot E_{k-1}f - (E_k b)^{-1} E_k f]\cdot b\\ &= (E_{k-1}b)^{-1}\cdot[E_{k-1}f - E_k f]\cdot b + E_k f\cdot[(E_{k-1}b)^{-1} - (E_k b)^{-1}]\cdot b\\ &= (E_{k-1}b)^{-1}\Delta_k f\cdot b - E_k f\cdot \frac{\Delta_k b}{E_k b\cdot E_{k-1}b}\cdot b.\end{aligned}$$

Since $b\in L^\infty$, and since b is weakly accretive,

$$\sum_k \|(E_{k-1}b)^{-1}\Delta_k f\cdot b\|^2 \leqslant \delta^{-2}\|b\|^2_\infty \cdot \|f\|^2_{L^2(\mu)}.$$

To estimate the second sum, notice that according to Lemma 4.2 below, the family $a_Q := \mu(Q)^{-1}\cdot \|\Delta_Q b\|^2_{L^2(\mu)}$, $Q\in\mathcal{D}$, satisfies the Carleson measure condition (iii) from Theorem 3.1 above. Therefore Theorem 3.1 (for $p=2$) implies

$$\sum_k \|E_k f\|^2_{L^2(\mu)}\cdot \|\Delta_k b\|^2_{L^2(\mu)} = \sum_{Q\in\mathcal{D}} \|E_Q f\|^2_{L^2(\mu)}\cdot \|\Delta_Q b\|^2_{L^2(\mu)} \leqslant C\|f\|^2_{L^2(\mu)},$$

and we are done with the estimate from above.

Notice that for $p=2$ Theorem 3.1 is well known: essentially it is a dyadic version of the famous Carleson Embedding Theorem. One of the possible proofs can be found in [21], see the proof of Theorem 3.1 there.

The estimate from below follows from a standard duality argument. First of all notice that

$$(E_k b)^* f = (E_k b)^{-1} \cdot E_k(bf) = b^{-1} E_k^b(bf),$$

and so $(\Delta_k^b)^* f = b^{-1}\Delta_k^b(bf)$ (here we use *bilinear* duality $\langle f,g\rangle = \int fg\,d\mu$). Since $b, b^{-1} \in L^\infty$, it follows from (4.1) that for any $f\in L^2(\mu)$,

$$\sum_k \|(\Delta_k^b)^* f\|^2_{L^2(\mu)} = \sum_k \|b^{-1}\Delta_k^b(bf)\|^2_{L^2(\mu)} \leqslant C\|f\|^2_{L^2(\mu)}.$$

Take

$$f = E_n^b f + \sum_{k\leqslant n} \Delta_k^b f$$

(to avoid complications with the convergence, assume that the sum contains only finitely many terms). Since $\Delta_k^b E_n^b = 0$, $\Delta_k^b \Delta_l^b = 0$ for $k<n$ and $l\neq k$, we have

$$\begin{aligned}\|f\|^2_{L^2(\mu)} &= \langle f, \bar f\rangle = \langle E_n^b f, (E_n^b)^* \bar f\rangle + \sum_{k\leqslant n} \langle \Delta_k^b f, (\Delta_k^b)^* \bar f\rangle \\ &\leqslant \biggl(\|E_n^b f\|^2_{L^2(\mu)} + \sum_k \|\Delta_k^b f\|^2_{L^2(\mu)}\biggr)^{1/2} \biggl(\|(E_n^b)^* \bar f\|^2_{L^2(\mu)} + \sum_k \|(\Delta_k^b)^* \bar f\|^2_{L^2(\mu)}\biggr)^{1/2}.\end{aligned}$$

The second factor is bounded from above by $C\|f\|_{L^2(\mu)}$, so the first one is bounded from below.

Since the estimate from below holds for all f in a dense set, it holds for all $f\in L^2(\mu)$. □

Now Lemma 4.1 is proved modulo the following simple lemma.

LEMMA 4.2. *Let $f\in L^\infty$. Define $a_Q := \mu(Q)^{-1} \cdot \|\Delta_Q b\|^2_{L^2(\mu)}$, $Q\in\mathcal{D}$. Then the family $\{a_Q\}_{Q\in\mathcal{D}}$ satisfies the Carleson measure condition*

$$\sum_{R\subset Q} a_R \mu(R) \leqslant C\mu(Q) \quad \textit{for all } Q\in\mathcal{D}.$$

Proof.

$$\sum_{R\subset Q} \|\Delta_R b\|^2_{L^2(\mu)} \leqslant \int_Q |b|^2\,d\mu \leqslant \|b\|^2_\infty \cdot \mu(Q).$$ □

5. $\mathrm{BMO}^2_\lambda(\mu)$ and a Carleson measure condition

If the measure is doubling, a function in BMO can be characterized in terms of a Carleson measure condition on its Haar coefficients.

For general measures some characterization of this type is given in the lemma below.

For technical reasons, in what follows, it is convenient for us to consider two different dyadic lattices, say $\mathcal{D}$ and $\mathcal{D}'$. Suppose that the sides of the cubes in both lattices are exactly 2^{-k}, $k\in\mathbf{Z}$, and that the lattices are shifted with respect to one another.

Fix r large enough so that $2^r\geqslant 4\lambda$. For a function φ and a dyadic cube $Q\in\mathcal{D}$ define

$$a_Q = a_Q^b(\varphi) = \sum_{\substack{Q'\in\mathcal{D}':\, l(Q')=2^{-r}l(Q)\\ \operatorname{dist}(Q',\partial Q)\geqslant\lambda l(Q')}} \|\Delta^b_{Q'}\varphi\|^2_{L^2(\mu)}.$$

Notice that $Q\in\mathcal{D}$, and the smaller cubes Q' are taken from another dyadic lattice $\mathcal{D}'$.

LEMMA 5.1. *Let b be a weakly accretive function. If $\varphi\in\mathrm{BMO}^2_\lambda(\mu)$, then for any $n>1$ the family $\{a^b_Q(\varphi)\}_{Q\in\mathcal{D}}$ defined above satisfies the Carleson measure condition*

$$\sum_{R\subset Q} a_R \leqslant C\mu(Q) \quad \textit{for all } Q\in\mathcal{D}.$$

Proof. It is sufficient to prove that for any dyadic cube $Q\in\mathcal{D}$,

$$\sum_{\substack{Q'\in\mathcal{D}':\, Q'\subset Q\\ l(Q')\leqslant 2^{-r}l(Q)\\ \operatorname{dist}(Q',\partial Q)\geqslant\lambda l(Q')}} \|\Delta^b_{Q'}\varphi\|^2_{L^2(\mu)} \leqslant C\mu(Q) \tag{5.1}$$

(all terms in the sum we want to estimate are contained in the above sum).

Consider the following Whitney-type covering of the cube Q by cubes $R\subset\mathcal{D}'$: Take all cubes $R\subset Q$ of size $2^{-r}l(Q)$ such that $\operatorname{dist}(R,\partial Q)\geqslant\lambda l(R)$ (the assumption $2^r\geqslant 4\lambda$ guarantees that there exists at least one such R), then take the layer around them consisting of all cubes of size $2^{-r-1}l(Q)$ such that $\operatorname{dist}(R,\partial Q)\geqslant\lambda l(R)$, then the layer of cubes of size 2^{-r-2}, etc., see Figure 5. Let us call the collection of such Whitney cubes $\mathcal{W}$.

Pick a cube $R\in\mathcal{W}$. By the definition of $\mathrm{BMO}^2_\lambda(\mu)$,

$$\int_R |\varphi-\varphi_R|^2\,d\mu \leqslant C\mu(\lambda R).$$

Lemma 4.1 implies

$$\sum_{\substack{Q'\in\mathcal{D}'\\ Q'\subset R}} \|\Delta^b_{Q'}\varphi\|^2_{L^2(\mu)} \leqslant C\mu(\lambda R). \tag{5.2}$$

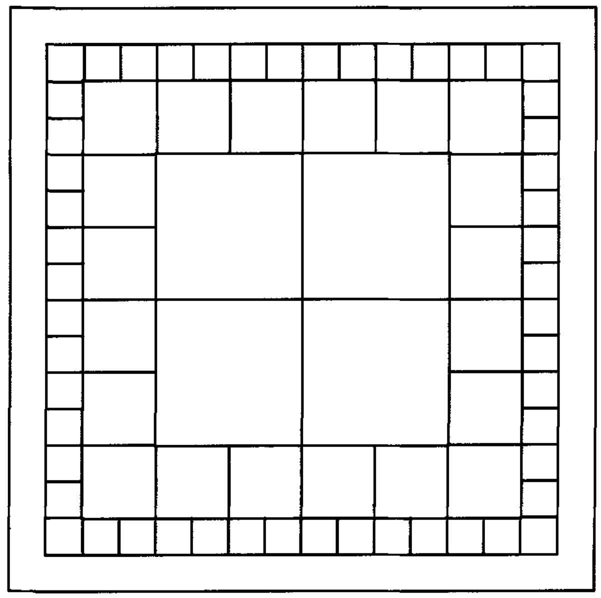

Fig. 5. Whitney type decomposition of the cube Q (here $N=2$, so cubes are squares). There are four squares R of size $2^{-2}l(Q)$ (here $r=2$), around are squares of size $2^{-3}l(R)$, then squares of size $2^{-4}l(R)$.

Estimate (5.2) implies

$$\sum_{R\in\mathcal{W}} \sum_{\substack{Q'\in\mathcal{D}'\\ Q\subset R}} \|\Delta_{Q'}^{b}\varphi\|_{L^2(\mu)}^2 \leqslant C\sum_{R} \mu(\lambda R). \tag{5.3}$$

Since for any cube R from the Whitney-type decomposition $\mathcal{W}$ we have $\operatorname{dist}(R,\partial Q)\geqslant \lambda l(R)$, any point in Q is covered by at most $M=M(N,\lambda)$ cubes λR, $R\in\mathcal{W}$. Therefore $\sum_R \mu(\lambda R)\leqslant M\mu(Q)$.

To complete the proof of the lemma, it is enough to notice that the sum on the left-hand side of (5.3) coincides with the sum in (5.1). □

6. Estimates of $\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle$ for disjoint Q and R

The idea of the proof of the main results is pretty simple. We would like to estimate $\langle Tf,g\rangle$. To do that, let us take two dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$, decompose f and g in the martingale difference decomposition given by Lemma 4.1, then estimate the matrix $\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle$, $Q\in\mathcal{D}$, $R\in\mathcal{D}'$, and conclude that the operator T is bounded.

LEMMA 6.1. *Let Q, R be two cubes, $l(Q) \leqslant l(R)$, and let $\operatorname{dist}(Q,R) \geqslant l(Q)$. Let $\varphi_Q, \psi_R \in L^2(\mu)$ be functions supported by the cubes Q and R respectively. Suppose also that φ_Q is orthogonal to constants. Then*

$$|\langle T\varphi_Q, \psi_R\rangle| \leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(Q,R)^{d+\alpha}} \mu(Q)^{1/2} \mu(R)^{1/2} \|\varphi_Q\|_{L^2(\mu)} \|\psi_R\|_{L^2(\mu)}.$$

Proof. Let s_0 be the center of the cube Q. Then we get

$$\begin{aligned} |\langle T\varphi_Q, \psi_R\rangle| &= \left| \iint K(t,s) \varphi_Q(s) \psi_R(t)\, d\mu(s)\, d\mu(t) \right| \\ &= \left| \iint [K(t,s) - K(t,s_0)] \varphi_Q(s) \psi_R(t)\, d\mu(s)\, d\mu(t) \right| \\ &\leqslant C \iint \frac{|s-s_0|^\alpha}{|t-s_0|^{d+\alpha}} |\varphi_Q(s)| \cdot |\psi_R(t)|\, d\mu(s)\, d\mu(t) \\ &\leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(Q,R)^{d+\alpha}} \|\varphi_Q\|_{L^1(\mu)} \|\psi_R\|_{L^1(\mu)} \\ &\leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(Q,R)^{d+\alpha}} \mu(Q)^{1/2} \mu(R)^{1/2} \|\varphi_Q\|_{L^2(\mu)} \|\psi_R\|_{L^2(\mu)}. \end{aligned}$$

□

Definition 6.2. Let $\gamma = \alpha/(2\alpha + 2d)$, and so $\gamma d + \gamma\alpha = \frac{1}{2}\alpha$. Let r be some positive integer to be fixed later. Consider a pair of cubes Q and R such that $\operatorname{dist}(Q,R) > 0$. Suppose for definiteness that $l(Q) \leqslant l(R)$. We will call this pair *singular* if $\operatorname{dist}(Q,R) \leqslant l(Q)^\gamma \cdot l(R)^{1-\gamma}$, and *essentially singular* if, in addition, $l(Q) \leqslant 2^{-r} l(R)$.

Definition 6.3. Let $D(Q,R)$ denote the so called *long distance* between cubes:

$$D(Q,R) := \operatorname{dist}(Q,R) + l(Q) + l(R).$$

LEMMA 6.4. *Let T be a Calderón–Zygmund operator and let $\varphi_Q, \psi_R \in L^2(\mu)$ be functions supported by the cubes Q and R respectively and normalized by $\|\varphi_Q\|_{L^2(\mu)} = \mu(Q)^{-1/2}$, $\|\psi_R\|_{L^2(\mu)} = \mu(R)^{-1/2}$. Suppose also that $l(Q) \leqslant l(R)$ and that φ_Q is orthogonal to constants. Then*

$$|\langle T\varphi_Q, \psi_R\rangle| \leqslant C \frac{l(Q)^{\alpha/2} l(R)^{\alpha/2}}{D(Q,R)^{d+\alpha}},$$

provided that $\operatorname{dist}(Q,R) \geqslant \min(l(Q), l(R))$ and the pair Q, R is not essentially singular.

Proof. Without loss of generality one can assume that $l(Q) \leqslant l(R)$. If $\operatorname{dist}(Q,R) \geqslant l(R)$, then $D(Q,R) \leqslant 3 \operatorname{dist}(Q,R)$; thus, the estimate from Lemma 6.1 implies

$$|\langle T\varphi_Q, \psi_R\rangle| \leqslant C \frac{l(Q)^\alpha}{D(Q,R)^{d+\alpha}} \leqslant C \frac{l(Q)^{\alpha/2} l(R)^{\alpha/2}}{D(Q,R)^{d+\alpha}}.$$

Now let us suppose that $\operatorname{dist}(Q,R)\leqslant l(R)$, but the pair Q,R is not singular. That means

$$\operatorname{dist}(Q,R)\geqslant l(Q)^{\gamma}\, l(R)^{1-\gamma}.$$

The estimate of Lemma 6.1 and the identity $\gamma d+\gamma\alpha=\frac{1}{2}\alpha$ imply

$$|\langle T\varphi_Q,\psi_R\rangle|\leqslant \frac{C\cdot l(Q)^{\alpha}}{l(Q)^{\alpha/2}l(R)^{d+\alpha/2}}=\frac{C\cdot l(Q)^{\alpha/2}l(R)^{\alpha/2}}{l(R)^{d+\alpha}}\leqslant C\frac{l(Q)^{\alpha/2}l(R)^{\alpha/2}}{D(Q,R)^{d+\alpha}}. \qquad \square$$

Note that if we do not normalize the functions φ_Q and ψ_R, the estimate from Lemma 6.4 can be rewritten as

$$|\langle T\varphi_Q,\psi_R\rangle|\leqslant C\frac{l(Q)^{\alpha/2}l(R)^{\alpha/2}}{D(Q,R)^{d+\alpha}}\mu(Q)^{1/2}\mu(R)^{1/2}\|\varphi_Q\|_{L^2(\mu)}\|\psi_R\|_{L^2(\mu)}.$$

The following theorem shows that the matrix $\{T_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$ defined by

$$T_{Q,R}:=\frac{l(Q)^{\alpha/2}l(R)^{\alpha/2}}{D(Q,R)^{d+\alpha}}\mu(Q)^{1/2}\mu(R)^{1/2}$$

generates a bounded operator on l^2.

THEOREM 6.5. *Let the measure μ satisfy $\mu(Q)\leqslant Cl(Q)^d$ for all squares Q. Then for the matrix $\{T_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$ defined above, one has*

$$\sum_{\substack{Q\in\mathcal{D}\\ R\in\mathcal{D}'}} T_{Q,R}\,x_Q y_R\leqslant C\biggl(\sum_{Q\in\mathcal{D}}x_Q^2\biggr)^{1/2}\biggl(\sum_{R\in\mathcal{D}'}y_R^2\biggr)^{1/2}$$

for any sequences of non-negative numbers $\{x_Q\}_{Q\in\mathcal{D}},\{y_R\}_{R\in\mathcal{D}'}\in l^2$.

Proof. The symmetry of Q and R implies that it is enough to consider only the sum over Q,R such that $l(Q)\leqslant l(R)$. So we can just assume that $T_{Q,R}=0$ if $l(Q)>l(R)$.

Let us "slice" the matrix $\{T_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$. Namely, for any $n=0,1,2,\ldots$ define the matrix $\{T^{(n)}_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$ by putting

$$T^{(n)}_{Q,R}=\begin{cases} T_{Q,R}, & l(Q)=2^{-n}l(R),\\ 0, & \text{otherwise.}\end{cases}$$

If we show that the norms of the operators $T^{(n)}$ decrease as a geometric progression, i.e., that

$$\sum_{\substack{Q\in\mathcal{D}\\ R\in\mathcal{D}'}} T^{(n)}_{Q,R}\,x_Q y_R\leqslant 2^{-n\beta}C\biggl(\sum_{Q\in\mathcal{D}}x_Q^2\biggr)^{1/2}\biggl(\sum_{R\in\mathcal{D}'}y_R^2\biggr)^{1/2}$$

for some $\beta>0$, then we are done.

We can split the matrices $T^{(n)}$ into layers $T^{(n,k)}$, where

$$T^{(n,k)}_{Q,R}=\begin{cases} T^{(n)}_{Q,R}, & l(R)=2^k,\\ 0, & \text{otherwise.}\end{cases}$$

Clearly, the layers $T^{(n,k)}$ of $T^{(n)}$ do not interfere, therefore it is enough to estimate each layer separately. So, it is enough to show that for any sequences of non-negative $x=\{x_Q\}_{Q\in\mathcal{D}}$, $y=\{y_R\}_{R\in\mathcal{D}'}\in l^2$,

$$\langle T^{(n,k)}x,y\rangle=\sum_{\substack{Q\in\mathcal{D},\, R\in\mathcal{D}'\\ l(Q)=2^{k-n},\, l(R)=2^k}} T^{(n,k)}_{Q,R}x_Q y_R\leqslant 2^{-n\beta}C\biggl(\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^{k-n}}} x_Q^2\biggr)^{1/2}\biggl(\sum_{\substack{R\in\mathcal{D}'\\ l(R)=2^k}} y_R^2\biggr)^{1/2}.$$

One can rewrite the matrix $T^{(n,k)}$ as an integral operator. Namely, if we define

$$X:=\sum_{Q\in\mathcal{D}:\, l(Q)=2^{k-n}}\mu(Q)^{-1/2}x_Q\chi_Q,\quad Y:=\sum_{R\in\mathcal{D}':\, l(R)=2^k}\mu(R)^{-1/2}y_R\chi_R,$$

then

$$\|X\|^2_{L^2(\mu)}=\sum_{Q\in\mathcal{D}:\, l(Q)=2^{k-n}}x_Q^2,\quad \|Y\|^2_{L^2(\mu)}=\sum_{R\in\mathcal{D}':\, l(R)=2^k}y_R^2.$$

Now the estimate we need can be rewritten as

$$\sum_{\substack{l(Q)=2^{k-n}\\ l(R)=2^k}} T^{(n,k)}_{Q,R}x_Q y_R=\iint K^{(n)}_k(s,t)X(s)\cdot Y(t)\,d\mu(s)\,d\mu(t)\leqslant C\|X\|_{L^2(\mu)}\|Y\|_{L^2(\mu)},$$

where the kernel $K^{(n)}_k(s,t)$ is defined by

$$K^{(n)}_k(s,t)=\sum_{\substack{Q\in\mathcal{D}:\, l(Q)=2^{k-n}\\ R\in\mathcal{D}':\, l(R)=2^k}} T_{Q,R}\,\mu(Q)^{-1/2}\mu(R)^{-1/2}\chi_Q(s)\chi_R(t).$$

Note that for each pair s,t, the sum has only one non-zero term, so the kernel $K^{(n)}_k(s,t)$ can be easily estimated:

$$K^{(n)}_k(s,t)\leqslant C2^{-n\alpha/2}\cdot\frac{2^{k\alpha}}{(2^k+|t-s|)^{d+\alpha}}=C2^{-n\alpha/2}\mathcal{K}_k(t-s),$$

where $\mathcal{K}_k(s)=2^{k\alpha}/(2^k+|s|)^{d+\alpha}$. Using the Comparison Lemma (Lemma 2.1) one can show that

$$\sup_k\int\mathcal{K}_k(s)\,d\mu(s)\leqslant\text{const}<\infty.$$

So, by the Schur Lemma the integral operators with kernels $\mathcal{K}_k(s-t)$ are uniformly bounded, therefore the norms of the operators $T^{(n,k)}$ (and hence of $T^{(n)}$) decrease as a geometric progression, and we are done. $\square$

7. Paraproducts and the estimate of $\langle T\varphi_Q, \psi_R\rangle$ when $Q\subset R$

As usual in the theory of singular integral operators, to estimate $\langle T\varphi_Q, \psi_R\rangle$ when $Q\subset R$, one can use the so-called *paraproducts*. The classical construction will not work in our case, and we will slightly modify it.

7.1. Paraproducts

Let b_1, b_2 be weakly accretive functions from the statement of the Tb-theorem (Theorem 0.4). Let r be a positive integer to be defined later (it is the same number we used in the definition of essentially singular pairs, see Definition 6.2). We define a *paraproduct* $\Pi=\Pi_{T^*}$ by

$$\Pi f := \sum_{R\in\mathcal{D}'} \sum_{\substack{Q\in\mathcal{D}:\, l(Q)=2^{-r}l(R)\\ \operatorname{dist}(Q,\partial R)\geqslant \lambda l(Q)}} (E_R b_2)^{-1}\cdot E_R f\cdot (\Delta_Q^{b_1})^* T^* b_2.$$

If we are working with a "nice" operator T, then $T^* b_2$ is well defined. Note that even if $T^* b_2$ is not well defined, we still can define $(\Delta_Q^{b_1})^* T^* b_2$ by duality as the function f satisfying

$$\langle f, g\rangle = \langle b_2, T\Delta_Q^{b_1} g\rangle \quad \text{for all } g\in L^2(\mu).$$

Let us study the matrix of Π. Let $Q\in\mathcal{D}$, $R\in\mathcal{D}'$. Let φ_Q and ψ_R be functions of the form

$$\varphi_Q(x) = \sum_{\substack{Q'\in\mathcal{D}:\, Q'\subset Q\\ l(Q')=l(Q)/2}} A_{Q'}\cdot \chi_{Q'}(x)\cdot b_1(x), \tag{7.1}$$

$$\psi_R(x) = \sum_{\substack{R'\in\mathcal{D}':\, Q'\subset R\\ l(R')=l(R)/2}} B_{R'}\cdot \chi_{R'}(x)\cdot b_2(x), \tag{7.2}$$

where $A_{Q'}, B_{R'}$ are some constants. Suppose also that the functions φ_Q, ψ_R are orthogonal to constants, i.e., $\int \varphi_Q\, d\mu=0$, $\int \psi_R\, d\mu=0$.

The above representation, together with orthogonality to constants, means simply that $\Delta_Q^{b_1}\varphi_Q=\varphi_Q$ and $\Delta_R^{b_2}\psi_R=\psi_R$. One should think of φ_Q, ψ_R as terms in the martingale difference decompositions, $\varphi_Q=\Delta_Q^{b_1} f$, $\psi_R=\Delta_R^{b_2} g$, $f, g\in L^2(\mu)$.

Notice that $\langle \varphi_Q, \Pi\psi_R\rangle$ is non-zero only if $Q\subset R$, $l(Q)<2^{-r}l(R)$. Moreover, there should exist a dyadic cube $S\in\mathcal{D}'$, $l(S)=2^r l(Q)$, $Q\subset S\subset R$, and for this cube S the inequality $\operatorname{dist}(Q,\partial S)\geqslant \lambda l(Q)$ should hold. Let $R_1\in\mathcal{D}'$ be the dyadic cube of size $\frac{1}{2}l(R)$ containing S (it may coincide with S).

In this case,

$$\langle\varphi_Q, \Pi\psi_R\rangle = \langle\varphi_Q, (\Delta_Q^{b_1})^* T^* b_2\rangle B_{R_1} = \langle T\varphi_Q, b_2\rangle B_{R_1}, \tag{7.3}$$

where B_{R_1} is the corresponding constant $B_{R'}$ in (7.2).

THEOREM 7.1. *Let b_1 and b_2 be weakly accretive functions. If $T^*b_2 \in \mathrm{BMO}_\lambda^2(\mu)$, then the paraproduct Π is bounded on $L^2(\mu)$.*

Proof. First notice that $|E_R b_2| \leqslant 1/\delta$. Therefore Lemma 4.1 and a standard duality argument imply that it is sufficient to prove the following embedding theorem:

$$\sum_{R\in\mathcal{D}'} |\mathbf{f}_R|^2 \sum_{\substack{Q\in\mathcal{D}:\, l(Q)=2^{-r}l(R)\\ \mathrm{dist}(Q,\partial R)\geqslant\lambda l(Q)}} \|(\Delta_Q^{b_1})^* T^* b_2\|_{L^2(\mu)}^2 \leqslant C\|f\|_{L^2(\mu)}^2;$$

here $\mathbf{f}_R$ denotes the average of f, $\mathbf{f}_R := \mu(R)^{-1}\int_R f\,d\mu$.

Let

$$a_R = \sum_{\substack{Q\in\mathcal{D}:\, l(Q)=2^{-r}l(R)\\ \mathrm{dist}(Q,\partial R)\geqslant\lambda l(Q)}} \|(\Delta_Q^{b_1})^* T^* b_2\|_{L^2(\mu)}^2.$$

Since $T^*b_2 \in \mathrm{BMO}_\lambda^2(\mu)$, Lemma 5.1 implies that the family $\{a_R\}_{R\in\mathcal{D}'}$ satisfies the Carleson measure condition

$$\sum_{R'\subset R} a_{R'} \leqslant C\mu(R).$$

Therefore the Carleson Embedding Theorem (Theorem 3.1) implies

$$\sum_{R\in\mathcal{D}'} |\mathbf{f}_R|^2 a_R \leqslant C\|f\|_{L^2(\mu)}^2. \qquad \square$$

Since we know that the paraproduct Π is bounded, we only need to estimate the matrix $\langle (T-\Pi^*)\varphi_Q, \psi_R\rangle$, $Q\in\mathcal{D}$, $R\in\mathcal{D}'$.

Definition 7.2. Let Q, R be a pair of cubes. Suppose for the definiteness that $l(Q)\leqslant l(R)$. We call this pair *singular* if

$$\mathrm{dist}(Q,\partial R) \leqslant l(Q)^\gamma l(R)^{1-\gamma},$$

or

$$\mathrm{dist}(Q,\partial R_k) \leqslant l(Q)^\gamma l(R_k)^{1-\gamma}$$

for some subcube $R_k \subset R$ of size $\frac{1}{2}l(R)$; here $\gamma=\alpha/(2\alpha+2d)$, and so $\gamma d+\gamma\alpha=\frac{1}{2}\alpha$. We call the singular pair Q, R *essentially singular* if, in addition, $l(Q)<2^{-r}l(R)$.

Note that the definitions are consistent with the ones we had for disjoint Q and R, see Definition 6.2.

7.2. Estimates of the matrix

From here on we assume that r in the definition of essentially singular pairs is large enough such that $2^{r(1-\gamma)} \geqslant \lambda$. Suppose that we have two dyadic cubes $Q \in \mathcal{D}$, $S \in \mathcal{D}'$, $Q \subset S$, $l(Q) = 2^{-r} l(S)$. Suppose also that $\operatorname{dist}(Q, \partial S) \geqslant l(Q)^\gamma l(S)^{1-\gamma}$. Then the inequality $2^{r(1-\gamma)} \geqslant \lambda$ implies that

$$\operatorname{dist}(Q, \partial S) \geqslant l(Q)^\gamma l(S)^{1-\gamma} = l(Q) 2^{r(1-\gamma)} \geqslant \lambda l(Q).$$

Therefore, if R is a dyadic cube of size at least $2l(S)$, $Q \subset S \subset R$, and the pair Q, R is not singular, then $\langle \varphi_Q, \Pi \psi_R \rangle$ is given by (7.3).

Let φ_Q, ψ_R be two functions of the form (7.1), (7.2), and let φ_Q be orthogonal to constants. Suppose also that the functions φ_Q, ψ_R are normalized in $L^2(\mu)$:

$$\|\varphi_Q\|^2_{L^2(\mu)} = 1, \quad \|\psi_R\|^2_{L^2(\mu)} = 1.$$

Let $R_k \in \mathcal{D}'$, $k = 1, 2, \ldots, 2^N$, be the dyadic cubes of size $\frac{1}{2} l(R)$ contained in R. Then ψ_R can be written as

$$\psi_R(x) = \sum_{k=1}^{2^N} B_k \cdot \chi_{R_k}(x) \cdot b_2(x).$$

Without loss of generality one can assume that $Q \subset R_1$. Then (see (7.3)),

$$\begin{aligned} |\langle (T - \Pi')\varphi_Q, \psi_R \rangle| &= |\langle T\varphi_Q, \psi_R - B_1 b_2 \rangle| \\ &\leqslant |B_1| \cdot |\langle T\varphi_Q, (\chi_{R_1} - 1) b_2 \rangle| + \sum_{k=2}^{2^N} |\langle T\varphi_Q, B_k \cdot \chi_{R_k} \cdot b_2 \rangle| \end{aligned}$$

The first term is easy to estimate. Using property (ii) of Calderón–Zygmund kernels and the orthogonality of φ_Q to constants, we can write for $x \in \mathbf{R}^N \setminus Q$,

$$|(T\varphi_Q)(x)| \leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(x, Q)^{d+\alpha}} \cdot \|\varphi_Q\|_{L^1(\mu)} \leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(x, Q)^{d+\alpha}} \cdot \mu(Q)^{1/2}.$$

Applying the Comparison Lemma (Lemma 2.1) one can get

$$|\langle T\varphi_Q, (\chi_{R_1} - 1) b_2 \rangle| \leqslant \int_{\mathbf{R}^N \setminus R_1} |T\varphi_Q| \cdot |B_2| \, d\mu \leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(Q, \partial R_1)^\alpha} \cdot \mu(Q)^{1/2}.$$

Since $\|\psi_R\|_{L^2(\mu)} = 1$, we have $|B_1| \leqslant \mu(R_1)^{-1/2}$ and therefore

$$|B_1| \cdot |\langle T\varphi_Q, (\chi_{R_1} - 1) b_2 \rangle| \leqslant C \frac{l(Q)^\alpha}{\operatorname{dist}(Q, \partial R_1)^\alpha} \cdot \left(\frac{\mu(Q)}{\mu(R_1)} \right)^{1/2}.$$

The pair Q,R is not singular, which implies

$$\operatorname{dist}(Q,\partial R_1)\geqslant l(Q)^\gamma l(R_1)^{1-\gamma}\geqslant l(Q)^{1/2}l(R_1)^{1/2},$$

and therefore

$$|B_1|\cdot|\langle T\varphi_Q,(\chi_{R_1}-1)b_2\rangle|\leqslant C\cdot\left(\frac{l(Q)}{l(R_1)}\right)^{\alpha/2}\left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}.$$

To estimate $\langle T\varphi_Q,B_k\chi_{R_k}b_2\rangle$, $k=2,3,...,2^N$, we can use Lemma 6.4. It implies (if we take into account that in our case $D(Q,R)\asymp l(R)$, and that $l(R_1)=\frac{1}{2}l(R)$) that

$$\begin{aligned}|\langle T\varphi_Q,B_k\chi_{R_k}b_2\rangle|&\leqslant C\cdot\frac{l(Q)^{\alpha/2}}{l(R)^{d+\alpha/2}}\mu(Q)^{1/2}\mu(R_k)^{1/2}\\&\leqslant C\cdot\left(\frac{l(Q)}{l(R)}\right)^{\alpha/2}\left(\frac{\mu(Q)}{l(R)}\right)^{1/2}\leqslant C\cdot\left(\frac{l(Q)}{l(R)}\right)^{\alpha/2}\left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}.\end{aligned}$$

So we have proved the following lemma.

LEMMA 7.3. *Let r be large enough so that $2^r\geqslant 4\lambda$ (see Lemma 5.1) and $2^{r(1-\gamma)}\geqslant\lambda$. Let $Q\in\mathcal{D}$, $R\in\mathcal{D}'$ be dyadic cubes, $Q\subset R$, $l(Q)<2^{-r}l(R)$. Suppose also that the pair Q,R is not singular. Let φ_Q and ψ_R be functions of the form (7.1), (7.2), and let φ_Q be orthogonal to constants. Let also $R_1\in\mathcal{D}'$ be the dyadic cube of size $\frac{1}{2}l(R)$ containing Q (clearly $R_1\subset R$). Then for the Calderón–Zygmund operator T,*

$$|\langle(T-\Pi^*)\varphi_Q,\psi_R\rangle|\leqslant C\cdot\left(\frac{l(Q)}{l(R)}\right)^{\alpha/2}\left(\frac{\mu(Q)}{\mu(R_1)}\right)^{1/2}\|\varphi_Q\|_{L^2(\mu)}\|\psi_R\|_{L^2(\mu)}.$$

Let the matrix $\{T_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$ be defined by

$$T_{Q,R}=\begin{cases}(l(Q)/l(R))^{\alpha/2}(\mu(Q)/\mu(R_1))^{1/2}, & Q\subset R,\ l(Q)<2^{-r}l(R),\\ 0, & \text{otherwise},\end{cases}$$

where R_1 is the subcube of R of the first generation ($l(R_1)=\frac{1}{2}l(R)$) containing Q.

LEMMA 7.4. *The matrix $\{T_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$ defined above generates a bounded operator on l^2, i.e.,*

$$\sum_{\substack{Q\in\mathcal{D}\\R\in\mathcal{D}'}}T_{Q,R}\,x_Q\cdot y_R\leqslant C\biggl(\sum_{Q\in\mathcal{D}}x_Q^2\biggr)^{1/2}\biggl(\sum_{R\in\mathcal{D}'}y_R^2\biggr)^{1/2}$$

for any sequences of non-negative numbers $\{x_Q\}_{Q\in\mathcal{D}},\{y_R\}_{R\in\mathcal{D}'}\in l^2$.

Proof. Let us "slice" the matrix $\{T_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$. Namely, for $n=r+1,r+2,r+3,...$, define the matrix $\{T^{(n)}_{Q,R}\}_{Q\in\mathcal{D},R\in\mathcal{D}'}$ by

$$T^{(n)}_{Q,R}=\begin{cases}T_{Q,R}, & l(Q)=2^{-n}l(R),\\ 0, & \text{otherwise.}\end{cases}$$

If we show that the norms of the operators $T^{(n)}$ decrease as a geometric progression, i.e., that

$$\sum_{\substack{Q\in\mathcal{D}\\ R\in\mathcal{D}'}} T^{(n)}_{Q,R}x_Q y_R \leqslant 2^{-n\beta}C\bigg(\sum_{Q\in\mathcal{D}} x_Q^2\bigg)^{1/2}\bigg(\sum_{Q\in\mathcal{D}} y_Q^2\bigg)^{1/2}$$

for some $\beta>0$, then we are done.

We can split the matrices $T^{(n)}$ into layers $T^{(n,k)}$, where

$$T^{(n,k)}_{Q,R}=\begin{cases} T^{(n)}_{Q,R}, & l(Q)=2^k,\\ 0, & \text{otherwise.}\end{cases}$$

Clearly, the layers $T^{(n,k)}$ of $T^{(n)}$ do not interfere; therefore it is enough to estimate each layer separately.

Note that the "rows" $\{T^{(n,k)}_{Q,R}:Q\subset R\}$ (R is fixed, $l(R)=2^{k+n}$) are uniformly (in R) bounded on l^2:

$$\sum_{\substack{Q:Q\subset R\\ l(Q)=2^k}} (T^{(n,k)}_{Q,R})^2 \leqslant C\cdot\bigg(\frac{l(Q)}{l(R)}\bigg)^{\alpha} \sum_{\substack{R_1:R_1\subset R\\ l(R_1)=l(R)/2}}\ \sum_{\substack{Q:Q\subset R_1\\ l(Q)=2^k}} \frac{\mu(Q)}{\mu(R_1)} = 2^N C\cdot\bigg(\frac{l(Q)}{l(R)}\bigg)^{\alpha} = 2^N C 2^{-n\alpha}.$$

Note that the supports of the "rows" of $T^{(n,k)}$ are pairwise disjoint. Therefore the rows do not interfere, and so the norm of $T^{(n,k)}$ is bounded by $C2^{-n\alpha/2}$. We are done. □

8. Estimates of the regular part of the matrix

Let dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$ be given. A dyadic square Q in one lattice (say, in $\mathcal{D}$) is called "bad" if there exists a bigger square R in the other lattice (in $\mathcal{D}'$ in this case) such that the pair Q, R is essentially singular; otherwise the square is called "good".

Let a function $f\in L^2(\mu)$ be supported by a cube of size 2^n. We call the function f "good" ($\mathcal{D}$-good) if $\Delta^{b_1}_Q f=0$ for any "bad" square $Q\in\mathcal{D}$, $l(Q)<2^n$.

If one replaces $\mathcal{D}$ by $\mathcal{D}'$ and b_1 by b_2, one gets the definition of $\mathcal{D}'$-good functions.

Here and in what follows, to avoid notation like $(n,\mathcal{D},b_1)$-good function, we assume that n is fixed, and we will always associate the dyadic lattice $\mathcal{D}$ with the function b_1, and the lattice $\mathcal{D}'$ with b_2.

In the following lemma we assume that r from the definition of completely singular pairs (Definition 6.2) is given. As in §7.2 we assume that r is large enough so that $2^{r(1-\gamma)}\geqslant\lambda$ and $2^r\geqslant 4\lambda$.

Also, let two dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$ be fixed.

LEMMA 8.1. *Suppose that T is a Calderón–Zygmund operator such that $Tb_1, T^*b_2 \in \mathrm{BMO}^2_\lambda(\mu)$, where b_1, b_2 are weakly accretive functions from Theorem* 0.4. *Suppose also that*

$$|\langle Tb_1\chi_Q, b_2\chi_R\rangle| \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2} \tag{8.1}$$

for cubes Q, R of comparable size which are close, i.e., for Q, R such that $2^{-r} \leqslant l(Q)/l(R) \leqslant 2^r$, $\operatorname{dist}(Q,R) \leqslant \min(l(Q), l(R))$.

Then, for any $\mathcal{D}$-good function f and any $\mathcal{D}'$-good function g ($f, g \in L^2(\mu)$, both supported by some cubes of size 2^n) we have

$$|\langle Tf, g\rangle| \leqslant C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}.$$

Proof. We can write the decomposition (see Lemma 4.1)

$$f = \sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}} E_Q^{b_1} f + \sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}} \Delta_Q^{b_1} f,$$

and similarly for g,

$$g = \sum_{\substack{R\in\mathcal{D}'\\ l(R)\leqslant 2^n}} E_R^{b_2} g + \sum_{\substack{R\in\mathcal{D}'\\ l(R)=2^n}} \Delta_R^{b_2} g.$$

Let us estimate the sum $\sum_{Q\in\mathcal{D}, R\in\mathcal{D}'} \langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle$. First notice that the condition (8.1) implies

$$|\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle| \leqslant C\|\Delta_Q f\|_{L^2(\mu)}\|\Delta_R g\|_{L^2(\mu)}.$$

Therefore

$$\begin{aligned}\sum_{\substack{2^{-r}l(R)\leqslant l(Q)\leqslant 2^r l(R)\\ \operatorname{dist}(Q,R)\leqslant\min(l(Q),l(R))}} |\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle| &\leqslant C\Big(\sum_{Q\in\mathcal{D}} \|\Delta_Q^{b_1} f\|^2_{L^2(\mu)}\Big)^{1/2}\Big(\sum_{R\in\mathcal{D}'} \|\Delta_R^{b_2} g\|^2_{L^2(\mu)}\Big)^{1/2}\\ &= C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}\end{aligned}$$

(finitely many bounded diagonals).

On the other hand, Lemma 6.4 and Theorem 6.5 imply that

$$\begin{aligned}\sum_{\substack{2^{-r}l(R)\leqslant l(Q)\leqslant 2^r l(R)\\ \operatorname{dist}(Q,R)\geqslant\min(l(Q),l(R))}} |\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle| &\leqslant C\Big(\sum_{Q\in\mathcal{D}} \|\Delta_Q^{b_1} f\|^2_{L^2(\mu)}\Big)^{1/2}\Big(\sum_{R\in\mathcal{D}'} \|\Delta_R^{b_2} g\|^2_{L^2(\mu)}\Big)^{1/2}\\ &= C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}.\end{aligned}$$

So, we need to estimate the sums over $l(Q)<2^{-r}l(R)$ and $l(R)\leqslant 2^{-r}l(Q)$. Due to the symmetry of the conditions of the lemma, it is enough to estimate only the sum over $l(Q)\leqslant 2^{-r}l(R)$.

It remains to estimate the sum

$$\sum_{l(Q)\leqslant 2^{-r}l(R)} \langle T\Delta_Q f, \Delta_R g\rangle = \sum_{\substack{Q\subset R\\ l(Q)\leqslant 2^{-r}l(R)}} \dots + \sum_{\substack{Q\cap R=\varnothing\\ l(Q)\leqslant 2^{-r}l(R)}} \dots .$$

The second sum can be estimated by Lemma 6.4 and Theorem 6.5:

$$\begin{aligned}\sum_{\substack{Q\cap R=\varnothing\\ l(Q)\leqslant 2^{-r}l(R)}} &|\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle| \\ &\leqslant \sum_{\substack{Q\cap R=\varnothing\\ l(Q)\leqslant 2^{-r}l(R)}} C\frac{l(Q)^{\alpha/2}l(R)^{\alpha/2}}{D(Q,R)^{d+\alpha}}\mu(Q)^{1/2}\mu(R)^{1/2}\|\Delta_Q^{b_1}f\|_{L^2(\mu)}\|\Delta_R^{b_2}g\|_{L^2(\mu)} \\ &\leqslant C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}\end{aligned}$$

(since the functions f,g are "good", the entries $\langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle$ corresponding to essentially singular pairs Q,R are zero, and all others can be estimated as above, see Lemma 6.4).

To estimate the first sum, notice that Π has a very special "triangular" matrix. Namely, in the sum $\langle f,\Pi g\rangle = \sum_{Q,R}\langle \Delta_Q^{b_1} f, \Pi\Delta_R^{b_2} g\rangle$ only the terms with $Q\subset R$, $l(Q)\leqslant 2^{-r}l(R)$ may be non-zero. Thus

$$\sum_{\substack{Q\subset R\\ l(Q)\leqslant 2^{-r}l(R)}} \langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle = \sum_{\substack{Q\subset R\\ l(Q)\leqslant 2^{-r}l(R)}} \langle (T-\Pi^*)\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle + \langle f, \Pi g\rangle.$$

We know that the paraproduct Π is bounded, so we have to estimate the sum. The estimate of the sum follows immediately from Lemmas 7.3 and 7.4.

The sums of terms with $E_Q^{b_1} f$ or $E_R^{b_2} g$,

$$\sum_{\substack{Q\in\mathcal{D},\, l(Q)=2^n\\ R\in\mathcal{D}'}} |\langle TE_Q^{b_1} f, \Delta_R^{b_2} g\rangle|, \qquad \sum_{\substack{R\in\mathcal{D}',\, l(R)=2^n\\ Q\in\mathcal{D}}} |\langle T\Delta_Q^{b_1} f, E_R^{b_2} g\rangle|,$$

can be estimated similarly.

And finally, the sum

$$\sum_{\substack{Q\in\mathcal{D},\, l(Q)=2^n\\ R\in\mathcal{D}',\, l(R)=2^n}} |\langle TE_Q^{b_1} f, E_R^{b_2} g\rangle|$$

is bounded because it contains at most 2^{2N} non-zero terms (recall that f,g are supported on a cube of size 2^n). $\square$

9. The Tb-theorem with a stronger weak boundedness assumption

In this section we will prove the following, weaker version of the Tb-theorem (Theorem 0.4), using a stronger version of the weak boundedness assumption. In this section we assume that the operator T is well defined on compactly supported functions and satisfies the conditions (0.4) above in the Introduction (one should think of the truncated operators T_ε here).

THEOREM 9.1. *Let T be a Calderón–Zygmund operator such that Tb_1, T^*b_2 are in $\mathrm{BMO}^2_\lambda(\mu)$ for some weakly accretive functions b_1, b_2. Suppose also that*

$$|\langle Tb_1\chi_Q, b_2\chi_R\rangle| \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2} \tag{9.1}$$

for all cubes Q, R such that

$$\tfrac{1}{2}l(R) \leqslant l(Q) \leqslant 2l(R) \quad \text{and} \quad \operatorname{dist}(Q,R) < 0.1\cdot\min(l(Q), l(R))$$

(*this assumption is a bit stronger than weak boundedness of b_2Tb_1*).

Then the operator T is bounded on $L^2(\mu)$.

First notice that the assumptions of the theorem imply that inequality (9.1) holds for all cubes Q, R satisfying

$$2^{-r}l(R) \leqslant l(Q) \leqslant 2^r l(R) \quad \text{and} \quad \operatorname{dist}(Q,R) < 0.1\cdot\min(l(Q), l(R)),$$

with constant depending on r, of course.

We will need this estimate for r satisfying

$$r \geqslant \frac{1}{\gamma}\log_2\left(\frac{2^9N2^{4N}}{1-2^{-\gamma}}A^2\right),$$

where $A=\max(A(b_1), A(b_2))$, $A(b_1)$, $A(b_2)$ are equivalence constants from Lemma 4.1.

Note that it is an easy exercise to check that condition (9.1) implies

$$|\langle T\Delta_Q^{b_1}f, \Delta_R^{b_2}g\rangle| \leqslant C\|\Delta_Q^{b_1}f\|_{L^2(\mu)}\|\Delta_R^{b_2}g\|_{L^2(\mu)}.$$

To prove the theorem we would like to estimate the bilinear form $\langle Tf, g\rangle$. We already estimated it for "good" functions f and g, see Lemma 8.1.

After we have proved the estimate for "good" functions, the question arises: What should we do about the "bad" ones? And the surprising answer is—nothing, just ignore them! The point is that "bad" cubes are extremely rare, so we do not have to worry about them.

Let us explain why.

9.1. A random dyadic lattice

Our random lattice will contain the dyadic cubes of standard size 2^k ($k\in\mathbf{Z}$), but will be "randomly shifted" with respect to the standard dyadic lattice $\mathcal{D}_0$. The simplest idea would be to pick up a random variable ξ uniformly distributed over $\mathbf{R}^N$ and to define the random lattice as $\xi+\mathcal{D}_0$. Unfortunately, there exists no such ξ, and we have to act in a little bit more sophisticated way.

Let us construct a random lattice of dyadic intervals on the real line $\mathbf{R}$, and then define a random lattice in $\mathbf{R}^N$ as the product of the lattices of intervals.

Let Ω_1 be some probability space and let $x(\omega)$ be a random variable uniformly distributed over the interval $[0,1)^N$.

Let $\xi_j(\omega)$ be random variables satisfying $\mathbf{P}\{\xi_j=+1\}=\mathbf{P}\{\xi_j=-1\}=\frac{1}{2}$. Assume also that $x(\omega), \xi_j(\omega)$ are independent. Define the random lattice $\mathcal{D}(\omega)$ as follows:

(i) Let $I_0(\omega)=[x(\omega)-1, x(\omega)]\in\mathcal{D}(\omega)$. This uniquely determines all intervals in $\mathcal{D}(\omega)$ of length 2^k where $k\leqslant 0$.

(ii) The intervals $I_k(\omega)\in\mathcal{D}(\omega)$ of length 2^k with $k>0$ are determined inductively: if $I_{k-1}(\omega)\in\mathcal{D}$ is already chosen, $I_k(\omega)$ is determined by the following rule: $(I_k(\omega))_+=I_{k-1}(\omega)$ if $\xi_k(\omega)=+1$ and $(I_k(\omega))_-=I_{k-1}(\omega)$ if $\xi_k(\omega)=-1$. In other words, at every step we extend the interval $I_{k-1}(\omega)$ to the left if $\xi_k(\omega)=+1$ and to the right otherwise. Clearly, to know one interval of length 2^k in the lattice is enough to determine all of them.

To get a random dyadic lattice in $\mathbf{R}^N$ we just take a product of N independent random lattices in $\mathbf{R}$.

It is easy to check that the random lattice $\mathcal{D}(\omega)$ in $\mathbf{R}^N$ constructed in this way is uniformly distributed over $\mathbf{R}^N$ and satisfies the following

Equidistribution property. For $x\in\mathbf{R}^N$, $k\in\mathbf{Z}$, the probability that $\operatorname{dist}(x,\partial Q)\geqslant\varepsilon l(Q)$ for some cube of size 2^k is exactly $(1-2\varepsilon)^N$.

9.2. Bad cubes

Let $\mathcal{D}(\omega)$ and $\mathcal{D}'(\omega')$ ($(\omega,\omega')\in\Omega\times\Omega$) be two independent random dyadic lattices, constructed above. We will call a cube $Q\in\mathcal{D}(\omega)$ *bad* if there exists a cube $R\in\mathcal{D}'(\omega')$ of length $l(R)\geqslant l(Q)$ such that the pair Q, R is essentially singular. Otherwise we will call the cube Q *good*.

The definition of bad cubes in $\mathcal{D}'(\omega')$ is the same (now we look for a bigger cube in $\mathcal{D}(\omega)$).

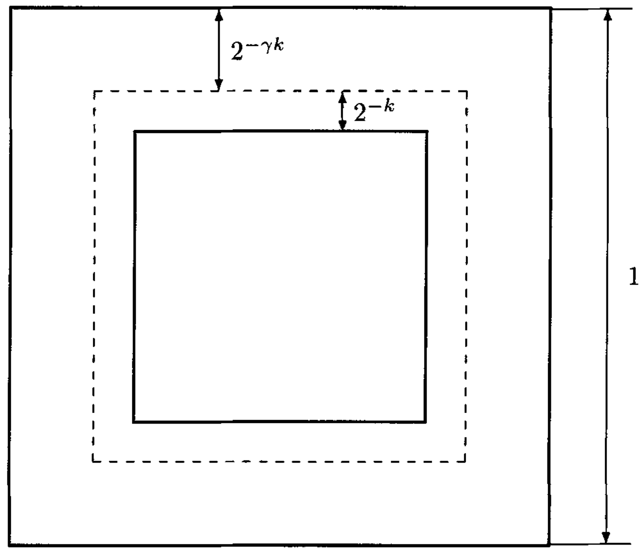

Fig. 6. Estimate of probability P_k.

LEMMA 9.2. *Let r,γ be from the definition of essentially singular pairs, see Definition 6.2. Then for any fixed ω and a cube $Q\in\mathcal{D}(\omega)$ we have*

$$P:=\mathbf{P}_{\omega'}\{Q \text{ is bad}\}\leqslant 2N\frac{2^{-r\gamma}}{1-2^{-\gamma}}.$$

Proof. Given a cube $Q\in\mathcal{D}(\omega)$ (ω is fixed) the probability P^k that there exists a cube $R\in\mathcal{D}'(\omega')$, $Q\subset R$, of size $2^k l(Q)$ such that

$$\operatorname{dist}(Q,\partial R)\leqslant l(Q)^\gamma l(R)^{1-\gamma}$$

can be estimated as

$$P^k\leqslant 1-(1-(2^{-k}+2^{-\gamma k}))^N\leqslant 2N2^{-\gamma k},$$

see Figure 6. So, the probability P can be estimated as

$$P=\sum_{k\geqslant r}P_k\leqslant 2N\sum_{k\geqslant r}2^{-\gamma k}=2N\frac{2^{-r\gamma}}{1-2^{-\gamma}}. \qquad \square$$

9.3. With large probability "bad" parts are small

Consider functions f and g supported on some cube of size 2^n. One can write down the decomposition

$$f=\sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n}}\Delta_Q^b f+\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^n}}E_Q^b f,$$

where the series converges in $L^2(\mu)$, see Lemma 4.1.

Let us split $f=f_{\rm good}+f_{\rm bad}$, where

$$f_{\rm bad}:=\sum_{\substack{Q\in\mathcal{D}\\ l(Q)\leqslant 2^n\\ Q\text{ is bad}}}\Delta_Q^{b_1}f.$$

Here "bad" means "$\mathcal{D}'$-bad" where $\mathcal{D}'=\mathcal{D}'(\omega')$ is the other random dyadic lattice.

Similarly, one can decompose $g=g_{\rm good}+g_{\rm bad}$, where

$$g_{\rm bad}:=\sum_{\substack{Q\in\mathcal{D}'\\ l(Q)\leqslant 2^n\\ Q\text{ is bad}}}\Delta_Q^{b_2}g;$$

here "bad" means "$\mathcal{D}$-bad".

Let us estimate the mathematical expectation $\mathbf{E}\|f_{\rm bad}\|^2_{L^2(\mu)}$ (taken over the random dyadic lattices constructed above). To do that, let us consider (for a fixed dyadic lattice $\mathcal{D}$) the so-called *square function* $S(x)$ defined for $x\in\mathbf{R}^n$ by

$$Sf(x)=S_{\mathcal{D}}f:=\sum_{\substack{Q\in\mathcal{D}:Q\ni x\\ l(Q)\leqslant 2^n}}\|\Delta_Q^{b_1}f\|^2_{L^2(\mu)}\mu(Q)^{-1}\chi_Q+\sum_{\substack{Q\in\mathcal{D}:Q\ni x\\ l(Q)=2^n}}\|E_Q^{b_1}f\|^2_{L^2(\mu)}\mu(Q)^{-1}\chi_Q.$$

Clearly,

$$\int_{\mathbf{R}^N}Sf(x)\,d\mu(x)=\sum_{\substack{Q\in\mathcal{D}:Q\ni x\\ l(Q)\leqslant 2^n}}\|\Delta_Q^{b_1}f\|^2_{L^2(\mu)}+\sum_{\substack{Q\in\mathcal{D}:Q\ni x\\ l(Q)=2^n}}\|E_Q^{b_1}f\|^2_{L^2(\mu)}\asymp\|f\|^2_{L^2(\mu)},$$

where $\asymp$ means equivalence in the sense of two-sided estimate, see Lemma 4.1. Note that $\int_{\mathbf{R}^N}Sf(x)\,d\mu(x)\leqslant A(b_1)\|f\|^2_{L^2(\mu)}$, where $A(b_1)$ is the constant from Lemma 4.1.

Consider the average square function $\mathbf{E}_\omega Sf(x)$ (for each $x\in\mathbf{R}^N$ take the mathematical expectation over all dyadic lattices $\mathcal{D}=\mathcal{D}(\omega)$). Changing the order of integration, one can see that $\int_{\mathbf{R}^N}\mathbf{E}_\omega Sf(x)\,d\mu(x)\leqslant A(b_1)\|f\|^2_{L^2(\mu)}$.

The (conditional, ω is fixed) probability $\mathbf{P}_{\omega'}$ that a square Q is bad, is at most $2N2^{-r\gamma}/(1-2^{-\gamma})\leqslant A^{-2}2^{-8}2^{-4N}$, where $A=\max(A(b_1),A(b_2))$, see Lemma 9.2, so

$$\mathbf{E}_{\omega'}Sf_{\rm bad}(z)\leqslant A^{-2}2^{-8}2^{-4N}Sf(z).$$

Since

$$\begin{aligned}\mathbf{E}_{\omega'}\|f_{\rm bad}\|^2\leqslant A\mathbf{E}_{\omega'}\left(\int Sf_{\rm bad}\,d\mu\right)&=A\int\mathbf{E}_{\omega'}Sf_{\rm bad}\,d\mu\\&\leqslant A^{-1}2^{-8}2^{-4N}\int_{\mathbf{R}^N}Sf\,d\mu\leqslant 2^{-8}2^{-4N}\|f\|^2_{L^2(\mu)},\end{aligned}$$

we get $\mathbf{E}_{\omega,\omega'}\|f_{\text{bad}}\|^2=\mathbf{E}_\omega\mathbf{E}_{\omega'}\|f_{\text{bad}}\|^2\leqslant 2^{-8}2^{-4N}\|f\|^2_{L^2(\mu)}$.

The probability that $\|f_{\text{bad}}\|^2_{L^2(\mu)}\geqslant 4\cdot 2^{-8}2^{-4N}\|f\|^2_{L^2(\mu)}$ cannot be more than $\frac{1}{4}$, and therefore with probability $\frac{3}{4}$ we have

$$\|f_{\text{bad}}\|_{L^2(\mu)}\leqslant 2\cdot 2^{-4}2^{-2N}\|f\|_{L^2(\mu)}.$$

So, if we have *two* functions f and g, and two random dyadic lattices $\mathcal{D}(\omega)$ and $\mathcal{D}'(\omega')$, then with probability at least $\frac{1}{2}$ we have simultaneously

$$\|f_{\text{bad}}\|_{L^2(\mu)}\leqslant 2^{-3}2^{-2N}\|f\|_{L^2(\mu)},\quad \|g_{\text{bad}}\|_{L^2(\mu)}\leqslant 2^{-3}2^{-2N}\|g\|_{L^2(\mu)}.$$

9.4. Pulling yourself up by the hair: proof of Theorem 9.1 under the a priori assumption that T is bounded

Let us now prove Theorem 9.1 under the assumption that we know a priori that T is bounded. Let us pick functions $f,g\in L^2(\mu)$, $\|f\|=\|g\|=1$, such that $|\langle Tf,g\rangle|\geqslant\frac{1}{2}\|T\|$. Since compactly supported functions are dense in $L^2(\mu)$, we can always assume that both functions are supported by some cube of size 2^n.

Pick dyadic lattices $\mathcal{D},\mathcal{D}'$ such that

$$\|f_{\text{bad}}\|_{L^2(\mu)}\leqslant 2^{-3}2^{-2N}\|f\|_{L^2(\mu)}\quad\text{and}\quad \|g_{\text{bad}}\|_{L^2(\mu)}\leqslant 2^{-3}2^{-2N}\|g\|_{L^2(\mu)}.$$

We can always pick such a lattice because, as we have shown above, a random pair of lattices fits with probability at least $\frac{1}{2}$.

First, let us recall that by Lemma 8.1 we have the estimate

$$|\langle Tf_{\text{good}},g_{\text{good}}\rangle|\leqslant C\|f_{\text{good}}\|_{L^2(\mu)}\|g_{\text{good}}\|_{L^2(\mu)}.$$

We can write

$$|\langle Tf,g\rangle|\leqslant|\langle Tf_{\text{good}},g\rangle|+|\langle Tf_{\text{bad}},g\rangle|\leqslant|\langle Tf_{\text{good}},g_{\text{good}}\rangle|+|\langle Tf_{\text{good}},g_{\text{bad}}\rangle|+|\langle Tf_{\text{bad}},g\rangle|.$$

We have

$$\begin{aligned}|\langle Tf_{\text{good}},g_{\text{good}}\rangle|&\leqslant C\|f_{\text{good}}\|_{L^2(\mu)}\|g_{\text{good}}\|_{L^2(\mu)}\leqslant C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)}\leqslant C,\\ |\langle Tf_{\text{good}},g_{\text{bad}}\rangle|&\leqslant 2^{-3}2^{-2N}\|T\|,\\ |\langle Tf_{\text{bad}},g\rangle|&\leqslant 2^{-3}2^{-2N}\|T\|,\end{aligned}$$

because $\|f_{\text{bad}}\|_{L^2(\mu)}\leqslant 2^{-3}2^{-2N}$, $\|f_{\text{good}}\|_{L^2(\mu)}\leqslant\|f\|_{L^2(\mu)}\leqslant 1$, and the same is true for g. Therefore, since $|\langle Tf,g\rangle|\geqslant\frac{1}{2}\|T\|$, and $2^{-2N}\leqslant 1$,

$$\tfrac{1}{2}\|T\|\leqslant C+2\cdot 2^{-3}\|T\|.$$

So $\|T\|\leqslant 4C$ and we are done. □

Remark 9.3. As one could see from the proof, to prove the limited version of Theorem 9.1, it was enough to assume that

$$r \geqslant \frac{1}{\gamma} \log_2 \left(\frac{2^9 A^2}{1-2^{-\gamma}} \right).$$

We will need the term 2^{4N} below, in the proof of the full version of Theorem 9.1.

9.5. Pulling yourself up by the hair: proof of the full version of Theorem 9.1

Now let us discuss what we should do to prove the theorem without the a priori assumption that the operator T is bounded.

The easiest way to do that is to restrict the operator T on a subspace where we know that it is bounded.

For example, let us consider a fixed dyadic grid of cubes of size 2^{-n_0}, and let a set X consist of all functions $f \in L^2(\mu)$, $\|f\| \leqslant 1$, constant on the grid and supported by a cube of size 2^n. Define

$$M(n_0, n) = \sup\{|\langle Tf, g\rangle| : f, g \in X\}$$

(f and g can be supported by different cubes).

Clearly, if we show that $M(n_0, n) \leqslant C$ (C independent of n_0, n), then we are done.

It looks like everything works fine in this case. The construction of random dyadic lattices, for example, even gets simpler. We start with the fixed grid of cubes of size 2^{-n_0} (base), and we want to construct grids of bigger cubes. There are 2^N possibilities of how to position a grid of size $2 \cdot 2^{-n_0}$, and we assign each of them probability 2^{-N}. For each choice of the grid of size $2 \cdot 2^{-n_0}$, there are 2^N possibilities of how to arrange a grid of size $2^2 \cdot 2^{-n_0}$: assign to each of them probability 2^{-N}, etc.

Pick functions $f, g \in X$ such that $|\langle Tf, f\rangle| \geqslant \frac{1}{2} M(n_0, n)$, split them into "good" and "bad" parts, pick dyadic lattices so that the norms $\|f_{\text{bad}}\|$, $\|g_{\text{bad}}\|$ are small, and pull yourself out.

There is only one little problem here: $f_{\text{bad}}, g_{\text{bad}}$ are not in X anymore: their support can become bigger. However, this problem is not hard to take care of.

Namely, the support of f_{bad} cannot be too big. Let R be a cube of size 2^n, supporting f. Then (for any dyadic lattice $\mathcal{D}$) R can be covered by at most 2^N dyadic cubes $Q_k \in \mathcal{D}$, $l(Q_k) = 2^n$. Therefore f_{bad} and f_{good} are supported by the union of the cubes Q_k.

Similarly, g_{bad} is supported by a union of at most 2^N cubes Q'_k, $l(Q'_k) = 2^n$.

As in the proof of the limited version of Theorem 9.1, we split the functions into good and bad parts, and write the estimate

$$|\langle Tf, g\rangle| \leqslant |\langle Tf_{\text{good}}, g\rangle| + |\langle Tf_{\text{bad}}, g\rangle| \leqslant |\langle Tf_{\text{good}}, g_{\text{good}}\rangle| + |\langle Tf_{\text{good}}, g_{\text{bad}}\rangle| + |\langle Tf_{\text{bad}}, g\rangle|.$$

We have

$$|\langle Tf_{\text{good}}, g_{\text{good}}\rangle| \leqslant C\|f_{\text{good}}\|_{L^2(\mu)}\|g_{\text{good}}\|_{L^2(\mu)} \leqslant C\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)} \leqslant C.$$

Since f_{bad} is supported by 2^N cubes of size 2^n, we can split it into a sum of 2^N functions such that each function is supported by a cube of size 2^n. Therefore

$$|\langle Tf_{\text{bad}}, g\rangle| \leqslant 2^N 2^{-3} 2^{-2N} M(n_0, n) \leqslant \tfrac{1}{8} M(n_0, n),$$

because $\|f_{\text{bad}}\|_{L^2(\mu)} \leqslant 2^{-3}2^{-2N}$. Similarly, since both f_{bad} and g_{good} are supported by 2^N cubes of size 2^n,

$$|\langle Tf_{\text{good}}, g_{\text{bad}}\rangle| \leqslant (2^N)^2 2^{-3} 2^{-2N} M(n_0, n) = \tfrac{1}{8} M(n_0, n),$$

because $\|f_{\text{good}}\|_{L^2(\mu)} \leqslant \|f\|_{L^2(\mu)} \leqslant 1$ and $\|g_{\text{bad}}\|_{L^2(\mu)} \leqslant 2^{-3}2^{-2N}$. Now, since $|\langle Tf, g\rangle| \geqslant \frac{1}{2}M(n_0, n)$, we get

$$\tfrac{1}{2}M(n_0, n) \leqslant C + 2\cdot\tfrac{1}{8}M(n_0, n).$$

Therefore, $M(n_0, n) \leqslant 4C$. □

10. Proof of the full version of the Tb-theorem

Now we are in a position to prove the Tb-theorem (Theorem 0.4). Again, we first consider a special, simpler case of the theorem (see §10.1 below), and then treat the general case.

10.1. Special case of the Tb-theorem: weak boundedness on parallelepipeds

Let us first consider a special case, namely, let us suppose that we have a stronger assumption of weak boundedness:

$$|\langle T\chi_Q b_1, \chi_Q b_2\rangle| \leqslant C\mu(Q) \quad \text{for any } \textit{parallelepiped } Q.$$

Recall that we assume that we have some kind of a priori estimate on the norm of the operator T (for example, we have a sequence of regularized operators), and we would like to get an estimate depending only on quantities in the theorem (independent of the parameter of regularization). Let us point out also that in §11 we will get rid of the assumption of a priori boundedness of T (at least sometimes). But now, in this section T is always already bounded (one should think of two-sided truncations of a Calderón–Zygmund operator), and we are proving only the correct estimate of its norm.

The case of the weaker a priori boundedness assumption, when T is bounded on compactly supported functions (one-sided truncations), is treated in §10.3.

We can pick functions $f, g \in L^2(\mu)$, $\|f\|=\|g\|=1$, such that $\langle Tf, g\rangle \geqslant \frac{3}{4}\|T\|$. As above we can assume that each function is supported by a cube of size 2^n. As in the previous section we can split the functions into "good" and "bad" parts, and write the estimate

$$\begin{aligned} |\langle Tf, g\rangle| &\leqslant |\langle Tf_{\mathrm{good}}, g\rangle| + |\langle Tf_{\mathrm{bad}}, g\rangle| \\ &\leqslant |\langle Tf_{\mathrm{good}}, g_{\mathrm{good}}\rangle| + |\langle Tf_{\mathrm{good}}, g_{\mathrm{bad}}\rangle| + |\langle Tf_{\mathrm{bad}}, g\rangle|. \end{aligned} \tag{10.1}$$

As we have shown in the previous section (§9), we can pick dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$ such that $\|f_{\mathrm{bad}}\|_{L^2(\mu)} \leqslant 2^{-3}2^{-2N}$, $\|g_{\mathrm{bad}}\|_{L^2(\mu)} \leqslant 2^{-3}2^{-2N}$, and therefore

$$|\langle Tf_{\mathrm{good}}, g_{\mathrm{bad}}\rangle| + |\langle Tf_{\mathrm{bad}}, g\rangle| \leqslant \tfrac{1}{4}\|T\|. \tag{10.2}$$

Unfortunately, now we cannot estimate $|\langle Tf_{\mathrm{good}}, g_{\mathrm{good}}\rangle| \leqslant C$, because in the sum

$$\sum_{\substack{Q\in\mathcal{D}\\ R\in\mathcal{D}'}} \langle T\Delta_Q^{b_1} f, \Delta_R^{b_2} g\rangle$$

we have infinitely many terms with Q and R of comparable size such that $Q\cap R \neq \varnothing$. And we do not have any estimate for such terms!

10.1.1. *Idea of the proof.* Recall that in the weak version of the Tb-theorem (Theorem 9.1), we did not have any good estimate for terms where the pair Q, R is essentially singular. We dumped these terms into "bad" parts of the functions, and we were able to "pull ourselves up by the hair". We will try to do the same trick with $|\langle Tf_{\mathrm{good}}, g_{\mathrm{good}}\rangle|$ now.

Namely, we want to get the estimate

$$|\langle Tf_{\mathrm{good}}, g_{\mathrm{good}}\rangle| \leqslant C + \tfrac{1}{4}\|T\|. \tag{10.3}$$

Together with (10.1), (10.2) this implies

$$|\langle Tf, g\rangle| \leqslant \tfrac{1}{2}\|T\| + C.$$

Since $|\langle Tf, g\rangle| \geqslant \frac{3}{4}\|T\|$, we get

$$\tfrac{1}{4}\|T\| \leqslant C,$$

and we are done!

To estimate $|\langle Tf_{\mathrm{good}}, g_{\mathrm{good}}\rangle|$ it is enough to estimate the sum

$$\sum_{l(Q), l(R) \leqslant 2^n} |\langle T\Delta_Q^{b_1} f, \Delta_Q^{b_2} g\rangle| \tag{10.4}$$

over all cubes Q, R of comparable size

$$2^{-r}l(Q) \leqslant l(R) \leqslant 2^r l(Q),$$

where r is the same as in Theorem 9.1. Here $\Delta_Q^{b_1} f$ should be replaced by $E_Q^{b_1} f$ if $l(Q)=2^n$, and similarly for R. Let us recall that since f is supported by a cube of size 2^n there are at most 2^N terms $E_Q^{b_1} f$, $l(Q)=2^n$, in the decomposition of f, and similarly for g.

If in the above sum (10.4) we consider only the terms such that the cubes Q and R are separated ($\operatorname{dist}(Q,R) \geqslant \varepsilon \min(l(Q), l(R))$, $\varepsilon>0$), then the sum is bounded by a constant $C=C(\varepsilon)$. Therefore, we only need to estimate the sum over all cubes Q, R such that $\operatorname{dist}(Q,R) < \varepsilon \min(l(Q), l(R))$, and $\varepsilon>0$ can be as small as we want. Of course, the estimate of $\|T\|$ we finally obtain will increase as $\varepsilon \to 0$, but we are not after the optimal estimate, so we can stop at arbitrary small ε.

To estimate the sum (10.4) over all cubes of comparable size ($2^{-r}l(Q) \leqslant l(R) \leqslant 2^r l(Q)$), $\operatorname{dist}(Q,R) < \varepsilon \min(l(Q), l(R))$, it is convenient to write it in a different form. Namely, we can rewrite the layer

$$\Delta_k^{b_1} f = \sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^k}} \Delta_Q^{b_1} f, \quad k \leqslant n,$$

as

$$\Delta_k^{b_1} f = \sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^{k-1}}} c_Q(f) b_1, \quad k \leqslant n,$$

where $c_Q(f)$ are some constants. We can write

$$f = \sum_{k\leqslant n} f^k = \sum_{k\leqslant n} \sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^k}} c_Q(f) b_1,$$

where the "top layer" $f^n = \sum_{Q\in\mathcal{D}, l(Q)=2^n} c_Q(f) b_1$ is given by $f^n = E_k^{b_1} f$.

Let us remind the reader that by Lemma 4.1,

$$A^{-1}\|f\|^2_{L^2(\mu)} \leqslant \sum_k \|f^k\|^2_{L^2(\mu)} \leqslant A\|f\|^2_{L^2(\mu)},$$

where the constant $A=A(b_1)$ depends only on the accretive function b_1.

Similarly,

$$\Delta_k^{b_2} g = \sum_{\substack{R\in\mathcal{D}'\\ l(R)=2^k}} \Delta_R^{b_2} g = \sum_{\substack{R\in\mathcal{D}'\\ l(R)=2^{k-1}}} c'_R(g) b_2, \quad k \leqslant n,$$

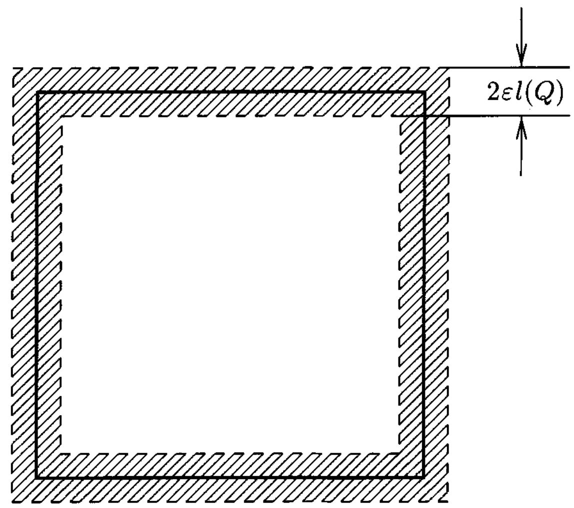

Fig. 7. The set δ_Q (the shaded one).

and

$$g=\sum_{k\leqslant n} g^k=\sum_{k\leqslant n}\sum_{\substack{R\in\mathcal{D}'\\ l(R)=2^k}} c'_R(g)b_2.$$

To estimate the sum (10.4) it is enough to estimate the sum

$$\sum_{k\in\mathbf{Z}}\sum_{Q,R}|c_Q(f)c'_R(g)\langle T\chi_Q b_1,\chi_R b_2\rangle|$$

over all $Q\in\mathcal{D}$ and $R\in\mathcal{D}'$ such that $l(Q),l(R)\leqslant 2^n$, $2^{-n}l(Q)\leqslant l(R)\leqslant 2^n l(Q)$, $\operatorname{dist}(Q,R)\leqslant 10\max(l(Q),l(R))$.

Since for each cube Q there are finitely many (at most $C(N,r)$) cubes $R\in\mathcal{D}'$ satisfying the above condition, and since for separated cubes Q,R (i.e. for cubes such that $\operatorname{dist}(Q,R)\geqslant\varepsilon\min(l(Q),l(R))$ we have the estimate $|\langle T\chi_Q b_1,\chi_R b_2\rangle|\leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}$, it is enough to consider the pairs Q,R satisfying $\operatorname{dist}(Q,R)<\varepsilon\min(l(Q),l(R))$.

10.1.2. *"Cutting out" the "bad" part f^k_{b}.* For a cube Q let $\delta_Q:=(1+2\varepsilon)Q\setminus(1-2\varepsilon)Q$, see Figure 7. For a fixed point $x\in\mathbf{R}^n$ and fixed k, let p_ε be the probability that $x\in\delta_R$ for some cube $R\in\mathcal{D}'(\omega')$, $2^{k-r}\leqslant l(R)\leqslant 2^{k+r}$, where $\mathcal{D}'(\omega')$ is the random dyadic lattice constructed above in §9. Note that p_ε does not depend on k, and that $p_\varepsilon\to 0$ as $\varepsilon\to 0$. Of course, if we consider the random dyadic lattice $\mathcal{D}(\omega)$, we get the same probability p_ε. Note that one can compute the probability p_ε, but we only need the fact that it can be arbitrarily small.

For a cube $Q\in\mathcal{D}$ let Q_{b} be its "bad" part,

$$Q_{\mathrm{b}}=Q\cap\Bigl(\bigcup_{\substack{R\in\mathcal{D}'\\ 2^{-k}l(Q)\leqslant l(R)\leqslant 2^k l(Q)}}\delta_R\Bigr).$$

For a function $f\in L^2(\mu)$ define the "bad" parts f_{b}^k of f^k as

$$f_{\mathrm{b}}^k:=\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^k}} c_Q(f)\chi_{Q_{\mathrm{b}}}b_1;$$

here we use the subscript "b" instead of "bad" to avoid confusion with f_{bad}.

Let us estimate the mathematical expectation $\mathbf{E}_{\omega'}\bigl(\sum_k \|f_{\mathrm{b}}^k\|_{L^2(\mu)}^2\bigr)$ over all random lattices $\mathcal{D}'(\omega')$ (the lattice $\mathcal{D}=\mathcal{D}(\omega)$ is fixed). First of all notice that for a fixed $x\in\mathbf{R}^n$,

$$\mathbf{E}_{\omega'}|f_{\mathrm{b}}^k(x)|^2\leqslant p_\varepsilon|f^k(x)|^2,$$

where p_ε is the probability that a point x belongs to δ_R for some cube $R\in\mathcal{D}'(\omega')$ of fixed size 2^k, see above. Therefore, changing the order of integration we get

$$\begin{aligned}\mathbf{E}_{\omega'}\Bigl(\sum_{k\leqslant n}\|f^k\|_{L^2(\mu)}^2\Bigr)&=\sum_{k\leqslant n}\int_{\mathbf{R}^N}\mathbf{E}|f_{\mathrm{b}}^k(x)|^2\,d\mu(x)\leqslant p_\varepsilon\sum_{k\leqslant n}\int_{\mathbf{R}^N}|f^k(x)|^2\,d\mu(x)\\&=p_\varepsilon\sum_{k\leqslant n}\|f^k\|_{L^2(\mu)}^2\leqslant p_\varepsilon A(b_1)\|f\|_{L^2(\mu)}^2,\end{aligned}$$

where $A(b_1)$ is the equivalence constant from Lemma 4.1.

Since the above inequality holds for any dyadic grid $\mathcal{D}=\mathcal{D}(\omega)$, we get for the mathematical expectation $\mathbf{E}=\mathbf{E}_{\omega,\omega'}$,

$$\mathbf{E}\Bigl(\sum_{k\leqslant n}\|f_{\mathrm{b}}^k\|_{L^2(\mu)}^2\Bigr)\leqslant p_\varepsilon A(b_1)\|f\|_{L^2(\mu)}^2=p_\varepsilon A(b_1).$$

Similarly, for the "bad" parts g_{b}^k of the functions g^k,

$$g_{\mathrm{b}}^k:=\sum_{\substack{R\in\mathcal{D}'\\ l(R)=2^k}} c_R'(g)\sum_{\substack{Q\in\mathcal{D}\\ l(Q)=2^k\\ \delta_Q\cap R\neq\varnothing}}\chi_{\delta_Q\cap R}b_2,$$

we get

$$\mathbf{E}\Bigl(\sum_{k\leqslant n}\|g_{\mathrm{b}}^k\|_{L^2(\mu)}^2\Bigr)\leqslant p_\varepsilon A(b_2)\|g\|_{L^2(\mu)}^2=p_\varepsilon A(b_2).$$

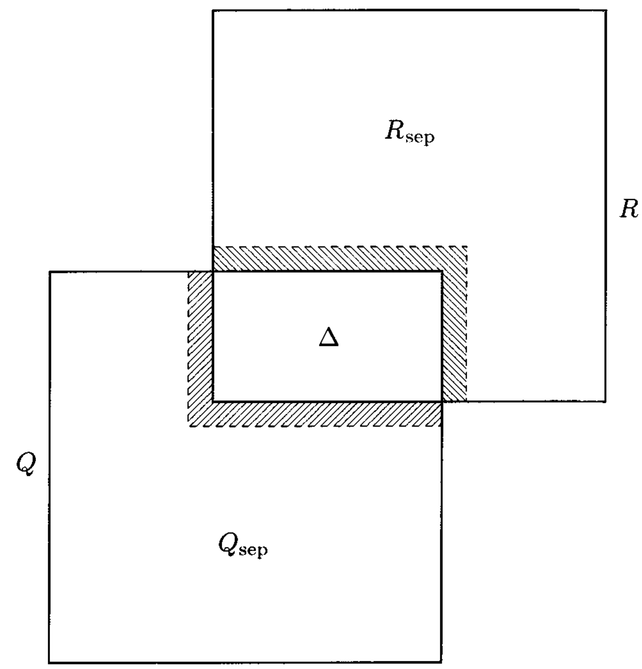

Fig. 8. Cutting out the bad part. The sets Q_∂ and R_∂ are the shaded parts of the squares Q and R respectively.

So, for $A = \max(A(b_1), A(b_2))$ we can estimate the probability

$$\mathbf{P}_{\omega,\omega'}\left\{\sum_{k\leqslant n}\|f_{\mathrm{b}}^k\|^2_{L^2(\mu)} \geqslant 8Ap_\varepsilon\right\} \leqslant \frac{1}{8},$$

and similarly for g. So, with probability at least $1-\frac{1}{4}-\frac{1}{4}-\frac{1}{8}-\frac{1}{8}=\frac{1}{4}$ we get

$$\|f_{\mathrm{bad}}\|^2_{L^2(\mu)} \leqslant 2^{-3}2^{-2N}, \quad \|g_{\mathrm{bad}}\|^2_{L^2(\mu)} \leqslant 2^{-3}2^{-2N} \tag{10.5}$$

and

$$\sum_{k\leqslant n}\|f_{\mathrm{b}}^k\|^2_{L^2(\mu)} \leqslant 8Ap_\varepsilon, \quad \sum_{k\leqslant n}\|g_{\mathrm{b}}^k\|^2_{L^2(\mu)} \leqslant 8Ap_\varepsilon. \tag{10.6}$$

10.1.3. *Estimates of* $|\langle T\chi_Q b_1, \chi_R b_2\rangle|$. Take two dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$ such that all the above inequalities hold (with probability at least $\frac{1}{4}$ random lattices would fit).

Consider two squares

$$Q\in\mathcal{D}, \quad R\in\mathcal{D}', \quad 2^{-k}l(Q)\leqslant l(R)\leqslant 2^k l(Q), \quad \operatorname{dist}(Q,R)<\varepsilon\min(l(Q),l(R)),$$

see Figure 8. We would like to estimate

$$|\langle T\chi_Q b_1, \chi_R b_2\rangle|.$$

Consider first the case when the cubes Q and R are in general position, as in Figure 8: the estimate for cases when $Q\cap R=\varnothing$ or one of the cubes contains the other can be done similarly.

Let $\Delta:=Q\cap R$, $Q_{\rm sep}=Q\setminus\Delta\setminus\delta_R$ (the square Q without Δ and without the shaded part in Figure 8), "sep" means separated (from R and from Δ). Let also $Q_\partial=Q\setminus\Delta\setminus Q_{\rm sep}$ (the shaded part of Q in Figure 8). The symbol ∂ here means boundary, i.e., this set touches R and Δ. Note that $Q_\partial\subset Q\cap\delta_R$.

Similarly, let us split R as $R=R_{\rm sep}\cup R_\partial\cup\Delta$, where all sets are disjoint. Then

$$\langle T\chi_Q b_1,\chi_R b_2\rangle=\langle T\chi_Q b_1,\chi_{R_{\rm sep}} b_2\rangle+\langle T\chi_Q b_1,\chi_{R_\partial} b_2\rangle+\langle T\chi_Q b_1,\chi_\Delta b_2\rangle.$$

The first two terms are easy to estimate: since Q and $R_{\rm sep}$ are separated,

$$|\langle T\chi_Q b_1,\chi_{R_{\rm sep}} b_2\rangle|\leqslant C\mu(Q)^{1/2}\mu(R_{\rm sep})^{1/2}\leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}$$

(the constant C here of course depends on ε). The second term can be estimated as

$$|\langle T\chi_Q b_1,\chi_{R_\partial} b_2\rangle|\leqslant \|T\|\cdot\|\chi_Q b_1\|_{L^2(\mu)}\|\chi_{R_{\rm b}} b_2\|_{L^2(\mu)},$$

because $R_\partial\subset R_{\rm b}$.

To estimate the last term, let us write it as

$$\langle T\chi_Q b_1,\chi_\Delta b_2\rangle=\langle T\chi_\Delta b_1,\chi_\Delta b_2\rangle+\langle T\chi_{Q_\partial} b_1,\chi_\Delta b_2\rangle+\langle T\chi_{Q_{\rm sep}} b_1,\chi_\Delta b_2\rangle.$$

The first term is bounded by $C\mu(Q)^{1/2}\mu(R)^{1/2}$ by the assumption of the theorem. The other two can be estimated as above (the measure of Q_∂ is small, and $Q_{\rm sep}$ and Δ are separated), so summarizing all we get

$$\begin{aligned}|\langle T\chi_Q b_1,\chi_R b_2\rangle|&\leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}\\&\quad+\|T\|\big(\|\chi_Q b_1\|_{L^2(\mu)}\|\chi_{R_{\rm b}} b_2\|_{L^2(\mu)}+\|\chi_{Q_{\rm b}} b_1\|_{L^2(\mu)}\|\chi_R b_2\|_{L^2(\mu)}\big).\end{aligned}$$

10.1.4. *Final estimates.* We know that

$$\sum|c_Q(f)|^2\|\chi_{Q_{\rm b}} b_1\|^2_{L^2(\mu)}=\sum\|f^k_{\rm b}\|^2_{L^2(\mu)}\leqslant 8Ap_\varepsilon,$$
$$\sum|c'_R(g)|^2\|\chi_{R_{\rm b}} b_2\|^2_{L^2(\mu)}=\sum\|g^k_{\rm b}\|^2_{L^2(\mu)}\leqslant 8Ap_\varepsilon,$$

and that

$$\sum|c_Q(f)|^2\mu(Q)\leqslant C\|f\|^2_{L^2(\mu)}=C,$$
$$\sum|c'_R(g)|^2\mu(R)\leqslant C\|g\|^2_{L^2(\mu)}=C.$$

Since for a cube $Q\in\mathcal{D}$ there are at most $M(N,r)$ cubes $R\in\mathcal{D}'$, $2^{-r}l(Q)\leqslant l(R)\leqslant 2^r l(Q)$ satisfying $\operatorname{dist}(Q,R)\leqslant\varepsilon\min(l(Q),l(R))$, we get, using the Cauchy–Schwarz inequality,

$$\begin{aligned}\sum|c_Q(f)c'_R(g)|\cdot|\langle T\chi_Q b_1,\chi_R b_2\rangle|\\ &\leqslant C\|f\|\cdot\|g\|+M(N,r)\|T\|\biggl(\Bigl(\sum_{k\leqslant n}\|f_{\mathrm{b}}^k\|^2\Bigr)^{1/2}\sqrt{A}\,\|g\|+\sqrt{A}\,\|f\|\cdot\Bigl(\sum_{k\leqslant n}\|g_{\mathrm{b}}^k\|^2\Bigr)^{1/2}\biggr)\\ &\leqslant C\|f\|\cdot\|g\|+M(N,r)A\cdot 4\sqrt{2p_\varepsilon}\,\|T\|\cdot\|f\|\cdot\|g\|=C+M(N,r)A\cdot 4\sqrt{2p_\varepsilon}\,\|T\|,\end{aligned}$$

where the sum is taken over all $Q\in\mathcal{D}$, $R\in\mathcal{D}'$ such that $l(Q),l(R)\leqslant 2^n$, $2^{-r}l(Q)\leqslant l(R)\leqslant 2^r l(Q)$, $\operatorname{dist}(Q,R)<\varepsilon\min(l(Q),l(R))$.

As we said above, this is enough to get the estimate

$$|\langle Tf_{\mathrm{good}},g_{\mathrm{good}}\rangle|\leqslant C+4A\sqrt{2\varepsilon}\,M(N,r)\|T\| \tag{10.7}$$

(of course, C here depends on ε). Taking ε sufficiently small so that $4A\sqrt{2\varepsilon}\,M(N,r)<\frac14$, we get

$$|\langle Tf_{\mathrm{good}},g_{\mathrm{good}}\rangle|\leqslant C+\tfrac14\|T\|,$$

and we are done. □

10.2. The Tb-theorem under the a priori assumption that T is bounded

Now we are going to prove the full version of the Tb-theorem (Theorem 0.4), assuming that the operator T is bounded. The case when the operator is only well defined for compactly supported functions is treated later in §10.3.

We are going to prove the theorem under the definition that *weak boundedness* means that for some $\Lambda>1$ the inequality $|\langle T\chi_Q b_1,\chi_Q b_2\rangle|\leqslant C\mu(\Lambda Q)$ holds for all cubes Q.

To do this we need to modify a little the estimate of $|\langle T\chi_Q b_1,\chi_R b_2\rangle|$, where Q and R are intersecting cubes of comparable size.

The construction goes as above. Let us recall that we have picked f,g in $L^2(\mu)$, $\|f\|_{L^2(\mu)}=\|g\|_{L^2(\mu)}=1$, such that $|\langle Tf,g\rangle|\geqslant\frac34\|T\|$, and we now want to estimate $|\langle Tf,g\rangle|$.

First we pick r in the definition of essentially singular pairs such that with large probability the norms $\|f_{\mathrm{bad}}\|_{L^2(\mu)}$, $\|g_{\mathrm{bad}}\|_{L^2(\mu)}$ are small, which implies the estimate

$$|\langle Tf_{\mathrm{good}},g_{\mathrm{bad}}\rangle|+|\langle Tf_{\mathrm{bad}},g\rangle|\leqslant\tfrac14\|T\|$$

for any dyadic lattice where the norms are small, cf. (10.2). Also, with large probability $\bigl($at least $\frac14\bigr)$, not only the norms of "bad parts" of f and g are small, but also the sums $\sum_k\|f_{\mathrm{b}}^k\|_{L^2(\mu)}$ and $\sum_k\|g_{\mathrm{b}}^k\|_{L^2(\mu)}$ are small, see (10.5), (10.6).

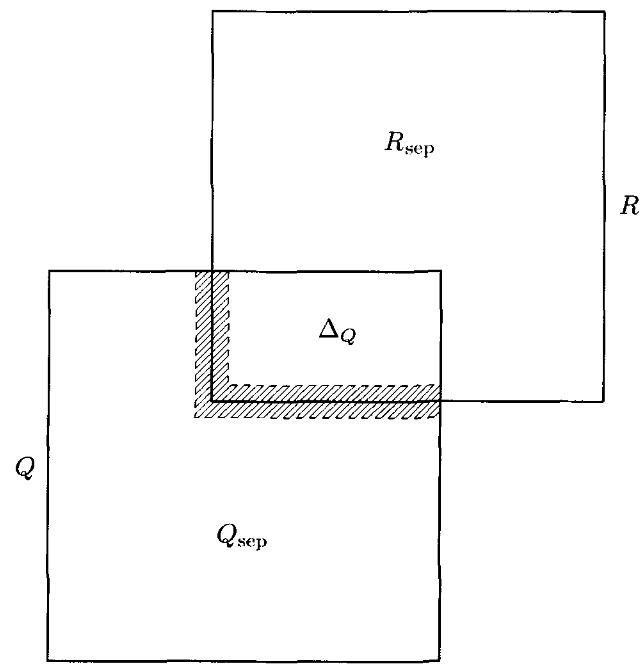

Fig. 9. Cutting out the bad part. Q_∂ is the shaded part.

Take a sufficiently small ε such that $4A\sqrt{2\varepsilon}\,M(N,r)<\frac{1}{4}$. Here, as above, $M(N,r)$ is the upper bound on the number of cubes $R\in\mathcal{D}'$, of comparable size with a given cube Q $(2^{-r}l(Q)\leqslant l(R)\leqslant 2^r l(Q))$, and such that $\operatorname{dist}(Q,R)\leqslant\varepsilon\min(l(Q),l(R))$.

So, now we have ε fixed, as well as two dyadic lattices $\mathcal{D}$ and $\mathcal{D}'$ such that inequalities (10.5), (10.6) hold.

10.2.1. *Cutting out more of the bad stuff.* Fix now two intersecting cubes R and Q of comparable size $((2^{-r}l(Q)\leqslant l(R)\leqslant 2^r l(Q)))$. Fix the size

$$s=(10\Lambda)^{-1}\varepsilon\min(l(Q),l(R)),$$

and "drop" on the set $\Delta:=Q\cap R$ a random grid G of cubes of size s. We want this random grid to be uniformly distributed over $\mathbf{R}^N$, for example we can take a fixed grid and consider all its shifts by $\xi(\omega)$, where ξ is a random vector uniformly distributed over the cube $[0,s)^N$.

For $\varepsilon'>0$ let $G_{\varepsilon'}$ be an $\varepsilon's$-neighborhood of the boundaries of the cubes in the grid G. Then for a fixed point $x\in\mathbf{R}^N$ the probability that $x\in G_{\varepsilon'}$ is $\varphi(\varepsilon')$, where $\varphi(\varepsilon')\to 0$ as $\varepsilon'\to 0$. (Again, here one can write a formula for $\varphi(\varepsilon')$, but we only need the fact that $\varphi(\varepsilon')\to 0$.)

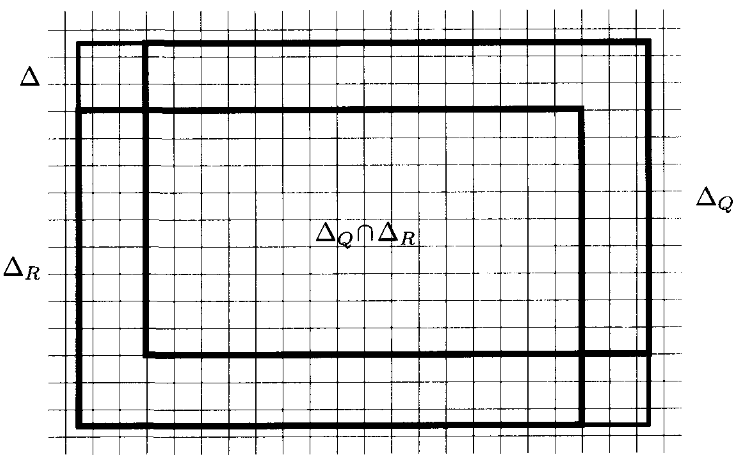

Fig. 10. The intersection $\Delta := Q \cap R$ and the grid $G_{\varepsilon'}$ (grid of small squares). Δ_Q and Δ_R are rectangles bounded by thick lines, Δ is the rectangle bounded by a thinner line. Notice that the boundary of the intersection $\Delta_Q \cap \Delta_R$ goes along the grid $G_{\varepsilon'}$.

Clearly, the expectation $\mathbf{E}(\mu(G_{\varepsilon'} \cap \Delta) = \varphi(\varepsilon')\mu(\Delta)$, so with positive probability $\mu(G_{\varepsilon'} \cap \Delta) \leqslant \varphi(\varepsilon')\mu(\Delta)$. So, for a given ε' (and Δ) one can always find at least one grid G such that the above inequality holds.

10.2.2. *Estimates of* $|\langle T\chi_Q b_1, \chi_R b_2\rangle|$. To estimate $|\langle T\chi_Q b_1, \chi_R b_2\rangle|$ let us split the cubes Q and R into three parts. As above, define Q_{sep} by $Q_{\mathrm{sep}} := Q \setminus \Delta \setminus \delta_R$, where we recall that $\delta_Q := (1+2\varepsilon)Q \setminus (1-2\varepsilon)Q$, see Figure 7.

The main difference with the previous case is in the definition of Q_∂. We want it now to be almost $\delta_R \cap Q$, see Figure 9. By "almost" we mean the following. We want that the boundary hyperplanes of Q_∂ that lie inside Δ do not cut the cubes of the grid G, but go along the boundaries of the grid, see Figure 10. One can always pick hyperplanes such that the distance to the corresponding (parallel) side of R is between $\frac{1}{2}\varepsilon l(R)$ and $\varepsilon l(R)$. It is possible because we assumed that the size s of the cubes of the grid G is at most $(10\Lambda)^{-1}\varepsilon l(R)$.

So, that is how we define Q_∂, and let us call the rest Δ_Q, $\Delta_Q := Q \setminus Q_{\mathrm{sep}} \setminus Q_\partial$, see Figure 9. Note also that $Q_\partial \subset Q_{\mathrm{b}}$.

Let us now estimate

$$\langle T\chi_Q b_1, \chi_R b_2\rangle = \langle T\chi_Q b_1, \chi_{R_{\mathrm{sep}}} b_2\rangle + \langle T\chi_Q b_1, \chi_{R_\partial} b_2\rangle + \langle T\chi_Q b_1, \chi_{\Delta_R} b_2\rangle.$$

The first two terms are easy to estimate: since Q and R_{sep} are separated,

$$|\langle T\chi_Q b_1, \chi_{R_{\mathrm{sep}}} b_2\rangle| \leqslant C\mu(Q)^{1/2}\mu(R_{\mathrm{sep}})^{1/2} \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}$$

(the constant C here of course depends on ε). The second term can be estimated as

$$|\langle T\chi_Q b_1, \chi_{R_\partial} b_2\rangle| \leqslant \|T\|\cdot\|\chi_Q b_1\|_{L^2(\mu)}\|\chi_{R_\mathrm{b}} b_2\|_{L^2(\mu)},$$

because $R_\partial \subset R_\mathrm{b}$.

To estimate the last term, let us write it as

$$\langle T\chi_Q b_1, \chi_{\Delta_R} b_2\rangle = \langle T\chi_{\Delta_Q} b_1, \chi_{\Delta_R} b_2\rangle + \langle T\chi_{Q_\partial} b_1, \chi_{\Delta_R} b_2\rangle + \langle T\chi_{Q_\mathrm{sep}} b_1, \chi_{\Delta_R} b_2\rangle.$$

Clearly we have the estimates

$$|\langle T\chi_{Q_\partial} b_1, \chi_{\Delta_R} b_2\rangle| \leqslant \|T\|\cdot\|\chi_{Q_\mathrm{b}} b_1\|_{L^2(\mu)}\|\chi_R b_2\|_{L^2(\mu)}$$

and

$$|\langle T\chi_{Q_\mathrm{sep}} b_1, \chi_{\Delta_R} b_2\rangle| \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}$$

since Q_sep and Δ_R are separated.

Now we only need to estimate the first term. Let us denote $\Delta'_Q := \Delta_Q \cap G_{\varepsilon'}$, $\tilde{\Delta}_Q := \Delta_Q \setminus G_{\varepsilon'}$, and similarly for Δ_R. Then

$$\langle T\chi_{\Delta_Q} b_1, \chi_{\Delta_R} b_2\rangle = \langle T\chi_{\Delta'_Q} b_1, \chi_{\Delta_R} b_2\rangle + \langle T\chi_{\tilde{\Delta}_Q} b_1, \chi_{\Delta'_R} b_2\rangle + \langle T\chi_{\tilde{\Delta}_Q} b_1, \chi_{\tilde{\Delta}_R} b_2\rangle. \tag{10.8}$$

The first two terms are easy to estimate:

$$\begin{aligned}
|\langle T\chi_{\Delta'_Q} b_1, \chi_{\Delta_R} b_2\rangle| &\leqslant \|T\|\cdot\|\chi_{\Delta'_Q} b_1\|_{L^2(\mu)}\|\chi_{\Delta_R} b_2\|_{L^2(\mu)} \\
&\leqslant \|T\|\cdot\|b_1\|_\infty\|b_2\|_\infty\cdot\mu(\Delta'_Q)^{1/2}\mu(\Delta_R)^{1/2} \\
&\leqslant \|T\|\cdot\|b_1\|_\infty\|b_2\|_\infty\cdot\sqrt{\varphi(\varepsilon')}\cdot\mu(\Delta_Q)^{1/2}\mu(\Delta_R)^{1/2} \\
&\leqslant \|T\|\cdot\|b_1\|_\infty\|b_2\|_\infty\cdot\sqrt{\varphi(\varepsilon')}\cdot\mu(Q)^{1/2}\mu(R)^{1/2},
\end{aligned}$$

and similarly

$$|\langle T\chi_{\tilde{\Delta}_Q} b_1, \chi_{\Delta'_R} b_2\rangle| \leqslant \|T\|\cdot\|b_1\|_\infty\|b_2\|_\infty\sqrt{\varphi(\varepsilon')}\,\mu(Q)^{1/2}\mu(R)^{1/2}.$$

The last term $\langle T\chi_{\tilde{\Delta}_Q} b_1, \chi_{\tilde{\Delta}_R} b_2\rangle$ is bounded by

$$C\mu(\Delta) \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2},$$

where the constant C depends on the parameters in the theorem, as well as on $\varepsilon, r, \varepsilon'$. Indeed, the set $\tilde{\Delta}_Q \cup \tilde{\Delta}_Q$ consists of finitely many disjoint parallelepipeds S_k (most of which are cubes). Moreover, the set $\tilde{\Delta}_Q$ is just a union of some of these parallelepipeds, and similarly for $\tilde{\Delta}_R$.

Since any two disjoint parallelepipeds S_1 and S_2 are separated, and $b_1, b_2 \in L^\infty$, we have

$$|\langle T\chi_{S_1} b_1, \chi_{S_2} b_2\rangle| \leqslant C\mu(S_1)^{1/2}\mu(S_2)^{1/2} \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}.$$

If a parallelepiped S belongs to both $\tilde{\Delta}_Q$ and $\tilde{\Delta}_R$, then it must be a cube, see Figure 10. Then by the assumption of weak boundedness,

$$|\langle T\chi_S b_1, \chi_S b_2\rangle| \leqslant C\mu(\Lambda S) \leqslant C\mu(\Delta) \leqslant C\mu(Q)^{1/2}\mu(R)^{1/2}.$$

Since the number of the parallelepipeds S_k is bounded above by a constant depending only on $r, \varepsilon, \Lambda, \varepsilon'$, then taking the sum over all the parallelepipeds we get the desired estimate.

Summarizing all, we get

$$\begin{aligned}|\langle T\chi_Q b_1, \chi_R b_2\rangle| \leqslant{} & C_1\mu(Q)^{1/2}\mu(R)^{1/2} \\ & + \|T\| \big(\|\chi_Q b_1\|_{L^2(\mu)} \|\chi_{R_b} b_2\|_{L^2(\mu)} + \|\chi_{Q_b} b_1\|_{L^2(\mu)} \|\chi_R b_2\|_{L^2(\mu)}\big) \\ & + C_2\|T\| \cdot \sqrt{\varphi(\varepsilon')} \cdot \mu(Q)^{1/2}\mu(R)^{1/2}.\end{aligned}$$

Here only the last term is new in comparison with the estimate (10.7) from §10.1.

10.2.3. *Final estimates.* Acting as in the previous section (i.e., taking the sum over all Q, R, see above), we can get the estimate

$$|\langle Tf_{\text{good}}, g_{\text{good}}\rangle| \leqslant C + 4A\sqrt{2\varepsilon}\, M(N,r)\|T\| + C'\|T\| \cdot \sqrt{\varphi(\varepsilon')};$$

here again, only the last term is new.

Let us remind the reader that ε was chosen to be small enough such that the second term is bounded by $\frac{1}{4}\|T\|$.

Let us also remind the reader that

$$\tfrac{3}{4}\|T\| \leqslant |\langle Tf, g\rangle| \leqslant \tfrac{1}{4}\|T\| + |\langle Tf_{\text{good}}, g_{\text{good}}\rangle|.$$

So, if we pick ε' to be sufficiently small such that $C'\sqrt{\varphi(\varepsilon')} \leqslant \frac{1}{8}$, we get

$$\tfrac{3}{4}\|T\| \leqslant C + \tfrac{1}{4}\|T\| + \tfrac{1}{4}\|T\| + \tfrac{1}{8}\|T\|,$$

and therefore $\|T\| \leqslant 8C$.

We are done! □

10.3. The full Tb-theorem

Now let us discuss the proof of the full version of the Tb-theorem. We need to relax the assumption that T is bounded, i.e., to replace it by the weaker assumption that for compactly supported functions f, g,

$$|\langle Tf, g\rangle| \leqslant C(A)\|f\|_{L^2(\mu)}\|g\|_{L^2(\mu)},$$

where

$$A = \max\{\operatorname{diam}(\operatorname{supp} f), \operatorname{diam}(\operatorname{supp} g)\}.$$

The definition of weak boundedness remains the same as in §10.2.

To prove the theorem under the above assumptions, we combine ideas from §§ 9.5 and 10.2.

Namely, let us introduce a set X consisting of all functions $f\in L^2(\mu)$, $\|f\|\leqslant 1$, supported by a cube of size 2^n (each function can be supported by its own cube, so X is not a linear space). Define

$$\mathcal{M}(n) = \sup\{|\langle Tf, g\rangle| : f, g \in X\}$$

(f, g can be supported by different cubes).

Clearly, if we show that $\mathcal{M}(n)\leqslant C$ (C independent of n), then we are done.

Pick functions $f, g\in X$ such that $|\langle Tf, g\rangle| \geqslant \frac{3}{4}\mathcal{M}(n)$. Acting as in §9.5, split the functions f and g into "good" and "bad" parts, then get the estimates

$$|\langle Tf_{\mathrm{bad}}, g\rangle| \leqslant 2^N 2^{-3} 2^{-2N}\mathcal{M}(n) \leqslant \tfrac{1}{8}\mathcal{M}(n)$$

and

$$|\langle Tf_{\mathrm{good}}, g_{\mathrm{bad}}\rangle| \leqslant (2^N)^2 2^{-3} 2^{-2N}\mathcal{M}(n) = \tfrac{1}{8}\mathcal{M}(n).$$

Then, acting as in §10.2, we get the estimate

$$|\langle Tf_{\mathrm{good}}, g_{\mathrm{good}}\rangle| \leqslant C + 4A\sqrt{2\varepsilon}\, M(N, r)\mathcal{M}(n) + C'\mathcal{M}(n)\cdot\sqrt{\varphi(\varepsilon')}\,.$$

Note that the crucial part of the above estimate is the estimate of $|\langle T\chi_Q b_1, \chi_R b_2\rangle|$. Since by the construction both χ_Q and χ_R are supported by cubes of size 2^n, one does not need to change anything in the reasoning, except replacing $\|T\|$ by $\mathcal{M}(n)$.

We leave the rest and all details to the reader as an easy exercise. One does not need even to change constants.

10.4. Remarks about other weak boundedness conditions

All the results of the above §§ 10.2 and 10.3 remain true if we consider a different weak boundedness condition,

$$|\langle Tb_1\chi_{\lambda Q}, b_2\chi_Q\rangle| \leqslant C\mu(\Lambda Q)$$

for some $\Lambda \geqslant \lambda > 1$. This kind of weak boundedness appears when we regularize (consider truncations of) Calderón–Zygmund operators, defined initially on Lipschitz or smooth functions, see §11 below.

Clearly, if the above condition holds for some $\lambda > 1$, it holds for all $\lambda \in (1, \Lambda)$, so we can assume that λ is as close to 1 as we want.

The only modification one has to do to the proof concerns §10.2.2. One just has to cut off different neighborhoods of the grid G (see §10.2.1) from the cubes Q and R. For example, cut $G_{\varepsilon'}$ off R, but cut only $G_{\varepsilon'/2}$ off Q.

More precisely, in doing estimate (10.8) one has to define Δ'_R, $\tilde{\Delta}_R$ exactly as they were defined, but put $\Delta'_Q := \Delta_Q \cap G_{\varepsilon'/2}$ and $\tilde{\Delta}_Q := \Delta_Q \setminus G_{\varepsilon'/2}$.

The rest of the proof remains the same.

11. Reduction to the case of a priori bounds

In this section we are going to consider the case when the bilinear form is defined for smooth functions or for Lipschitz functions, as in §§ 0.3.2 and 0.3.1 respectively.

We are going to reduce these cases to the case when we have a priori bounds on T. Namely, first we are going to show that if $Tb_1 \in \mathrm{BMO}^p_\lambda(\mu)$ for some p, $1 \leqslant p < \infty$ (in particular, if $Tb_1 \in \mathrm{BMO}^1_\lambda(\mu)$), and the operator T is weakly bounded, then $Tb_1 \in \mathrm{RBMO}$, and therefore $Tb_1 \in \mathrm{BMO}^2_\lambda(\mu)$.

Then we show that under the same assumptions the condition $Tb_1 \in \mathrm{BMO}^2_\lambda(\mu)$ implies $T_\varepsilon b_1 \in \mathrm{BMO}^2_\Lambda(\mu)$ (for some $\Lambda > \lambda$) for all truncated operators T_ε with uniform estimates on BMO-norms, and that the truncated operators $M_{b_2} T_\varepsilon M_{b_1}$ are weakly bounded (with uniform estimates on constants).

11.1. The bilinear form is defined on smooth functions

We assume that the bilinear form $\langle Tb_1 f, b_2 g\rangle$ of the operator $M_{b_2} T M_{b_1}$ is well defined for all smooth (say, C^∞) compactly supported f and g.

We consider the following version of the weak boundedness assumption. Fix a C^∞-function σ on $[0,\infty)$ such that $0 \leqslant \sigma \leqslant 1$, $\sigma \equiv 1$ on $[0,a]$ $(0 < a < 1)$ and $\sigma \equiv 0$ on $[1,\infty)$, see

Figure 2. The parameter a is not essential here, but we will already have too many parameters in what follows, so let us fix some a, say $a=0.9$.

For a ball $B=B(x_0,r)$, let $\sigma_B(x):=\sigma(|x-x_0|/r)$. Clearly, σ_B is supported by the ball B and is identically 1 on the ball $0.9B$. We will require that for any ball B,

$$|\langle T\sigma_B b_1, \sigma_{2B} b_2\rangle| \leqslant C\mu(3B), \quad |\langle T\sigma_{2B} b_1, \sigma_B b_2\rangle| \leqslant C\mu(3B). \tag{11.1}$$

The parameters 3 and 2 are not essential here, and can be replaced by any numbers $\beta>\alpha>1/a>1$. In the classical theory an even stronger version of this condition is assumed, see [1, p. 49]. We should also mention that for antisymmetric kernels (when the operator is treated as the *canonical value*) and $b_1=b_2=b$, this condition holds, see Corollary 11.4 below.

Let us recall that a function b is called *sectorial* if $b\in L^\infty$ and there exists a constant $\xi\in\mathbf{C}$, $|\xi|=1$, such that $\operatorname{Re}\xi b\geqslant\delta>0$.

THEOREM 11.1. *Let the bilinear form $\langle Tb_1 f, b_2 g\rangle$ be defined for smooth (C^∞) compactly supported functions, and let T_r be truncated operators. Suppose also that for a function $b_1\in L^\infty$ and a sectorial function b_2, the estimate* (11.1) *holds for any ball B.*

Then the condition $Tb_1\in\mathrm{BMO}^p_\Lambda(\mu)$ (for some p, $1\leqslant p<\infty$) implies that $Tb_1\in\mathrm{RBMO}(\mu)$ (and therefore, $Tb_1\in\mathrm{BMO}^2_\Lambda(\mu)$).

THEOREM 11.2. *Let T be a Calderón–Zygmund operator (with bilinear form $\langle Tb_1 f, b_2 g\rangle$ defined for smooth (C^∞) compactly supported functions), and let T_r be truncated operators. Suppose also that for a function $b_1\in L^\infty$ and a sectorial function b_2, the estimate* (11.1) *holds for any ball B. Then the condition $Tb_1\in\mathrm{BMO}^2_\Lambda(\mu)$ implies that*

$$T_r b_1\in\mathrm{BMO}^2_\Lambda(\mu),$$

uniformly in r, where $\Lambda=14\lambda$.

Moreover,

$$|\langle Tb_1\chi_{\Lambda Q}, b_2\chi_Q\rangle| \leqslant C\mu(\Lambda Q)$$

for all cubes Q.

As we said above, the estimate (11.1) holds for antisymmetric Calderón–Zygmund operators. Namely, let K be an antisymmetric Calderón–Zygmund kernel ($K(x,y)=-K(y,x)$). Let T be the corresponding operator defined in the sense of *canonical value*, i.e.,

$$\langle Tbf, bg\rangle = \frac{1}{2}\iint K(x,y)[f(y)g(x)-f(x)g(y)]\, b(x)b(y)\, d\mu(x)\, d\mu(y)$$

for Lipschitz compactly supported f and g.

LEMMA 11.3. *Let φ_1, φ_2 be Lipschitz functions, $|\varphi_{1,2}(x)-\varphi_{1,2}(y)| \leqslant L \cdot |x-y|$, supported by bounded sets D_1, D_2, respectively, and such that $\|\varphi_{1,2}\|_\infty \leqslant 1$. Then for $b \in L^\infty$,*

$$|\langle Tb\varphi_1, b\varphi_2\rangle| \leqslant CL \cdot \|b\|_\infty^2 \cdot \operatorname{diam}(D_1) \cdot \mu(D_2).$$

Proof. Notice that

$$\begin{aligned}|\varphi_1(y)\varphi_2(x)-\varphi_1(x)\varphi_2(y)| &= |\varphi_1(y)\varphi_2(x)-\varphi_1(y)\varphi_2(y)+\varphi_1(y)\varphi_2(y)-\varphi_1(x)\varphi_2(y)| \\ &\leqslant |\varphi_1(y)(\varphi_2(x)-\varphi_2(y))|+|\varphi_2(y)(\varphi_1(x)-\varphi_1(y)| \\ &\leqslant 2L|x-y|.\end{aligned}$$

By property (i) of Calderón–Zygmund kernels we have for the function

$$F(x,y) = K(x,y) \cdot [\varphi_1(y)\varphi_2(x)-\varphi_1(x)\varphi_2(y)] \cdot b(x)b(y)$$

the estimate $|F(x,y)| \leqslant CL \cdot \|b\|_\infty^2 \cdot |x-y|^{-d+1}$. One can estimate

$$|\langle Tb\varphi_1, b\varphi_2\rangle| \leqslant \iint_{D_1\times D_2} |F(x,y)|\, d\mu(x)\, d\mu(y) + \iint_{D_2\times D_1} |F(x,y)|\, d\mu(x)\, d\mu(y).$$

The Comparison Lemma (Lemma 2.1) implies

$$\int_{D_1} |F(x,y)|\, d\mu(x) \leqslant C'L \cdot \|b\|_\infty^2 \operatorname{diam}(D_1).$$

Integrating once more over D_2 with respect to $d\mu(y)$ we get

$$\iint_{D_1\times D_2} |F(x,y)|\, d\mu(x)\, d\mu(y) \leqslant C'L \operatorname{diam}(D_1)\mu(D_2).$$

The second integral can be estimated similarly, one only has to change the order of integration. □

COROLLARY 11.4. *For the antisymmetric operator T defined above as canonical value the inequality*

$$|\langle Tb\sigma_{B_1}, b\sigma_{B_2}\rangle| \leqslant C\mu(B_2) \quad (b \in L^\infty)$$

holds for concentric balls $B_1 \subset B_2$ of comparable diameter, $\operatorname{diam}(B_2) \leqslant 2\operatorname{diam}(B_1)$.

Proof. Let r be the radius of the ball B_1. The functions $\sigma_{B_{1,2}}$ are Lipschitz functions with norms at most C/r, i.e.,

$$|\sigma_1(x)-\sigma_1(y)| \leqslant \frac{C}{r}|x-y|, \quad |\sigma_2(x)-\sigma_2(y)| \leqslant \frac{C}{r}|x-y|,$$

and the result follows trivially from Lemma 11.3. □

The next lemma holds for an arbitrary integral operator (whose bilinear form $\langle Tf, g\rangle$ is defined on smooth functions with compact supports) with kernel K satisfying $|K(x,y)| \leqslant C|x-y|^{-d}$. We are going to apply it later to the operator $(M_{b_2}TM_{b_1})^*$ where T is a Calderón–Zygmund operator.

LEMMA 11.5. *Suppose that the operator T satisfies*

$$|\langle T\sigma_B, \sigma_{2B}\rangle| \leqslant C\mu(3B)$$

for any B.

Then for any two concentric balls $B_1 \subset B_2$ of radii r and R respectively, $R/r \geqslant 2$,

$$|\langle T\sigma_{B_1}, \sigma_{B_2}\rangle| \leqslant C\cdot\left(\mu(3B_1)+\mu(B_1)\log\frac{R}{r}\right). \tag{11.2}$$

Remark 11.6. Clearly, in the conclusion of the lemma one can replace σ_{B_2} by χ_{B_2}: the result will be the same.

Remark 11.7. In what follows, the exact expression $C\cdot(1+\log(R/r))$ for the multiplier at $\mu(3B_1)$ in the estimate is not essential. What is essential is that this expression depends only on the ratio R/r (which will be large but fixed in what follows), but does not depend on the ratio $\mu(B_2)/\mu(B_1)$, which can be arbitrary large, because the measure μ is not doubling.

Proof of Lemma 11.5. First, we can assume that $R>1.2r$, because otherwise the conclusion is trivial.

Let x_0 be the center of the balls B_1, B_2. Denote $\sigma_{1,2} := \sigma_{B_{1,2}}$, and let $\varphi := \sigma_{2B_1}$, $\psi := 1-\varphi$. Then

$$\langle T\sigma_1, \sigma_2\rangle = \langle T\sigma_1, \varphi\rangle + \langle T\sigma_1, \psi\sigma_2\rangle$$

because $\varphi\sigma_2 = \varphi$.

By the assumption, we have

$$|\langle T\sigma_1, \varphi\rangle| \leqslant C\mu(3B_1).$$

The second term is also easy to estimate. Due to the estimate on the kernel K,

$$|[T\sigma_1](x)| \leqslant \frac{C\mu(B_1)}{\operatorname{dist}(x, B_1)^d} = \frac{C\mu(B_1)}{(|x-x_0|-r)^d},$$

where x_0 is the center of the balls B_1, B_2. Since $\psi(x)=0$ for $|x-x_0|<1.8r$, we can write

$$\begin{aligned}|\langle T\sigma_1, \psi\sigma_2\rangle| &\leqslant \int_{1.8r\leqslant|x-x_0|<R} \frac{C\mu(B_1)}{(|x-x_0|-r)^d}\\ &\leqslant \int_{1.8r\leqslant|x-x_0|<R} \frac{C\mu(B_1)13^d}{|x-x_0|^d} \leqslant C'\mu(B_1)\log\frac{R}{r}.\end{aligned}$$

Adding the estimates, we get the desired conclusion. □

Let us remind the reader that in the following lemma BMO means the "ball" BMO, i.e., all averages are taken over balls, not over the cubes.

Let us also remind the reader that a function f is called *sectorial* if $f \in L^\infty$ and there exists $\xi \in \mathbf{C}$, $|\xi|=1$, such that $\operatorname{Re}\xi f \geqslant \delta > 0$.

LEMMA 11.8. *Let T be a Calderón–Zygmund operator, let $b_1 \in L^\infty$, and let b_2 be a sectorial function. Suppose also that*

$$|\langle Tb_1\sigma_{2B}, b_2\sigma_B\rangle| \leqslant C\mu(3B)$$

for any concentric balls $B\subset B'$. Suppose also that $Tb_1 \in \mathrm{BMO}^p_\lambda(\mu)$, $\lambda\geqslant 2$, for some p, $1\leqslant p<\infty$.

Then for a ball B,

$$\int_B |Tb_1\sigma_B|^p\,d\mu \leqslant C\mu(\mathcal{B}),$$

where $\mathcal{B}=2\lambda B$.

Proof. The idea of the proof is quite simple. First of all notice that the assumption $\lambda\geqslant 2$ is not a restriction. The condition $Tb_1\in\mathrm{BMO}^p_\lambda(\mu)$ implies that Tb_1 restricted to the ball B belongs "up to an additive constant" to $L^p(\mu|B)$, and the weak boundedness (11.1) will imply that the constant is not too big.

Let g be a smooth function supported by the ball B, $\|\varphi\|_{L^q(\mu)}=1$, $1/p+1/q=1$. We want to estimate $|\langle Tb_1\chi_B, b_2 g\rangle|$.

Pick a constant c such that

$$c\int_{2B}\sigma_{2B}b_2\,d\mu = \int_B b_2 g\,d\mu,$$

i.e., such that $\int(b_2g-cb_2\sigma_{2B})\,d\mu=0$.

Since b_2 is sectorial, $\left|\int_{2B}\sigma_{2B}b_2\,d\mu\right|\geqslant\delta\mu(B)$, and we have

$$|c|\leqslant\delta^{-1}\mu(B)^{-1}\int_B|b_2g|\,d\mu\leqslant\delta^{-1}\mu(B)^{-1}\|b_2\|_\infty\|g\|_{L^q(\mu)}\mu(B)^{1/p}=\delta^{-1}\|b_2\|_\infty\cdot\mu(B)^{-1/q}.$$

Since $|\sigma_{2B}|\leqslant 1$ and b_2 is sectorial,

$$\int|\sigma_{2B}|^q\,d\mu\leqslant\int|\sigma_{2B}|\,d\mu\leqslant\delta^{-1}\left|\int\sigma_{2B}b_2\,d\mu\right|.$$

On the other hand, we know that

$$|c|\cdot\left|\int_{2B}\sigma_{2B}b_2\,d\mu\right|=\left|\int_B b_2g\,d\mu\right|\leqslant\|b_2\|_\infty\|g\|_{L^2(\mu)}\mu(B)^{1/p}=\|b_2\|_\infty\mu(B)^{1/p}.$$

Combining this with the above estimate for $|c|$ we get

$$|c|^q\int|\sigma_{2B}|^q\,d\mu\leqslant|c|^{q-1}|c|\cdot\delta^{-1}\left|\int\sigma_{2B}b_2\,d\mu\right|\leqslant C^{1/q}\mu(B)^{-(q-1)/q}\mu(B)^{1/p}=C^{1/q},$$

i.e., $\|c\sigma_{2B}b_2\|_{L^q(\mu)} \leqslant C$.

Therefore for $\varphi = g - c\sigma_{2B}$ we have $\|\varphi\|_{L^q(\mu)} \leqslant C+1$ and $\int \varphi b_2 \, d\mu = 0$.

Then

$$\langle Tb_1, \varphi b_2\rangle = \langle T(1-\sigma_{\mathcal{B}})b_1, \varphi b_2\rangle + \langle Tb_1\sigma_{\mathcal{B}}, b_2\varphi\rangle.$$

Since the supports of φ and $1-\sigma_{\mathcal{B}}$ are separated, using Lemma 2.2 (for balls instead of cubes) and the Comparison Lemma (Lemma 2.1) we can estimate the first term as

$$|\langle T(1-\sigma_{\mathcal{B}})b_1, b_2\varphi\rangle| \leqslant C\|\varphi\|_{L^1(\mu)} \leqslant C\mu(B)^{1/p}\|\varphi\|_{L^q(\mu)} = C\mu(B)^{1/p}.$$

We know that $Tb_1 \in \mathrm{BMO}_\lambda^p(\mu)$, and therefore

$$|\langle Tb_1, b_2\varphi\rangle| \leqslant C\mu(2\lambda B)^{1/p}\|\varphi\|_{L^q(\mu)}$$

(φ is supported by $2B$). It follows that

$$|\langle Tb_1\sigma_{\mathcal{B}}, b_2\varphi\rangle| \leqslant C\mu(2\lambda B)^{1/p}\|\varphi\|_{L^q(\mu)}.$$

Lemma 11.5 implies that $|\langle Tb_1\sigma_{\mathcal{B}}, b_2\sigma_{\mathcal{B}}\rangle| \leqslant C\mu(3B) \leqslant C\mu(\mathcal{B})$, so

$$|\langle Tb_1\sigma_{\mathcal{B}}, cb_2\sigma_{\mathcal{B}}\rangle| \leqslant C'\mu(\mathcal{B})\mu(B)^{-1/q} \leqslant C'\mu(\mathcal{B})^{1/p}.$$

Thus

$$|\langle Tb_1\sigma_{\mathcal{B}}, b_2 g\rangle| \leqslant C\mu(\mathcal{B})^{1/p}.$$

We are done. □

Proof of Theorem 11.1. This proof follows the lines of the proof of Theorem 2.4 with the only modification that one has to use Lemma 11.8 instead of Lemma 2.5. We leave the details to the reader. □

Proof of Theorem 11.2. Fix some ball B. First of all notice that we need to prove the conclusion of the theorem only for small r, say for $r < 0.1\,\mathrm{diam}(B)$.

Indeed, let $r \geqslant 0.1\,\mathrm{diam}(B)$. Then

$$|T_r b_1 \chi_{2B}(x)| \leqslant \frac{C}{r^d}\mu(2B) \leqslant C',$$

and so

$$\int_B |T_r b_1 \chi_{2B}|^2 \, d\mu \leqslant C'\mu(B).$$

On the other hand, for φ supported by the ball B and satisfying $\int \varphi \, d\mu = 0$ we have (cf. Lemmas 2.2 and 2.1)

$$|\langle T_r(1-\chi_{2B}), \varphi\rangle| \leqslant C\|\varphi\|_{L^1(\mu)} \leqslant C\mu(B)^{1/2}\|f\|_{L^2(\mu)} \tag{11.3}$$

(this inequality holds for all r), so for $r \geqslant 0.1\,\mathrm{diam}(B)$ we even have inclusion in $\mathrm{BMO}_1^2(\mu)$.

So, let us suppose that $r<0.1\,\mathrm{diam}(B)$. Define $B_0:=7B$, and let $\mathcal{B}:=2\lambda B_0=\Lambda B$. We want to show that

$$\int_B |T_r b_1 \sigma_{\mathcal{B}}|^2\,d\mu \leqslant C\mu(\mathcal{B}). \tag{11.4}$$

This would imply $T_r b_1 \in \mathrm{BMO}_\lambda^2(\mu)$, because as we already know, for any φ supported by the ball B and satisfying $\int \varphi\,d\mu=0$, we have (cf. (11.3))

$$|\langle T_r(1-\chi_{\mathcal{B}}),\varphi\rangle| \leqslant C\|\varphi\|_{L^1(\mu)} \leqslant C\mu(B)^{1/2}\|\varphi\|_{L^2(\mu)}.$$

The condition (11.4) also implies the weak boundedness condition

$$|\langle Tb_1\chi_{\Lambda B}, b_2\chi_B\rangle| \leqslant C\mu(\Lambda B),$$

so if we prove (11.4), we are done.

To prove the inequality (11.4) we are going to apply a modification of what we called "the Guy David trick" in [22, §4].

Let $x\in B$ and $r<0.1\,\mathrm{diam}(B)$ be fixed. Consider a sequence of balls $B^j=B(x,r_j)$, $r_j=2^J r$. Let $\mu_j:=\mu(B^j)$. Let n be the smallest number such that either $\mu_n \leqslant 2\cdot 3^d \mu_{n-1}$ or $B\subset B^n$.

Let $R=r_{n-1}=3^{n-1}r$. Let us estimate the difference

$$\begin{aligned}|[T_r b_1\sigma_{\mathcal{B}}](x)-[T_{3R} b_1\sigma_{\mathcal{B}}](x)| &\leqslant \int_{B^n\setminus B^0} |K(x,y)b_1(y)\sigma_{\mathcal{B}}(y)|\,d\mu(y)\\ &\leqslant C\sum_{j=1}^n \int_{B^k\setminus B^{k-1}} |K(x,y)|\,d\mu(y)=\sum_{j=1}^n \mathcal{I}_j.\end{aligned}$$

Let us recall now that $|K(x,y)|\leqslant A|x-y|^d$, and therefore

$$\mathcal{I}_j \leqslant A\frac{\mu_j}{r_{j-1}^d}=A\frac{\mu_j}{r_{j-1}^d}, \quad j=1,\dots,n.$$

By construction, $\mu_j \leqslant [2\cdot 3^d]^{j+1-n}\mu_{n-1}$ for $j=0,\dots,n-1$, and therefore

$$\sum_{j=1}^{n-1}\mathcal{I}_j \leqslant A\sum_{j=1}^{n-1}\frac{\mu_j}{r_{j-1}^d} \leqslant A\cdot 2\cdot 3^d\sum_{j=1}^{n-1} 2^{j-k} \leqslant A\cdot 2\cdot 3^d.$$

The last term can be estimated as $\mathcal{I}_n \leqslant A\mu_n/r_{n-1}^d \leqslant C$, and therefore

$$|[T_r b_1\sigma_{\mathcal{B}}](x)-[T_{3R} b_1\sigma_{\mathcal{B}}](x)| \leqslant C.$$

Now we want to estimate $|[T_{3R}b_1\sigma_{\mathcal{B}}](x)|$. If we stopped because $B\subset B^n$, then $3R\geqslant \frac{1}{2}\operatorname{diam}(B)$, and in this case we know that $|[T_{3R}b_1\sigma_{\mathcal{B}}](x)|\leqslant C$. Therefore we now can assume that $\mu_n\leqslant 2\cdot 3^d\mu_{n-1}$, i.e., we are now in the *doubling* situation!

Let $\sigma:=\sigma_{B(x,1.2R)}$, so $\sigma\equiv 1$ on $B^{n-1}=B(x,R)$.

Denote $A:=\int b_2\sigma\,d\mu$, and let us compare $[T_{3R}b_1\sigma_{\mathcal{B}}](x)$ to the average

$$V(x):=V_R(x):=A^{-1}\int b_1\sigma[Tb_1\sigma_{\mathcal{B}}]\,d\mu. \tag{11.5}$$

Since b_2 is sectorial, $A\geqslant\delta\mu(B(x,R))$, and therefore

$$\begin{aligned}|V_R(x)|&\leqslant\frac{1}{\delta\cdot\mu(B(x,R))}\int_{B(x,1.2R)}|b_1\sigma[Tb_1\sigma_{\mathcal{B}}]|\,d\mu\\&\leqslant\frac{\mu(B(x,3R))}{\delta\cdot\mu(B(x,R)}\cdot\|b_2\|_\infty\cdot\widetilde{M}|\chi_{B(x,1.2R)}\cdot Tb_1\sigma_{\mathcal{B}}|\\&\leqslant\delta^{-1}2\cdot 3^d\|b_2\|_\infty\cdot\widetilde{M}|\chi_{B_0}\cdot Tb_1\sigma_{\mathcal{B}}|,\end{aligned}$$

where $\widetilde{M}$ is the maximal operator,

$$\widetilde{M}f(x):=\sup_{r>0}\mu(B(x,2.5r))^{-1}\cdot\int_{B(x,r)}|f(y)|\,d\mu(y)$$

(in the last inequality we replaced $\chi_{B(x,1.2R)}$ by χ_{B_0} because $B(x,1.2R)\subset B_0$).

We know that the operator $\widetilde{M}$ is bounded on $L^2(\mu)$, see Lemma 3.1 in [22].

We have

$$\begin{aligned}[T_{3R}b_1\sigma_{\mathcal{B}}](x)-V_R(x)&=\int_{\mathcal{B}\setminus B(x,3R)}b_1\sigma_{\mathcal{B}}[T^*\delta_x]\,d\mu(y)\\&\quad-A^{-1}\int\sigma b_2\cdot[Tb_1\cdot(1-\chi_{B(x,3R)})\sigma_{\mathcal{B}}]\,d\mu\\&\quad-A^{-1}\int\sigma b_2\cdot[Tb_1\cdot\chi_{B(x,3R)}\sigma_{\mathcal{B}}]\,d\mu\\&=\int_{\mathcal{B}\setminus B(x,3R)}b_1\sigma_{\mathcal{B}}\cdot[T^*(\delta_x-A^{-1}\sigma b_2)]\,d\mu\\&\quad-A^{-1}\int\sigma b_2\cdot[Tb_1\chi_{B(x,3R)}]\,d\mu.\end{aligned}$$

We know that $\int\delta_x-A^{-1}\sigma b_2\,d\mu=0$, and therefore the first term is bounded by

$$C|A|^{-1}\|\sigma b_2\|_{L^1(\mu)}\leqslant C'.$$

The second term also can be estimated using Lemma 11.5 (see Remark 11.6) by

$$A^{-1}C\cdot\mu(B(x,1.2^2R))\leqslant A^{-1}C\cdot\mu(B(x,3R))$$
$$\leqslant A^{-1}C\cdot 2\cdot 3^d\cdot\mu(B(x,R))\leqslant\delta^{-1}C\cdot 2\cdot 3^d.$$

Summarizing everything we get for $x\in B$ the estimate

$$|[T_r b_1\sigma_{\mathcal{B}}](x)|\leqslant C_1+C_2\widetilde{M}|\chi_{B_0}\cdot Tb_1\sigma_{\mathcal{B}}|.$$

By Lemma 11.8 for $p=2$,

$$\|\chi_{B_0}\cdot Tb_1\sigma_{\mathcal{B}}\|^2_{L^2(\mu)}\leqslant C\mu(\mathcal{B}).$$

Since the operator $\widetilde{M}$ is bounded on $L^2(\mu)$,

$$\int_B|[T_r b_1\sigma_{\mathcal{B}}](x)|^2\,d\mu(x)\leqslant C\mu(\mathcal{B}),$$

and we are done! □

LEMMA 11.9. *The modified maximal function operator $\widetilde{M}$ is bounded on $L^p(\mu)$ for each $p\in(1,+\infty]$ and acts from $L^1(\mu)$ to $L^{1,\infty}(\mu)$.*

Proof. The boundedness on $L^\infty(\mu)$ is obvious. To prove the weak type $(1,1)$ estimate, we will use the celebrated

VITALI COVERING THEOREM. *Let $\mathcal{X}$ be a separable measure space with measure. Fix some $R>0$. Let $E\subset\mathcal{X}$ be any set and let $\{B(x,r_x)\}_{x\in E}$ be a family of balls of radii $0<r_x<R$. Then there exists a countable subfamily $\{B(x_j,r_j)\}_{j=1}^\infty$ (where $x_j\in E$ and $r_j:=r_{x_j}$) of disjoint balls such that $E\subset\bigcup_j B(x_j,2.5r_j)$ (2.5 can be replaced by $2+\varepsilon$, $\varepsilon>0$, here).*

For the proof of the Vitali covering theorem, we refer the reader to his favorite textbook in geometric measure theory.

Now, to prove the lemma, fix some $t>0$. Pick $R>0$ and consider the set E of the points $x\in\operatorname{supp}\mu$ for which

$$\widetilde{M}^{(R)}f(x):=\sup_{0<r<R}\frac{1}{\mu(B(x,3r))}\int_{B(x,r)}|f|\,d\mu>t.$$

For every such x, there exists some radius $r_x\in(0,R)$ such that

$$\int_{B(x,r_x)}|f|\,d\mu>t\mu(B(x,3r_x)).$$

Choose the corresponding collection of pairwise disjoint balls $B(x_j, r_j)$. We have

$$\mu(E) \leqslant \sum_j \mu(B(x_j, 3r_j)) \leqslant \frac{1}{t} \sum_j \int_{B(x_j,r_j)} |f|\, d\mu \leqslant \frac{\|f\|_{L^1(\mu)}}{t}.$$

It remains only to note that $\widetilde{M}^{(R)} f \nearrow \widetilde{M} f$ as $R \to +\infty$.

The boundedness on $L^p(\mu)$ for $1<p<+\infty$ follows now from the Marcinkiewicz interpolation theorem. □

11.2. The bilinear form is defined on Lipschitz functions

In this section we assume that the bilinear form $\langle b_2 T b_1 f, g\rangle$ is well defined for compactly supported Lipschitz functions f, g.

Let $|\cdot|$ denote the "l^∞-norm" on $\mathbf{R}^N$, $|x| := \max\{|x_k| : 1 \leqslant k \leqslant N\}$, so the "balls" in this norm are just cubes. We fixed the "l^∞-norm" on $\mathbf{R}^N$ because we have to use cubes in the definition of weak accretivity. The results of this section hold for an arbitrary norm $|\cdot|$, if weak accretivity means that the averages over the balls in this norm are large.

By weak boundedness in this case we mean the following two conditions:

(i) For all pairs of Lipschitz functions φ_1, φ_2 satisfying $|\varphi_{1,2}(x) - \varphi_{1,2}(y)| \leqslant L \cdot |x-y|$, supported by bounded sets D_1, D_2, respectively, and such that $\|\varphi_{1,2}\|_\infty \leqslant 1$, the inequalities

$$|\langle T b_1 \varphi_1, b_2 \varphi_2\rangle|, \; |\langle T b_1 \varphi_2, b_2 \varphi_1\rangle| \leqslant CL \cdot \|b_1\|_\infty \cdot \|b_2\|_\infty \cdot \operatorname{diam}(D_1) \cdot \mu(D_2)$$

should hold for weakly accretive functions b_1, b_2 (this is for the Tb-theorem, for the $T1$-theorem $b_1 = b_2 = 1$).

As Lemma 11.3 above shows, this is true for antisymmetric kernels.

(ii) Let σ^ε be the function as in Figure 1. For a ball (cube) $Q = Q(x_0, r) = \{x \in \mathbf{R}^N : |x - x_0| \leqslant r\}$ let

$$\sigma_Q^\varepsilon := \sigma^\varepsilon(|x - x_0|/r).$$

(Clearly σ_Q^ε is a Lipschitz function with Lipschitz norm at most $C/r\varepsilon$.)

We will require that for all cubes Q,

$$|\langle T b_1 \sigma_Q^\varepsilon, b_2 \sigma_Q^\varepsilon\rangle| \leqslant C\mu(\lambda' Q)$$

for some $\lambda' \geqslant 1$, uniformly in ε and Q.

Definitely, the last condition holds for antisymmetric kernels, since $\langle T b \sigma_Q^\varepsilon, b \sigma_Q^\varepsilon\rangle = 0$.

THEOREM 11.10. *Let T be a Calderón–Zygmund operator such that the bilinear form $\langle Tb_1 f, b_2 g\rangle$ is defined for Lipschitz compactly supported f and g. Suppose that T is weakly bounded as above.*

If $Tb_1 \in \mathrm{BMO}^p_\lambda(\mu)$ for some $p\in[1,\infty)$, $\lambda>1$, then $Tb_1\in\mathrm{RBMO}(\mu)$ (and therefore $Tb_1\in\mathrm{BMO}^2_\lambda(\mu)$).

THEOREM 11.11. *Let T be a Calderón–Zygmund operator as in the previous theorem, $b_1\in L^\infty$, and let b_2 be a weakly accretive function. If $Tb_1\in\mathrm{BMO}^2_\lambda(\mu)$, then for truncated operators T_r we have $T_r b_1\in\mathrm{BMO}^2_\Lambda(\mu)$, $\Lambda=14\lambda$, with uniform estimates on the norms. Moreover, the weak boundedness condition*

$$|\langle T_\varepsilon b_1 \chi_{2Q}, b_2\chi_Q\rangle| \leqslant C\mu(3Q)$$

holds for all cubes Q.

Let us recall that weakly accretive means $\mu(Q)^{-1}\left|\int_Q b\,d\mu\right|\geqslant\delta$ for all cubes Q. Let us also recall that $|\cdot|$ means the "l^∞-distance" $|x-y|:=\|x-y\|_\infty:=\max\{|x_k-y_k|:1\leqslant N\}$ on $\mathbf{R}^N$, and the theorem implies that for "cubic" truncated operators T^{c}_r,

$$T^{\mathrm{c}}_r f(x) := \int_{\|x-y\|_\infty>r} K(x,y) f(y)\,d\mu(y),$$

we have $T^{\mathrm{c}}_r\in\mathrm{BMO}^2_\Lambda(\mu)$ (with uniform estimates on norms). However, since the differences $T_r-T^{\mathrm{c}}_r$ (T_r is the usual truncation, where one integrates over the set $\|x-y\|_2>r$) are uniformly bounded, the same holds for T_r.

The proof of the theorem is very similar to the proof of Theorem 11.2. Let us introduce functions σ^ε as in Figure 1. We denote $\sigma:=\sigma^{0.1}$.

For a cube (ball in the norm $|\cdot|$) $B=B(x_0,r)$ let $\sigma^\varepsilon_B:=\sigma^\varepsilon(|x-x_0|/r)$. Clearly σ^ε_B is a Lipschitz function with Lipschitz norm $1/r\varepsilon$.

LEMMA 11.12. *Let $M_{b_2}TM_{b_1}$ be weakly bounded, as it was defined in the beginning of this section, and let λ' be the blow-up constant in the definition of weak boundedness, i.e.,*

$$|\langle Tb_1\sigma^\varepsilon_Q, b_2\sigma^\varepsilon_Q\rangle| \leqslant C\mu(\lambda' Q)$$

for all cubes Q (uniformly in ε and Q).

Given $R>0$ let R_0, $R\leqslant R_0\leqslant 1.2R$, be as above in Lemma 2.8. Then for all $\varepsilon>0$,

$$|\langle Tb_1\sigma_{B(x_0,3R)}, b_2\sigma^\varepsilon_{B(x_0,R_0)}\rangle| \leqslant C\mu(B(x_0,\Lambda R)),$$

where $\Lambda=\max(1.2\lambda',3)$, and C does not depend on ε.

Proof. Since $M_{b_2}TM_{b_1}$ is weakly bounded,

$$|\langle Tb_1\sigma^\varepsilon_{B(x_0,R_0)}, b_2\sigma^\varepsilon_{B(x_0,R_0)}\rangle| \leqslant C\mu(B(x_0,\lambda' R_0)) \leqslant C\mu(B(x_0,\Lambda R)),$$

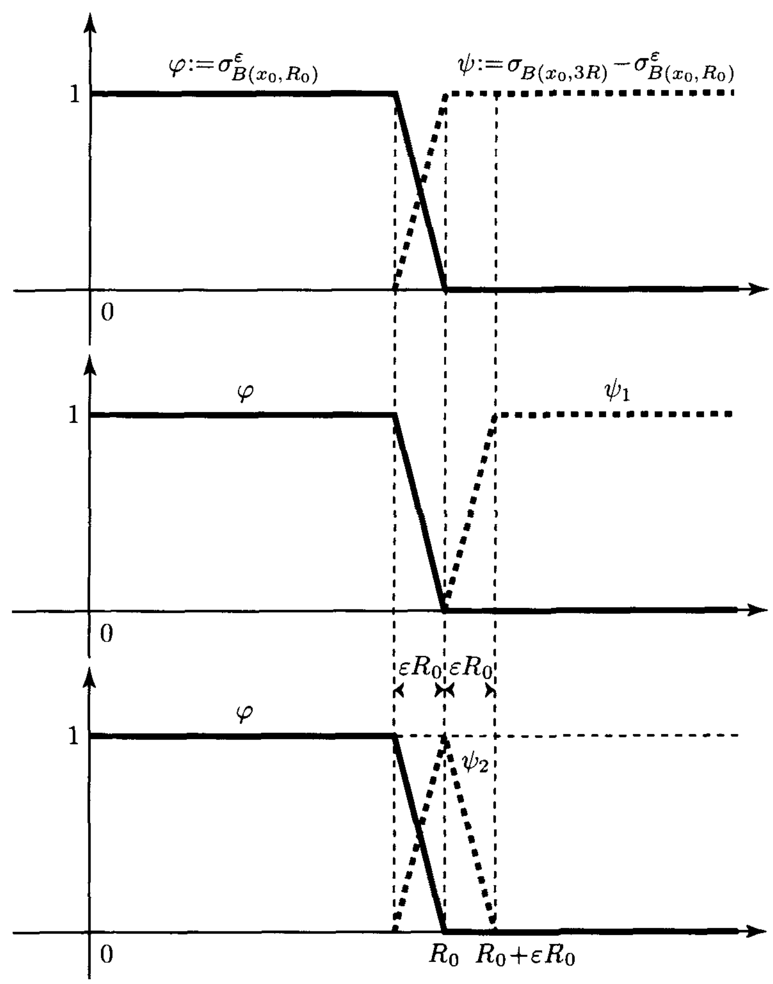

Fig. 11. Splitting of the function ψ.

so it remains to estimate

$$\langle Tb_1\cdot(\sigma_{B(x_0,3R)}-\sigma^{\varepsilon}_{B(x_0,R_0)}), b_2\sigma^{\varepsilon}_{B(x_0,R_0)}\rangle=\langle T\psi b_1,\varphi b_2\rangle,$$

where $\psi:=\sigma_{B(x_0,3R)}-\sigma^{\varepsilon}_{B(x_0,R_0)}$, $\varphi:=\sigma^{\varepsilon}_{B(x_0,R_0)}$.

Split $\psi=\psi_1+\psi_2$ as in Figure 11. Then

$$\langle Tb_1\psi, b_2\varphi\rangle=\langle Tb_1\psi_1, b_2\varphi\rangle+\langle Tb_1\psi_2, b_2\varphi\rangle.$$

Since $\|\varphi\|_\infty\leqslant 1$, $\|\psi_1\|_\infty\leqslant 1$, and the functions φ and ψ_1 are supported by $B(x_0,R_0)$ and $B(x_0,3R)\setminus B(x_0,R_0)$ respectively, the first term can be estimated by Lemma 2.9:

$$|\langle Tb_1\psi_1, b_2\varphi\rangle|\leqslant C\mu(B(x_0,3R)).$$

By condition (i) in the definition of weak boundedness the second term can be estimated as

$$|\langle Tb_1\psi_2, b_2\varphi\rangle| \leqslant C\cdot\frac{1}{\varepsilon R_0}\cdot R_0\cdot\mu(\{x:\operatorname{dist}(x,S_{R_0})<\varepsilon R_0\}),$$

where $S_{R_0}:=\{x:|x-x_0|=R_0\}$; here $\operatorname{diam}(\operatorname{supp}\varphi)\leqslant 2R_0$, and ψ_2 is supported by the "annulus" $\{x:\operatorname{dist}(x,S_{R_0})<\varepsilon R_0\}$. Lemma 2.8 implies that

$$\mu(\{x:\operatorname{dist}(x,S_{R_0})<\varepsilon R_0\}) \leqslant C\varepsilon\cdot\mu(B(x_0,3R)),$$

and therefore $|\langle Tb_1\psi_2, b_2\varphi\rangle|\leqslant C\mu(B(x_0,3R))$. □

To prove Theorem 11.11 we need the following analogues of Lemmas 2.5, 2.7 and 11.8.

LEMMA 11.13. *Under the assumptions of Theorem* 11.11, *for any cube* Q,

$$\int_Q |Tb_1\chi_{2Q}|^p\,d\mu \leqslant C\mu(\Lambda Q),$$

where $\Lambda=\max(2\lambda,2\lambda',3)$.

Proof. Pick a ball (cube) $B(x_0,R)$. Lemma 11.12 implies that

$$|\langle Tb_1\chi_{B(x_0,2R)}, b_2\sigma^{\varepsilon}_{B(x_0,R_0)}\rangle| \leqslant C\mu(B(x_0,\Lambda R))$$

uniformly in ε. Taking the limit as $\varepsilon\to 0$ we get

$$|\langle Tb_1\chi_{B(x_0,2R)}, b_2\chi_{B(x_0,R_0)}\rangle| \leqslant C\mu(B(x_0,\Lambda R)). \tag{11.6}$$

Now we just repeat the proof of Lemma 2.5.

Let g be a smooth (Lipschitz) function supported by the ball $B(x_0,R)$ and such that $\|g\|_q=1$, $1/p+1/q=1$. Pick a constant c such that

$$c\int_{B(x_0,R_0)} b_2\,d\mu = \int b_2 g\,d\mu,$$

so that $\int(b_2g-cb_2\chi_{B(x_0,R_0)})\,d\mu=0$.

Weak accretivity of b_2 implies(3) $\left|\int_{B(x_0,R_0)} b_2\,d\mu\right|\geqslant\delta\mu(B(x_0,R_0))$, therefore

$$\begin{aligned}|c| &\leqslant \delta^{-1}\mu(B(x_0,R_0))^{-1}\int|b_2g|\,d\mu\\ &\leqslant C\mu(B(x_0,R_0))^{-1}\|g\|_q\,\mu(B(x_0,R))^{1/p}\leqslant C\mu(B(x_0,R_0))^{-1/q},\end{aligned}$$

(3) There is a little detail here: In the definition of weak accretivity we deal with cubes that are obtained from the cube $[0,1)^N$ by shifts and dilations, but our cube (ball) $B(x_0,R_0)$ is an open one. However, Lemma 2.8 implies that the measure μ of the boundary of the ball $B(x_0,R_0)$ is 0, so this does not present a problem.

so $\|c\chi_{B(x_0,R)}\|_q \leqslant C$. Then $\|b_2\cdot(g-c\chi_{B(x_0,R)})\| \leqslant C+1$, and the condition $Tb_1 \in \mathrm{BMO}^p_\lambda(\mu)$ implies

$$|\langle Tb_1\chi_{B(x_0,2R)}, b_2\cdot(g-c\chi_{B(x_0,R_0)})\rangle| \leqslant C\mu(B(x_0,2\lambda R)) \leqslant C\mu((B(x_0,\Lambda R)).$$

This inequality together with estimate (11.6) implies

$$|\langle Tb_1\chi_{B(x_0,2R)}, b_2 g\rangle| \leqslant C\mu(B(x_0,\Lambda R)),$$

and that is exactly what we need. □

Proof of Theorems 11.10 *and* 11.11. The proof of Theorem 11.10 follows the proof of Theorem 2.4 without any modifications. One only has to use the above Lemma 11.13 instead of Lemma 2.5.

The proof of Theorem 11.11 follows the proof of Theorem 11.2, only instead of Lemma 11.5 one has to use Lemma 11.13. We leave all the details to the reader. □

References

[1] CHRIST, M., *Lectures on Singular Integral Operators.* CBMS Regional Conf. Ser. in Math., 77. Amer. Math. Soc., Providence, RI, 1990.

[2] — A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral. *Colloq. Math.*, 60/61 (1990), 601–628.

[3] COIFMAN, R. R., JONES, P. W. & SEMMES, S., Two elementary proofs of the L^2 boundedness of Cauchy integrals on Lipschitz curves. *J. Amer. Math. Soc.*, 2 (1989), 553–564.

[4] COIFMAN, R. R. & WEISS, G., *Analyse harmonique non-commutative sur certains espaces homogènes.* Lecture Notes in Math., 242. Springer-Verlag, Berlin–Heidelberg, 1971.

[5] DAVID, G., Opérateurs intégraux singuliers sur certaines courbes du plan complexe. *Ann. Sci. École Norm. Sup. (4)*, 17 (1984), 157–189.

[6] — *Wavelets and Singular Integrals on Curves and Surfaces.* Lecture Notes in Math., 1465. Springer-Verlag, Berlin–Heidelberg, 1991.

[7] — Unrectifiable 1-sets have vanishing analytic capacity. *Rev. Mat. Iberoamericana*, 14 (1998), 369–479.

[8] — Analytic capacity, Calderón–Zygmund operators, and rectifiability. *Publ. Mat.*, 43 (1999), 3–25.

[9] DAVID, G. & JOURNÉ, J.-L., A boundedness criterion for generalized Calderón–Zygmund operators. *Ann. of Math. (2)*, 120 (1984), 371–397.

[10] DAVID, G., JOURNÉ, J.-L. & SEMMES, S., Opérateurs de Calderón–Zygmund fonctions para-accrétives et interpolation. *Rev. Mat. Iberoamericana*, 1 (1985), 1–56.

[11] DAVID, G. & MATTILA, P., Removable sets for Lipschitz harmonic functions in the plane. *Rev. Mat. Iberoamericana*, 16 (2000), 137–215.

[12] DAVIE, A. M. & ØKSENDAL, B., Analytic capacity and differentiability properties of finely harmonic functions. *Acta Math.*, 149 (1982), 127–152.

[13] GARNETT, J. B. & JONES, P. W., BMO from dyadic BMO. *Pacific J. Math.*, 99 (1982), 351–371.

[14] MATTILA, P., Removability, geometric measure theory, and singular integrals, in *Potential Theory—ICPT 94* (Kouty, 1994), pp. 129–146. de Gruyter, Berlin, 1996.

[15] MATTILA, P., MELNIKOV, M. S. & VERDERA, J., The Cauchy integral, analytic capacity, and uniform rectifiability. *Ann. of Math. (2)*, 144 (1996), 127–136.

[16] MATTILA, P. & PARAMONOV, P. V., On geometric properties of harmonic Lip_1-capacity. *Pacific J. Math.*, 171 (1995), 469–491.

[17] MELNIKOV, M. S., Analytic capacity: a discrete approach and the curvature of measure. *Mat. Sb.*, 186 (1995), 57–76 (Russian); English translation in *Sb. Math.*, 186 (1995), 827–846.

[18] MELNIKOV, M. S. & VERDERA, J., A geometric proof of the L^2 boundedness of the Cauchy integral on Lipschitz graphs. *Internat. Math. Res. Notices*, 1995, 325–331.

[19] MURAI, T., *A Real Variable Method for the Cauchy Transform, and Analytic Capacity.* Lecture Notes in Math., 1307. Springer-Verlag, Berlin–Heidelberg, 1988.

[20] NAZAROV, F. & TREIL, S., The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis. *Algebra i Analiz*, 8 (1996), 32–162; English translation in *St. Petersburg Math. J.*, 8 (1997), 721–824.

[21] NAZAROV, F., TREIL, S. & VOLBERG, A., Cauchy integral and Calderón–Zygmund operators on nonhomogeneous spaces. *Internat. Math. Res. Notices*, 1997, 703–726.

[22] — Weak type estimates and Cotlar inequalities for Calderón–Zygmund operators on nonhomogeneous spaces. *Internat. Math. Res. Notices*, 1998, 463–487.

[23] — Accretive system Tb-theorems on non-homogeneous spaces. *Duke Math. J.*, 113 (2002), 259–312.

[24] SAWYER, E. T., A characterization of a two-weight norm inequality for maximal operators. *Studia Math.*, 75 (1982), 1–11.

[25] TOLSA, X., Cotlar's inequality without the doubling condition and existence of principal values for the Cauchy integral of measures. *J. Reine Angew. Math.*, 502 (1998), 199–235.

[26] — L^2-boundedness of the Cauchy integral operator for continuous measures. *Duke Math. J.*, 98 (1999), 269–304.

[27] — BMO, H^1 and Calderón–Zygmund operators for non-doubling measures. *Math. Ann.*, 319 (2001), 89–149.

[28] — Littlewood–Paley theory and the $T(1)$ theorem with non-doubling measures. *Adv. Math.*, 164 (2001), 57–116.

[29] TREIL, S. & VOLBERG, A., Wavelets and the angle between past and future. *J. Funct. Anal.*, 143 (1997), 269–308.

[30] UY, N. X., Removable sets of analytic functions satisfying a Lipschitz condition. *Ark. Mat.*, 17 (1979), 19–27.

[31] — A removable set for Lipschitz harmonic functions. *Michigan Math. J.*, 37 (1990), 45–51.

[32] VERDERA, J., On the $T(1)$-theorem for the Cauchy integral. *Ark. Mat.*, 38 (2000), 183–199.

F. NAZAROV
Department of Mathematics
Michigan State University
East Lansing, MI 48823
U.S.A.

S. TREIL
Department of Mathematics
Brown University
151 Thayer Street
Box 1917
Providence, RI 02912
U.S.A.

A. VOLBERG
Department of Mathematics
Michigan State University
East Lansing, MI 48823
U.S.A.

Received July 17, 2000
Received in revised form December 10, 2001